

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES

ÉCOLE DOCTORALE N° 637

Sciences de la Vie et de la Santé

Spécialité : *Génétique, Génomique, Bioinformatique*

Par

Louis LE NÉZET

**Recherche des bases génétiques de la Dysplasie Coxo-Fémorale
chez le chien guide comme modèle de la Dysplasie Développementale
de la Hanche chez l'humain**

Thèse présentée et soutenue à Rennes, le 3 novembre 2025

Unité de recherche : UMR 6290 Institut de Génétique et Développement de Rennes

Rapporteurs avant soutenance :

Hervé Perdry	DR	CESP INSERM	UMR 1018	Villejuif, France
Laurent Tiret	PR	IMRB INSERM	U 955	Créteil, France

Composition du Jury :

Présidente :	Emmanuelle Génin	DR	GGB INSERM	UMR 1078	Brest, France
Examinatrice :	Sarah Djebali	CR	IRSD INSERM	UMR 1220	Toulouse, France
Dir. de thèse :	Catherine André	DR	IGDR CNRS	UMR 6290	Rennes, France
Co-dir. de thèse :	Pascale Quignon	MCU	IGDR Univ. Rennes	UMR 6290	Rennes, France

Invité :

Jean-Pierre Genevois	PR	ENV VetAgro Sup	Lyon, France

REMERCIEMENTS

Remerciements institutionnels et scientifiques

Je tiens tout d'abord à remercier les membres du jury : Dr Hervé Perdry, Pr Laurent Tiret, Dr Emmanuelle Génin, Dr Sarah Djebali, Dr Catherine André, Dr Pascale Quignon et Pr Jean-Pierre Genevois, pour avoir accepté de consacrer du temps à l'évaluation de ce travail de thèse. Votre expertise dans vos domaines respectifs et vos retours, que j'attends avec impatience, permettront d'enrichir et de poursuivre la réflexion sur ce projet au long cours.

Je remercie tout particulièrement le Dr Hervé Perdry et le Pr Laurent Tiret pour avoir accepté la tâche exigeante de rapporteurs. À l'heure où j'écris ces lignes, vos retours ne me sont pas encore parvenus, mais je suis convaincu qu'ils seront précieux pour améliorer la qualité et la rigueur de ce travail.

Je tiens à exprimer toute ma reconnaissance aux Drs Pascale Quignon et Catherine André. Merci de m'avoir accordé votre confiance dès le départ, malgré mon absence de formation initiale en bio-informatique et en génétique. Vous m'avez offert la liberté d'explorer et d'apprendre en toute autonomie, tout en guidant le projet avec bienveillance. Cette confiance m'a permis de m'épanouir pleinement dans le développement bio-informatique, un domaine qui me passionne désormais.

Cette thèse s'inscrit dans une collaboration de longue date avec l'Association des Chiens Guides d'Aveugles de l'Ouest (ACGAO) et a été rendue possible grâce au soutien financier de la Fondation VISIO. Je vous adresse mes sincères remerciements pour votre confiance et votre engagement, aussi bien sur les aspects logistiques que financiers. Le choix audacieux de financer une première thèse n'a certainement pas été anodin, mais j'espère que ce travail saura témoigner du bien-fondé de cette décision. Je remercie tout particulièrement Pascale Humbert pour votre suivi et votre confiance ainsi que Pierre Wallet pour votre engouement tout au long du projet et lien avec le CESECAH.

Je remercie également la Fondation pour la Recherche Médicale (FRM) pour le financement de ma quatrième année de doctorat. Votre soutien m'a offert un temps supplémentaire indispensable pour finaliser l'ensemble des travaux présentés ici.

Je remercie les membres de mon comité de suivi de thèse : Pr Claude Férec, Dr Anne-Louise Leutenegger et Dr Alexandre Caron, pour les échanges constructifs que nous avons pu avoir lors de nos différentes réunions. Merci à vous pour vos conseils.

Collaborateurs et partenaires du projet

Ce travail n'aurait pu aboutir sans la participation active et bienveillante de nombreux collaborateurs, institutions et experts, que je tiens à remercier très sincèrement.

Je remercie tout d'abord les différents clubs de race impliqués dans ce projet : le Retriever Club de France, la Société des Amateurs de Dogues de Bordeaux, ainsi que l'Association Française des Bouviers Suisses. Votre soutien a été déterminant, notamment pour la communication du projet, l'organisation des prélèvements et la transmission des données cliniques. Une gratitude particulière à Didier Letienne et Joëlle Bardet (AFBS), Sylviane Tompousky et Dr Didier Fau (SADB), ainsi qu'à Aline Baron et Evelyne Robinot (RCF), pour votre implication constante.

Je tiens également à remercier les vétérinaires Pr Jean-Pierre Genevois, Dr Didier Fontaine et feu Dr Francis Legeard pour votre expertise et le partage des informations cliniques dans le cadre de ce projet. Un remerciement tout particulier au Pr Jean-Pierre Genevois pour nos échanges approfondis et votre regard critique éclairé sur l'ensemble du projet.

Je souhaite également souligner l'apport du groupe interdisciplinaire de réflexion, constitué autour de cette étude, qui a permis des échanges stimulants à l'interface de la génétique, de la statistique, de la clinique et de l'anatomie comparée. Merci à Claude Férec, Emmanuelle Génin, Anthony Herzig, Virginie Scotet, Karen Rouault, François Gaucher, Christian Lefèvre, Alexandre Caron et Jean-Pierre Genevois pour la richesse de nos discussions. Ce travail collectif a permis de mieux cerner les dimensions biologiques, anatomiques et biomécaniques de la dysplasie coxo-fémorale.

Je remercie également toutes les cliniques vétérinaires ayant contribué à la collecte de prélèvements, avec une mention particulière à la clinique du Dr Chapalain pour les échantillons de chiens guides, ainsi qu'à TrioVet, pour les prélèvements tissulaires, qui ouvriront des perspectives.

Enfin, ma reconnaissance va à la société *Antagene* et en particulier à Caroline Dufaure de Citres, pour la transmission conséquente d'échantillons. Votre contribution a été déterminante pour atteindre la puissance statistique nécessaire à nos analyses.

International bio-informatics community

During this PhD, I was fortunate to actively engage with two major bioinformatics communities that played a crucial role in my training and development.

First, I warmly thank the *Bioconductor* community, whose scientific rigor and kindness supported me throughout the development of the *Pedixplorer* package. Designing a new *R* *S4* class was a major challenge that I could not have overcome without your invaluable help. A special thanks to Hervé Pagès for your thorough review during the

package submission, as well as to Jason Sinnwell for your trust and for sharing the *kinship2* source code, which served as a foundational starting point for this project.

I also wish to express my admiration and gratitude to the *nf-core* community, which introduced me to the fundamentals of modern bioinformatics development : collaborative coding, the intricate workings of *git*, code review practices, continuous integration (CI/CD), and much more. This exemplary community embodies the values of transparency, rigor, and scientific collaboration that I aspire to uphold myself. Thanks to all those who contributed to the development of the *nf-core/phaseimpute* pipeline : developers Anabella Trigila, reviewers Maxime Garcia, James Fellows Yates, Matthias Hörtenhuber, and many others.

Today, as a member of the *nf-core* maintenance team, I appreciate the tremendous behind-the-scenes effort required to sustain this splendid initiative. I am proud to be part of this ecosystem and excited to continue contributing to a more open, reproducible, and collaborative science !

Mentors

I would like to especially thank Drs Anna Letko and Anthony Herzig for their guidance and the insightful discussions that influenced the direction of this project. Genomic data analysis can quickly become a tangled mess, and your advice helped illuminate the path, allowing me to achieve far more meaningful results.

Institut et ressources techniques

Je tiens à remercier chaleureusement l'ensemble de l'IGDR. Faire fonctionner une unité de recherche de cette envergure est un défi quotidien qui ne pourrait être relevé sans l'engagement et la bonne volonté de chacun. Un grand merci à toute l'équipe administrative pour le suivi rigoureux de mes missions, qui n'ont pas été sans rebondissements ni complexités.

Je remercie également le CRB Cani-DNA, Nadine Botherel et tout particulièrement Richard Guyon, pour la gestion de l'ensemble des prélèvements liés à ce projet. Plus de 1 000 échantillons sanguins ont été collectés et traités avec soin, un travail rigoureux en amont indispensable à la qualité des analyses bio-informatiques réalisées.

Enfin, je remercie le cluster de calcul Genouest pour la mise à disposition gracieuse des ressources informatiques nécessaires à ce projet, ainsi que pour leur soutien efficace lors des différents échanges et dépannages.

Équipe « Génétique du Chien »

Je remercie chaleureusement toute l'équipe « Génétique du Chien » ainsi que toutes les personnes que j'ai eu la chance de côtoyer au sein de cette équipe. Merci d'avoir pris le temps de m'expliquer les enjeux et les subtilités de la génétique, du monde canin et de la recherche. Votre passion est contagieuse et c'est avec d'excellents souvenirs que je poursuivrai ma carrière professionnelle, je l'espère, en recherche.

Je remercie tout particulièrement les membres du bureau 015, Caroline, Édouard et Armel, ainsi que Pascale et Lauranne, pour les moments de détente bien mérités. La recherche est parfois semée d'embûches et son parcours loin d'être linéaire, mais l'ambiance bienveillante instaurée au rez-de-chaussée est chose rare et précieuse.

Enfin, bonne chance à toi, Caroline, camarade de bureau et de frustrations. Le chemin de la thèse est tortueux mais gratifiant et j'espère sincèrement que tu parviendras à t'y épanouir malgré les quelques difficultés relationnelles que tu rencontres de l'autre côté.

Famille

Enfin, je souhaite remercier mes proches, pour les années passées, votre soutien indéfectible et votre présence sans faille. Vous êtes source de force et de réconfort qui m'ont été nécessaires depuis maintenant 30 ans révolus.

Merci à toi Maëlle, d'être à mes côtés depuis maintenant plus de 8 ans. La vie est loin d'être un long fleuve tranquille et ces dernières années nous l'ont rappelé à plusieurs reprises. Malgré les obstacles et les épreuves, nous cheminons ensemble. La vie trouve tout son sens lorsque l'on peut la partager avec une personne qui compte profondément.

DÉFINITIONS ET ABRÉVIATIONS

Définitions

- **Abduction / Adduction** : Mouvement qui écarte / rapproche un membre ou un segment de membre de la ligne sagittale du corps.
- **Arthrose** : Arthropathie chronique caractérisée par la destruction primitive des cartilages articulaires, une ostéophytose marginale et souvent une ostéosclérose sous-chondrale ; la synoviale est souvent inflammatoire.
- **Bassin ou ceinture pelvienne** : Ensemble osseux constitué du sacrum, du coccyx ainsi que des deux os coxaux anciennement appelé os iliaques. Ces derniers résultent de la fusion, lors de l'adolescence chez l'humain, de trois os distincts : l'ischium en position postérieur, l'ilium en position supérieur et le pubis en position antérieur.
- **Coaptation et congruence** : La coaptation qualifie l'ajustement immédiat et parfait de deux organes d'un même individu formés séparément. Ce terme reflète par voie de conséquence la capacité de la capsule articulaire à retenir la tête fémorale en place. De son côté, la congruence caractérise la qualité d'une articulation ou d'une anastomose dont les deux parties s'adaptent parfaitement. Deux surfaces sont congruentes lorsqu'il y a un emboîtement parfait. Une articulation peut ainsi être coapte (i.e., les éléments de l'articulation sont l'un dans l'autre) et non congruente (i.e., les surfaces sont déformées et ne correspondent pas ou peu) et inversement.
- **Congénitale et héréditaire** : Une caractéristique ou condition congénitale fait référence à ce qui existe dès la naissance. Une maladie congénitale se développe pendant la période de gestation et est présente à la naissance. Un caractère ou condition héréditaire fait référence quant à eux à ce qui est transmis de génération en génération par des facteurs génétiques. Une maladie héréditaire est causée par des anomalies génétiques spécifiques qui sont transmises des parents aux enfants. Bien qu'il puisse y avoir un certain recouvrement, les conditions congénitales ne sont pas nécessairement héréditaires et inversement.
- **Direction anatomique** : Un point d'attention particulier est à porter sur la nomenclature des directions utilisées. En effet, celles-ci sont dépendantes du modèle animal considéré. Ainsi, la couverture crâniale chez l'humain correspond à la couverture dorsale chez le chien du fait du changement d'orientation du pelvis.

- **Dysplasie** : Terme générique désignant toutes lésions liées à une anomalie du développement d'un tissu, d'un organe ou d'une partie d'organe. Du préfixe grec $\delta\upsilon\sigma$ / *dys* : difficulté, qui indique un fonctionnement défectueux ou difficile et de la racine $\pi\lambda\acute{\alpha}\sigma\iota\varsigma$ / *plasis* : façonner, qui qualifie ce qui a trait à la formation d'un tissu ou au développement d'un organe. Anomalie qui s'est produite pendant le développement d'un tissu ou d'un organe, engendrant malformation ou déformation. Habituellement congénitales et souvent héréditaires. Les dysplasies survenant au cours du développement embryonnaire, mieux appelées les dysembryoplasies, aboutissant à des ensembles malformatifs et déformatifs fixes et définitifs (e.g., dysplasie fibreuse des os, dysplasie rénale, dysplasie ectodermique, etc.) (Académie Nationale de Médecine, 2025a).
- **Dysplasie Coxo-Fémorale** : Trouble du développement de l'articulation coxofémorale qui se traduit par une incongruence, une laxité, un défaut de contact entre les surfaces articulaires, une déformation progressive des pièces articulaires et des lésions secondaires d'arthrose (Académie Nationale de Médecine, 2025b).
- **Fémur** : Segment proximal du membre inférieur, comprenant la diaphyse (i.e., partie centrale du fémur) ainsi que deux épiphyses dont la tête fémorale. Le grand et le petit trochanter permettent l'insertion de muscles.
- **Hanche** : Segment proximal du membre inférieur unissant la cuisse au pelvis. Il est centré par l'articulation coxale. Le terme *Coxal* signifie « relatif à la hanche ».
- **Laxité** : Elasticité de certains tissus comme les muscles ou les tissus conjonctifs de la peau ou des articulations : tendons et ligaments. Il est nécessaire de distinguer la laxité radiologique de la laxité clinique ainsi que la laxité passive et active. La laxité passive est celle observée lors de la radiologie en sédation profonde, les muscles sont le plus relâché possible et la laxité mesurée est donc seulement due à la configuration osseuse de l'articulation. Lorsque l'individu est éveillé, la tonicité des muscles fessiers compense en partie la laxité articulaire. C'est cette laxité active qui entraîne les déformations et celle qui de fait est lié aux aspects cliniques. La compensation est dépendante des individus, ainsi chez le chien, le Bouledogue possède une hanche luxée et un cotyle plat. Cette race est donc chondrodysplasique, mais du fait de sa très forte musculature ne présente que très peu de signes cliniques de dysplasie.
- **Luxation** : Perte totale des rapports anatomiques normaux entre les extrémités d'une articulation ; par extension, déplacement d'un organe hors de sa loge ou de sa position normale.
- **Subluxation** : Déplacement articulaire sans perte complète des rapports entre les surfaces en contact.

Termes *MeSH*

Un reflet de la complexité étymologique de la dysplasie de la hanche se retrouve dans le système de métadonnées médicales en langue anglaise (i.e., termes *MeSH*). La luxation congénitale de la hanche y est présentée comme une sous-catégorie de la dysplasie développementale de la hanche.

Chez l'humain

ID D000082602 : Dysplasie développementale de hanche / *Developmental dysplasia of the hip*

“Dislocation of the hip joint from an abnormal femoral head to the acetabulum relationship. It is most often due to ligamentous laxity, abnormal positioning of the joint and various other developmental, congenital factors, and method of delivery (e.g., oligohydramnios). When dislocation is diagnosed in neonates it is referred to as congenital hip dysplasia (D006618).”

ID D006618 : Luxation congénitale de la hanche / *Hip dislocation, congenital*

“Congenital dislocation of the hip generally includes subluxation of the femoral head, acetabular dysplasia, and complete dislocation of the femoral head from the true acetabulum. This condition occurs in approximately 1 in 1 000 live births and is more common in females than in males.”

Arborescence des termes *MeSH*

- C05 Musculoskeletal Diseases
 - C05.550 Joint Diseases
 - C05.550.518 Joint Dislocations
 - C05.550.518.384 Hip Dislocation
 - C05.550.518.384.500 Developmental Dysplasia of the Hip
 - C05.660 Musculoskeletal Abnormalities
 - C05.660.297 Developmental Dysplasia of the Hip
 - C05.660.297.500 Hip Dislocation, Congenital
- C16 Congenital, Hereditary, and Neonatal Diseases and Abnormalities
 - C16.131 Congenital Abnormalities
 - C16.131.621 Musculoskeletal Abnormalities
 - C16.131.621.297 Developmental Dysplasia of the Hip
 - C16.131.621.297.500 Hip Dislocation, Congenital
 - C16.131.621.449 Hip Dislocation, Congenital

Chez le chien

ID D006619 : Dysplasie de la hanche chez le chien / *Hip dysplasia, canine*

“A hereditary disease of the hip joints in dogs. Signs of the disease may be evident any time after 4 weeks of age.”

Arborescence des termes MeSH

-C22 Animal Diseases

-C22.268 Dog Diseases

-C22.268.485 Hip Dysplasia, Canine

Acronymes et abréviations

TABLE 1 – Listes des abréviations utilisées et leur signification en Anglais et Français

Acronyme	Anglais	Français
BMD	<i>Bernese Mountain Dog</i>	Bouvier Bernois
BP	<i>Before Present</i>	Avant notre ère
bp	<i>Base pairs</i>	Pairs de bases
CC	<i>Cane Corso</i>	Cane Corso
CFA	<i>Canis Familiaris chromosome</i>	Chromosome Canin
CHD/DCF	<i>Canine Hip Dysplasia</i>	Dysplasie Coxo-Fémorale
CNV	<i>Copy number variation</i>	Variation du nombre de copie
CRB	<i>Biological Resource Center</i>	Centre de Ressource Biologique
DBVDC	<i>Dog Biomedical Variant Database Consortium</i>	
DDBX	<i>Dogue de Bordeaux</i>	Dogue de Bordeaux
DDH	<i>Developmental Dysplasia of the Hip</i>	Dysplasie Développementale de la Hanche
DNA/ADN	<i>DeoxyriboNucleic Acid</i>	Acide DésoxyriboNucléique
EDTA	<i>Ethylene Diamine Tetra Acetic acid</i>	Acide Ethylène Diamine Tétra Acétique
FCI		Fédération Cynologique Internationale
Gol	<i>Golden Retriever</i>	Golden Retriever
GRM	<i>Genomic Relationship Matrix</i>	Matrice de relation génomique

Acronyme	Anglais	Français
GWAS	<i>Genome Wide Association Study</i>	Étude d'association pangénomique
IBD	<i>Identity By Descent</i>	Identité par descendant
IBS	<i>Identity By State</i>	Identité par état
InDels	<i>Insertions and Deletions</i>	Insertions et Délétions
Kb	<i>Kilobase</i>	Kilobase
Lab	<i>Labrador Retriever</i>	Labrador Retriever
LabxGol	<i>Labrador retriever mixed Golden</i>	Croisé Labrador et Golden
LD	<i>Linkage Disequilibrium</i>	Déséquilibre de liaison
LncRNA	<i>Long non coding RNA</i>	ARN long non codant
LOD	<i>Logarithm Of Odds</i>	Logarithme des probabilités
LOF		Livre des Origines Français
LP-WGS	<i>Low-Pass Whole Genome Sequencing</i>	Séquençage à faible couverture du génome complet
LTR	<i>Long Terminal Repeat sequence</i>	Séquence terminale longue répétée
MAF	<i>Minor Allele Frequency</i>	Fréquence de l'allèle mineure
Mb	<i>Megabase</i>	Mégabase
MDS	<i>Multi Dimentional Scaling</i>	Positionnement multidimensionnel
Mo	<i>Megaoctet</i>	Mégaoctet
NGS	<i>Next-Generation Sequencing</i>	Nouvelle Génération de séquençage
PCA	<i>Principal Component Analysis</i>	Analyse en Composantes Principales
PCR	<i>Polymerase Chain Reaction</i>	Amplification en chaîne par polymérase
QTL	<i>Quantitative Traits Loci</i>	Locus de caractères quantitatifs
RNA/ARN	<i>RiboNucleic Acid</i>	Acide ribonucléique
SINE/LINE	<i>Short/Long Interspersed Nucleotide Elements</i>	Eléments nucléaires dispersés courts/longs
SNP	<i>Single-nucleotide polymorphism</i>	Polymorphisme mononucléotidique
STR	<i>Short Tandem Repeat</i>	Répétition en tandem de séquence courte
SV	<i>Structural variant</i>	Variant de structure

Acronyme	Anglais	Français
UTR	<i>Untranslated region</i>	Région non traduite
VUS	<i>Variants of Unknown/Uncertain Significance</i>	Variants de signification inconnue/incertaine
WES	<i>Whole exome sequencing</i>	Séquençage complet de l'exome
WGS	<i>Whole genome sequencing</i>	Séquençage du génome complet

Sommaire

Remerciements	3
Définitions et abréviations	7
I Introduction	23
I Le modèle canin	24
A Description.....	24
B Domestication	24
1 Connaissances archéologiques	24
2 Connaissances génétiques	25
a Premières études sur l'ADN mitochondrial	25
b Études pangénomiques.....	26
c Études sur le chromosome Y et les lignées mâles.....	26
d Diversité régionale et signes de métissages	27
3 Un processus coévolutif ou une décision humaine délibérée ?	28
a L'hypothèse de l'adoption interspécifique	28
b L'hypothèse du charognard commensal	29
4 Conclusion	30
C Genèse et construction des races canines.....	31
1 La domestication : du mutualisme à l'élevage intentionnel	31
a Premières races façonnées par la fonction et l'environnement	32
2 Émergence des races modernes et maladies génétiques.....	33
3 Conclusion	35
D Le génome canin.....	37
1 Carte de liaison génétique et carte d'hybrides d'irradiation	37
2 Assemblage du génome	38
3 Émergence de nouvelles approches en génomique canine.....	38
4 Caractéristiques du génome canin	39
5 Annotation du génome canin	41
6 Évolution et conservation génomique.....	43
7 Déséquilibre de liaison (LD) et structure des haplotypes.....	45

8	Conclusion	46
II	Approche <i>One Health</i>	47
A	Définition et principes clés.....	47
B	Justification scientifique	47
C	Pertinence du chien dans une approche <i>One Health</i>	48
III	Etude des maladies génétiques	50
A	Les maladies génétiques	50
B	Obtention des données génomiques	51
C	Approches classiques d'analyse génétique	52
1	Études de liaison génétique (<i>Genetic Linkage Analysis</i>)	52
2	Études d'association pangénomiques (GWAS)	54
a	Particularités de l'espèce canine pour les études d'association	55
3	Études gènes candidats (<i>Candidate Gene Studies</i>)	58
D	Études de validation fonctionnelle	59
1	Validation fonctionnelle <i>in silico</i>	59
2	Validation fonctionnelle expérimentale	60
E	Conclusion.....	61
IV	Avantages du modèle canin dans l'étude des maladies génétiques complexes	62
A	Aspects pratiques	62
B	Aspects génomiques.....	64
1	Prédispositions par races	64
a	Effets fondateurs et co-ségrégation d'allèles	64
b	Prévalence élevée des maladies génétiques.....	64
2	Mode de transmission simplifié ?	65
3	Développer un test génétique de risque	66
C	Maladies homologues entre humain et chien	67
D	Exemple d'une maladie « simple » : l'ichtyose du Golden Retriever	68
V	La Dysplasie Coxo-Fémorale, une maladie « complexe »	69
A	Description anatomique	69
B	Impact de la bipédie	70
C	Pathogénèse et physiopathologie de la dysplasie de la hanche	70
D	Similarités et différences entre la dysplasie de la hanche chez le chien et chez l'humain.....	73
E	Facteurs de risque environnementaux de la dysplasie de la hanche	74
F	Aspects génétiques de la dysplasie de la hanche	74

VI Science Ouverte et FAIREST	75
A L'Open Science	76
B Les principes FAIR	77
1 Findable	77
2 Accessible	77
3 Interoperable	77
4 Reusable	78
C L'extension FAIREST	78
1 Engagement & Ethique / Equitable	78
2 Scalable	79
3 Social	79
4 Tested & Trust	79
D Inconvénients	80
E Avantages	80
F Les communautés Bioconductor et nf-core	81
1 Bioconductor	81
2 nf-core	82
 II Avant-propos	 83
I Historique du projet	84
II Diagnostic de la DCF	86
III Collaborations extérieures	88
IV Hypothèses et objectifs	90
V Travaux réalisés	92
A Collecte de prélèvements et des informations associées	92
B Comparaison de la dysplasie de la hanche chez le chien et l'humain	96
C Développement d'outils bio-informatiques	96
1 Préparation des données et constitution d'un fichier unique	96
2 Développement d'un package R Bioconductor de construction de pedigree : Pedixplorer	96
3 Développement d'un pipeline d'imputation : nf-core/phaseimpute	97
D Analyses d'association pangénomiques	98

III Résultats	103
I Étude bibliographique comparative	104
A La dysplasie de la hanche : une approche <i>One Health</i> entre humain et chien	106
1 Objectif	106
2 Description	106
3 Anatomie de la hanche chez l'humain et le chien	107
3.1 Développement d'une articulation synoviale	107
3.2 Adaptation, régénération et vascularisation	108
3.3 Composition et densité osseuse	108
3.4 Description de l'articulation coxo-fémorale chez l'humain	109
3.5 Description de l'articulation coxo-fémorale chez l'humain	111
4 Développement des dysplasies de la hanche	111
4.1 Hypothèses du développement anormal de la hanche	112
4.2 La dysplasie développementale la hanche chez l'humain	113
4.3 La dysplasie coxo-fémorale chez le chien	115
4.4 Arthrose	118
4.5 Conclusion	119
5 Diagnostic	120
5.1 Examen clinique	120
5.2 Imagerie	121
5.3 Scores composites	131
5.4 Biomarqueurs chez le chien	133
5.5 Tests génétiques chez le chien	134
5.6 Comparaison des mesures et scores en médecine vétérinaire	134
5.7 Différences et similarités entre humain et chien	135
6 Etiologie	138
6.1 Facteurs environnementaux	138
6.2 Facteurs hormonaux	140
6.3 Facteurs génétiques	141
7 Conclusion	148
8 Supplementary material	149
9 Bibliography	174
II Développement d'outils de bio-informatiques et de génétique computationnelle	199
A Développement d'un outil de concaténation de données hétérogènes	199
1 Introduction	199
2 Développement	199
a Phase 1 : Consolidation manuelle et macros <i>Excel</i>	200

b	Phase 2 : Automatisation par script et base de données	200
3	Données d'entrées et de sorties	202
4	Résultats	203
B	Développement d'un package <i>R Bioconductor</i> de construction de pedigree :	
	<i>Pedixplorer</i>	208
1	Présentation	208
2	<i>Pedixplorer</i> : a <i>Bioconductor</i> package to streamline pedigree design and visualization	214
1	Abstract	214
2	Introduction	215
3	Results	217
4	Discussion	220
5	Acknowledgements.....	221
6	Funding.....	221
7	Competing interests	221
8	Author contributions statement	221
9	Bibliography	222
10	Supplementary data.....	224
C	Développement d'un pipeline d'imputation :	
	<i>nf-core/phaseimpute</i>	236
1	<i>nf-core/phaseimpute</i> : A pipeline for genetic imputation	240
1	Abstract	240
2	Introduction	241
3	Materials and Methods	243
4	Results	249
5	Discussion	255
6	Availability and implementation.....	256
7	Acknowledgements.....	256
8	Funding.....	256
9	Competing interests	256
10	Bibliography	257
III	Étude génétique de la Dysplasie Coxo-Fémorale	263
A	Préparation des échantillons et séquençage <i>low-pass</i>	263
B	Alignement des séquences et imputation sur le génome de référence <i>canFam4</i>	264
C	Contrôles qualités des séquences LP-WGS obtenues	264
D	Analyses d'association pangénomiques (GWAS, <i>Genome-Wide Association Study</i>)	269
1	Première approche	270

2	Seconde approche	275
3	Identification of loci predisposing to canine hip dysplasia in four dog breeds using imputed genotypes from low-pass sequencing	280
1	Abstract	280
2	Introduction	280
3	Materials and Methods	282
4	Results	284
5	Discussion	291
6	Acknowledgements.....	292
7	Funding.....	292
8	Competing interests	293
9	Bibliography.....	293

IV Discussion & Conclusion 299

Discussion 300

A	Contraintes méthodologiques et logistiques	301
1	Recrutement des individus.....	301
2	Qualité et standardisation du phénotype	302
B	Validation du choix des races canines étudiées	303
C	Limites de l'étude et interprétation des résultats	305
1	Comparabilité inter espèces	305
2	Analyses génétiques	307
a	Données de génotypage	307
b	Analyses de liaison génétique	308
c	Analyses d'Association génétique : GWAS	309
D	Science ouverte et développement d'outils de génétique computationnelle....	312
1	Science ouverte	312
2	Démarche et communautés	313
3	Développement et essaimage des outils	314

Conclusion 316

Bibliography 319

Liste des figures

1	Schéma de domestication du chien et conséquences génomiques de la création des races	36
2	Photos des différents chiens dont l'ADN a été séquencé et utilisé comme génome de référence.....	39
3	Évolutions des annotations d' <i>Ensembl</i> de 2010 à 2025 pour les différentes versions d'assemblages du génome canin.....	42
4	Synténie (i.e., régions chromosomiques orthologues) entre le génome du chien (<i>canFam4</i>) et le génome humain (<i>hg38</i>).	44
5	Concept « Une Seule Santé » (<i>One Health</i>)	49
6	Stratégie de cartographie des maladies en deux étapes	57
7	Représentation anatomique d'une hanche humaine normale et d'une hanche atteinte de dysplasie de la hanche.	72
8	Logos des communautés <i>Bioconductor</i> et <i>nf-core</i>	82
9	Races collectées au sein du projet génétique Dysplasie Coxo-Fémorale ..	87
10	Déroulement du projet de thèse	93
11	Protocole de prélèvement sanguin transmis aux différents collaborateurs et accompagnants chaque kit de prélèvement.	94
12	Plaquette de communication destinée aux éleveurs et propriétaires de chiens atteints de DCF.....	95
13	Diagramme fonctionnel d'agrégation des données par le biais d'un script <i>Python</i> et de la base <i>NoSQL MongoDB</i>	201
14	Exemple de code utilisant le package <i>Pedexplorer</i>	211
15	Arbre généalogique obtenu en utilisant le package <i>Pedexplorer</i>	212
16	Logo du pipeline <i>nf-core/phaseimpute</i>	237
17	Proportion de chaque nucléotide pour chaque position le long de la séquence lue	265

18	Logigramme décisionnel de la sélection des individus à partir des échantillons du CRB Cani-DNA et de la société <i>Antagene</i> et des variants dans le cadre du projet DCF.	266
19	Graphique MDS (<i>Multi-Dimensional Scaling</i>) de l'ensemble des échantillons par race et statut dysplasique comme défini par la Fédération Cynologique Internationale (FCI) pour les dimensions 1 et 2 et un ensemble de variants indépendants obtenus via le logiciel <i>plink</i> v1.9.	267
20	Dendrogramme circulaire de proximité génomique entre les individus des différentes races (cercle intérieur) et des statuts cliniques (cercle extérieur)	268
21	Évolution de l'inflation génomique et du nombre de SNPs significatifs en fonction du modèle statistique utilisé pour chaque race et combinaison de races.	272
22	Représentation des différents loci identifiés par race et combinaison de races par région chromosomique pour les GWAS retenus parmi tous les modèles statistiques.	273
23	Représentation schématique des loci impliqués dans la DCF sous forme de caryotype canin pour la littérature humaine (HS) et canine (CF).	276
24	Ensemble des publications de la littérature utilisées pour représenter les régions chromosomiques (loci) impliquées dans la dysplasie de la hanche chez le chien et l'humain.	277
25	Cladogramme de 161 races canines, basé sur des données de séquençage tout génome (WGS), d'après (Parker et al., 2017).	304
26	Comparaison des différents os du squelette chez l'humain et le chien	306
27	Logos des trois outils développés et publiés lors de cette thèse	314

Liste des tableaux

1	Listes des abréviations utilisées et leur signification en Anglais et Français	10
2	Caractéristique de l'assemblage du génome canin (<i>canFam4</i>)	40
3	Bilan global des prélèvements disponibles en mai 2024 dans les races d'intérêt	205
4	Loci génomiques associés à la Dysplasie Coxo-Fémorale chez le chien en fonction des races.	274

Première partie

INTRODUCTION



LE MODÈLE CANIN

A Description

Le chien est un mammifère de l'ordre des carnivores appartenant à la famille des canidés regroupant 37 espèces en 13 genres. Le genre *Canis* contient la plupart des loups (i.e., doré, gris, d'Abyssinie), les coyotes, chacals et les dholes, présentant une répartition géographique distincte (Etats-Unis, Afrique, Asie du sud-est). Longtemps considéré comme une espèce à part entière du genre *Canis*, le chien est désormais classifié comme une sous-espèce du loup gris sous le nom de *Canis Lupus Familiaris*. Le chien et le loup gris sont en effet tous les deux descendants d'une même espèce désormais éteinte (Freedman et al., 2014). Tous les membres du genre *Canis* possèdent 78 chromosomes dont 38 paires d'autosomes acrocentriques et deux chromosomes sexuels métacentriques (Wurster-Hill & Centerwall, 1982). Les différentes espèces et sous-espèce du genre *Canis* sont théoriquement interfécondes (Vilà et al., 1999).

B Domestication

1 Connaissances archéologiques

Les premières traces archéologiques montrent que les humains et les loups partageaient déjà les mêmes territoires il y a environ 300 000 ans avant notre ère (BP), comme en attestent des ossements retrouvés en association étroite sur des sites tels que Zhoukoudian (300 000 BP, Chine) ou encore Lazaret (150 000 BP, France). Toutefois, cette proximité géographique n'atteste pas encore d'une domestication effective (Serpell, 1995). Les premiers éléments morphologiques attribuées à une domestication canine apparaissent entre 40 000 et 35 000 BP, notamment avec un crâne de chien paléolithique découvert à Hohle Fels, en Allemagne (Camarós et al., 2016). Des spécimens similaires ont été retrouvés dans la grotte de Goyet en Belgique (~36 500 BP) (Germonpré et al., 2009), à Razboinichya en Russie (~33 500 BP) (Ovodov et al., 2011) et à Předmostí en République tchèque (~31 000 BP) (Germonpré et al., 2012), ce qui suggère une possible domestication précoce dès la période aurignacienne (i.e., 43 000 à 33 000 BP). Cependant, l'identification de ces restes comme « chiens paléolithiques » fait toujours débat (Perri, 2016).

En effet, durant la transition loup-chien du Pléistocène, les marqueurs morphologiques sont subtils et se chevauchent : la forme du crâne et les dents (encombrement ou

réduction de la taille) peuvent indiquer une domestication, mais ces traits ne se fixent pas immédiatement et peuvent se produire dans une population locale de loups indépendante de l'humain. Les analyses génétiques des « chiens paléolithiques » montrent souvent qu'ils appartenaient à des lignées de loups divergentes et non à des clades de chiens déjà domestiqués. Ainsi, les données antérieures à $\approx 17\,000$ BP sont fragmentaires et demeurent ambiguës pour attester de restes canins (Perri et al., 2021). Ainsi certains chercheurs estiment qu'il s'agirait plutôt de loups apprivoisés (Serpell, 2021), attirés par les camps humains, dans un processus qualifié de proto-domestication, qui ne correspondrait pas encore à un processus de domestication dirigée ou intentionnel (Galibert et al., 2011).

Des spécimens plus récents, datés entre 17 000 et 13 500 ans BP, ont été retrouvés en France (Pionnier-Capitan et al., 2011 ; Morey, 2014), en Espagne (Vigne, 2005 ; Hervella et al., 2022) et en Allemagne (Street et al., 2014). Ils constituent à ce jour les premiers sites archéologiques attestant clairement de la présence de chiens domestiqués, sur la base de critères morphologiques et grâce aux apports conjoints de la géoarchéologie et de la taphonomie. Ces données fournissent des informations cruciales sur la formation, la préservation et l'interprétation et permettent ainsi la distinction entre individus sauvages et domestiques. Le site archéologique de Bonn-Oberkassel en Allemagne, datant d'environ 14 000 BP, est un exemple fascinant de la relation précoce entre les humains et les chiens. La découverte d'un individu canin inhumé aux côtés de restes humains offre des indices précieux sur les liens sociaux et culturels qui existaient déjà à cette époque (Benecke, 1987 ; Street et al., 2014). Cette période de la fin du Paléolithique coïncide avec une transformation des techniques de chasse, marquée notamment par l'introduction des armes à projectiles, ayant pu être favorisée par une collaboration plus étroite entre l'humain et le chien (Shipman, 2015).

2 Connaissances génétiques

a Premières études sur l'ADN mitochondrial

Les premières recherches génétiques sur la domestication du chien, dès les années 1995, se sont concentrées sur l'ADN mitochondrial, transmis uniquement par la lignée maternelle et donc utile pour retracer l'origine des femelles fondatrices. Une étude pionnière menée par Vilà et al. en 1997 a analysé l'ADN mitochondrial de 140 chiens domestiques et 162 loups gris. Leurs résultats ont confirmé que les chiens descendent bien du loup gris et ont identifié quatre clades distincts chez les chiens, sans correspondance avec les races modernes. L'un de ces clades semblait avoir divergé il y a plus de 100 000 ans, bien que cette estimation ait ensuite été considérée comme surestimée. L'étude suggérait également des cas d'hybridation entre chiens et loups après la domestication (Vilà et al., 1997).

Une étude plus large menée par Savolainen et al., portant sur 654 chiens et 38 loups, a proposé une origine des chiens domestiques en Asie du Sud-Est, à partir d'au moins cinq lignées femelles de loups, il y a environ 15 000 ans (Savolainen et al., 2002). Pang et al. en 2009 ont renforcé cette hypothèse en identifiant dix haplogroupes mitochondriaux majeurs, tous sont présents en Asie du Sud-Est. Selon leurs analyses, la domestication se serait déroulée entre 5 400 et 16 300 BP (Pang et al., 2009). Cependant, une étude menée par Thalmann et al. en 2013, basée sur l'ADN mitochondrial de canidés préhistoriques, a proposé une origine bien plus ancienne en Europe, estimée entre 18 800 et 32 100 BP (Thalmann et al., 2013).

b Études pangénomiques

L'évolution des technologies d'analyses génétiques a permis, dès les années 2000, d'élargir les analyses au-delà de l'ADN mitochondrial. Wang et al. ont ainsi utilisé des données pangénomiques pour estimer la date de la domestication du chien à environ 33 000 ans. Leurs résultats soutiennent une origine en Asie de l'Est, plus précisément dans le sud de la Chine, avec des migrations ultérieures vers l'ouest durant le Néolithique (Wang et al., 2013 ; Wang et al., 2016).

Cependant, Larson et al. ont remis en question l'idée selon laquelle les races canines dites « anciennes » peuvent être utilisées comme témoins de la domestication initiale. Ils ont démontré que la spécificité génétique de ces races reflète souvent un isolat géographique et culturel récent, plutôt qu'un héritage ancien (Larson et al., 2012).

c Études sur le chromosome Y et les lignées mâles

Alors que l'ADN mitochondrial permet uniquement de retracer les lignées maternelles, l'analyse du chromosome Y offre un aperçu complémentaire en suivant les lignées paternelles. Brown et al. en 2011 ont combiné des marqueurs mitochondriaux et du chromosome Y dans une étude portant sur environ 600 chiens modernes. Leurs résultats montrent que les chiens de village d'Asie du Sud-Est présentent une plus grande diversité génétique que ceux du Moyen-Orient, renforçant l'hypothèse d'une domestication précoce dans cette région (Brown et al., 2011). Ding et al. ont par la suite corroboré cette origine en séquençant le chromosome Y de 151 chiens, 12 loups et 2 coyotes (Ding et al., 2012).

Toutefois, Sacks et al. proposent une interprétation plus nuancée. Au lieu d'une origine unique, les chiens auraient connu selon eux une diversification importante en Asie du Sud-Est avant de s'étendre massivement vers l'ouest au cours du Néolithique, remplaçant partiellement les lignées locales (Sacks et al., 2013).

d Diversité régionale et signes de métissages

Fait intéressant, de nombreuses races de chiens modernes européennes et américaines se révèlent génétiquement plus proches des chiens de village d'Asie du Sud-Est que de ceux présents localement, ce qui suggère une influence génétique répandue des lignées d'Asie du Sud-Est. Ainsi Shannon et al. ont remarqué que certaines populations, notamment au Vietnam, en Inde et en Égypte, présentent peu de traces d'hybridation européenne et présentent des caractéristiques génétiques compatibles avec une origine en Asie centrale. D'autres, en revanche, montrent une influence européenne plus marquée (Shannon et al., 2015). Frantz et al. semblent confirmer cette distinction en mettant en évidence une profonde divergence génétique entre les chiens d'Asie de l'Est et ceux d'Eurasie occidentale, apparue entre $\approx 14\,000$ et $6\,400$ BP. Cette vaste étude suggère que les chiens ont été domestiqués indépendamment en deux régions distinctes de l'Eurasie : l'Est et l'Ouest. Cette conclusion est basée sur l'analyse génétique de 59 anciens chiens et du génome complet d'un chien du Néolithique tardif (daté d'environ $4\,800$ BP) trouvé en Irlande. Cette scission pourrait refléter des événements de domestication indépendante ou un remplacement partiel des populations européennes (Frantz et al., 2016). À l'inverse, Botigué et al. ont observé une continuité génétique chez les chiens néolithiques européens et ont resserré la fenêtre temporelle de domestication entre $20\,000$ et $40\,000$ BP, remettant en cause l'hypothèse d'une double origine (Botigué et al., 2017).

Skoglund et al., de leur côté, ont montré que certaines races arctiques modernes contiennent une part d'héritage génétique provenant d'une ancienne lignée de loups de Sibérie, indiquant une introgression (i.e., transfert de matériel génétique entre espèces par hybridation suivie de croisements répétés) à partir de populations régionales distinctes (Skoglund et al., 2015). Cela semble être confirmé par Perri et al. en 2021, qui suggèrent que la domestication du chien s'est produite en Sibérie, il y a environ $23\,000$ BP (Perri et al., 2021) ainsi que par Bergström et al. en 2022. Les travaux de Bergström et al 2022 sont issues d'une analyse génomique exhaustive de 72 génomes de loups anciens, couvrant une période de $100\,000$ BP avec des échantillons provenant de diverses régions géographiques, notamment l'Europe, la Sibérie et l'Amérique du Nord. Cette étude montre que les chiens domestiques modernes descendent d'au moins deux populations distinctes de loups. Cette dualité dans l'ascendance génétique des chiens suggère que la domestication des chiens n'a pas été un événement unique et localisé, mais un processus complexe impliquant plusieurs populations de loups à différents endroits et à différentes époques (Vonholdt et al., 2010 ; Bergström et al., 2022).

3 Un processus coévolutif ou une décision humaine délibérée ?

La domestication est un processus évolutif à l'échelle de la population, au cours duquel les humains influencent activement la reproduction et l'environnement d'une espèce animale, modifiant ainsi ses caractéristiques génétiques, physiologiques et comportementales afin de répondre à leurs besoins (Purugganan, 2022 ; Tancredi & Cardinali, 2023). Contrairement au simple apprivoisement, qui consiste à habituer un animal sauvage individuel à la présence humaine, la véritable domestication implique des changements héréditaires transmis sur plusieurs générations. En d'autres termes, ce n'est pas le comportement amical d'un seul individu qui suffit, mais un processus de sélection reproductrice à long terme et transmise génétiquement.

Deux scénarios principaux de domestication coexistent actuellement dans la littérature scientifique. Le premier, proposé dès 1865 par Francis Galton, postule une adoption interspécifique de louveteaux par les premiers chasseurs-cueilleurs (Galton, 1865). Le second envisage la domestication du chien non comme une invention délibérée de l'être humain, mais comme un sous-produit émergeant des interactions prolongées entre humains et loups.

a L'hypothèse de l'adoption interspécifique

Le premier scénario de domestication met l'accent sur le rôle actif des humains dans ce processus. Contrairement à l'hypothèse de cohabitation opportuniste, cette théorie décrit les chasseurs-cueilleurs paléolithiques comme les principaux agents de la domestication, enlevant délibérément des louveteaux de leur tanière pour les élever comme des dépendants au sein du groupe humain. Si la plupart de ces animaux auraient probablement repris une vie sauvage à l'âge adulte, une minorité présentant des traits socialement favorables (e.g., docilité, capacité d'apprentissage, sociabilité) aurait bénéficié d'un traitement privilégié, augmentant ainsi ses chances de transmettre ces caractéristiques à une descendance semi-domestique (Galton, 1865).

Cependant, cette hypothèse soulève plusieurs objections. La domestication semble s'être produite de manière très localisée, alors même que l'adoption de jeunes animaux aurait pu, en principe, être plus largement répandue (Downs, 1960). Pour répondre à cette contradiction, certains auteurs avancent que le comportement d'adoption interspécifique n'était lui-même ni universel ni spontané, mais probablement limité à certaines cultures humaines spécifiques du Paléolithique (Shipman, 2010 ; Serpell, 2021).

De plus l'élevage de loups nécessite un apport important en protéines animales, disponible uniquement dans certains environnements favorables, comme ceux du dernier maximum glaciaire (21 000 BP) (Lahtinen et al., 2021). Enfin, la faisabilité technique de l'apprivoisement est elle aussi remise en question. Les loups possèdent en effet une fenêtre de socialisation étroite et même apprivoisés, ils restent souvent difficiles à gérer

à l'âge adulte (Woolpy & Ginsburg, 1967). Toutefois, des données ethnographiques suggèrent que des humains ont su contourner ces limites par des pratiques intensives de soins, allant jusqu'à l'allaitement (Simoons & Baldwin, 1982), appuyées par une tendance humaine naturelle à l'allo-parentalité (Serpell & Paul, 2011).

b L'hypothèse du charognard commensal

En contraste avec cette hypothèse de domestication intentionnelle, un second scénario propose une explication plus passive et graduelle du processus. Selon cette théorie, la domestication résulterait non pas d'un projet humain délibéré, mais d'une réponse évolutive à une niche écologique créée involontairement par les humains en mettant à disposition des déchets autour des campements (Larson et al., 2012). La domestication du chien aurait ainsi été une forme de mutualisme accidentel (Tancredi & Cardinali, 2023). Sans véritable intention, les humains auraient commencé à accorder un traitement préférentiel aux individus présentant certains traits utiles, amorçant ainsi un processus de sélection inconsciente (Coppinger & Coppinger, 2001).

Ce modèle suppose néanmoins que les chasseurs-cueilleurs paléolithiques généraient une quantité suffisante de restes alimentaires pour entretenir une population de loups charognards (Jung & Pörtl, 2018). Un tel contexte paraît peu probable autour de 15 000 ans avant notre ère, période supposée de la domestication, car les humains vivaient alors en petits groupes mobiles. Toutefois, certains sites d'Europe centrale et orientale présentent d'imposantes accumulations de restes de mammouths datées entre 40 000 et 24 000 BP, ce qui pourrait localement valider cette hypothèse (Serpell, 2021).

En revanche, certaines pratiques culturelles documentées chez des groupes de chasseurs-cueilleurs actuels plaident contre la disponibilité de tels déchets : de nombreux peuples appliquent des rituels stricts visant à éviter que les restes d'animaux ne soient souillés par des charognards (Reo & Whyte, 2012).

Enfin, la présence de prédateurs semi-habitués mais non apprivoisés constitue un risque non négligeable. Ces animaux peuvent se montrer imprévisibles, audacieux et parfois agressifs, en particulier envers les individus vulnérables comme les enfants (Moriceau, 2007). Des témoignages historiques et contemporains rapportent des attaques parfois mortelles de canidés sauvages nourris ou habitués par l'humain (Schmidt & Timm, 2007).

4 Conclusion

Les contraintes évoquées montrent que ni l'adoption interspécifique ni la commensalité ne suffisent seules à expliquer la domestication du loup. Un scénario mixte apparaît plus plausible : la cohabitation opportuniste aurait ainsi favorisé des interactions régulières, rendant possible, dans certains contextes culturels et écologiques favorables, l'adoption volontaire de louveteaux. Ce processus rare aurait permis à certains groupes du Paléolithique de coexister avec des loups sur plusieurs générations, jusqu'à enclencher une sélection comportementale durable et irréversible (Serpell, 2021).

Ce cadre théorique trouve un écho dans les données archéologiques et génétiques disponibles. La domestication du chien apparaît comme un processus plurimillénaire et multirégional, oscillant entre proto-domestication et sélections successives, plutôt que comme un événement unique et ponctuel. Les données archéologiques, de la simple cohabitation humain-loup ($\approx 400\,000$ BP) aux inhumations conjointes du Paléolithique supérieur ($\approx 14\,000$ BP), témoignent d'une relation progressive. Les analyses génétiques, qu'il s'agisse de l'ADN mitochondrial, du chromosome Y, des panels de marqueurs génétiques SNP ou des génomes anciens, confirment l'ascendance de tous les chiens modernes d'un ancêtre commun avec le loup gris désormais éteint (Bergström et al., 2022).

Les données génétiques couplées aux indices archéologiques décrivent une séquence d'événements s'étant produits en différents endroits (Asie, Europe, Sibérie). Les loups auraient ainsi investi les niches humaines en Eurasie au cours du Pléistocène supérieur, probablement de manière plus marquée durant le Dernier Maximum Glaciaire dans certaines zones refuges et y auraient subi une première domestication entre 25 000 et 20 000 ans avant le présent, avant de se diffuser globalement en suivant les migrations humaines (Perri et al., 2021). Dès l'Holocène (i.e., 12 000 BP à aujourd'hui), les chiens domestiqués se seraient alors diversifiés en de nombreuses populations régionales et plus tard en races, chacune adaptée aux environnements et aux cultures humaines locales.

C Genèse et construction des races canines

1 La domestication : du mutualisme à l'élevage intentionnel

La domestication du chien fut un processus complexe en plusieurs étapes, étalé sur des millénaires et combina sélection naturelle et sélection opérée par l'humain de manière consciente ou non. La figure 1 schématise l'ensemble de ce processus. Les chiens du paléolithique co-évoluèrent tout d'abord avec les chasseurs-cueilleurs en tant que synanthropes (i.e., espèces non domestiquées bénéficiant d'un environnement anthropisé). Dans un second temps, la domestication consista certainement en la réduction de la peur et de l'anxiété, puis tendit à renforcer les capacités sociales et la coopération (Tancredi & Cardinali, 2023).

Le génome canin reflète les goulets d'étranglement génétiques et les expansions démographiques qui ont accompagné les migrations humaines, selon le principe bien résumé par l'adage "*where people went, dogs went*" (Ollivier et al., 2018 ; Perri et al., 2021). Ces goulets d'étranglement, survenus lors de la domestication puis des colonisations successives, correspondent à des périodes où un nombre limité d'individus ont contribué à la génération suivante. Ce rétrécissement drastique de la diversité génétique favorise la dérive génétique, qui peut entraîner la fixation aléatoire d'allèles neutres, mais aussi d'allèles délétères (Marsden et al., 2016 ; Bergström et al., 2020). Ainsi, lorsque de petits groupes de chasseurs-cueilleurs s'installèrent dans de nouveaux territoires, ils emmenèrent leurs chiens, provoquant des effets fondateurs profonds, dont on observe encore aujourd'hui les conséquences dans le patrimoine génétique des races actuelles. De nombreuses caractéristiques associées à la domestication, comme la réduction de l'agressivité ou le maintien de traits juvéniles, ne résultèrent pas nécessairement d'une sélection intentionnelle, mais émergèrent probablement comme des effets secondaires d'un relâchement des pressions de sélection naturelle (Dobney & Larson, 2006). L'influence humaine, souvent exercée de manière inconsciente ou par sélection négative (e.g., en éliminant les individus trop agressifs), a également dû jouer un rôle déterminant dans le développement des aptitudes cognitives et comportementales des chiens, facilitant ainsi leur intégration durable dans les sociétés humaines.

L'expérience soviétique de domestication du renard, menée par Belyaev, démontra de manière spectaculaire que la sélection comportementale, en l'occurrence la docilité, peut entraîner une cascade de modifications physiques et physiologiques appelé « syndrome de domestication ». En quelques générations seulement, des renards élevés pour leur moindre peur de l'humain développèrent des oreilles tombantes, un pelage tacheté, un museau raccourci et une réponse au stress modifiée, notamment avec des taux de cortisol nettement réduits (Trut, 1999 ; Trut et al., 2009). Ces transformations, observées chez de nombreux mammifères domestiqués, résultent vraisemblablement

de modifications du développement des crêtes neurales (Wilkins et al., 2014). Le système de réponse au stress des renards évolua ainsi au niveau de l'axe hypothalamo-hypophyso-surrénalien et des voies de neurotransmission, peut-être médié par des modifications épigénétiques et contribua à la diminution de la peur et de l'agressivité (Sajjanar et al., 2025).

Au-delà de l'apparence et du tempérament (arrivée de pelage blanc, enroulement de la queue, docilité, etc.), la domestication impacta d'autres fonctions biologiques clés. La reproduction saisonnière observée chez le loup devint plus fréquente et indépendante des saisons, avec l'apparition de deux *œstrus* chez les femelles (Lord et al., 2013). Sur le plan métabolique, les chiens s'adaptèrent aux régimes riches en amidon propres aux habitats humains. Les analyses génomiques montrent ainsi que les chiens ont acquis des copies supplémentaires du gène *AMY2B*, ce qui renforce la production d'amylase pancréatique pour une digestion efficace de l'amidon. Des gènes tels que *MGAM* et *SGLT1*, impliqués dans le métabolisme des glucides, présentent des signes de sélection positive, témoignant d'une adaptation poussée aux apports glucidiques (Arendt et al., 2016 ; Ollivier et al., 2016).

Ces adaptations physiologiques et métaboliques illustrent l'impact majeur de la domestication, qui permis aux chiens de s'épanouir aux côtés des humains dans des milieux très variés. L'humain orienta ensuite progressivement la sélection canine afin de favoriser des caractéristiques précises (Larson & Fuller, 2014).

a Premières races façonnées par la fonction et l'environnement

Avant l'apparition de standards de race officiels, la majorité des chiens domestiques existaient sous forme de populations locales semi-libres, souvent désignées aujourd'hui sous le terme de « chiens de village » (*village dogs*). Il ne s'agissait pas de races à proprement parler, mais de populations non sélectionnées de manière dirigée. Leurs caractéristiques morphologiques et comportementales résultaient alors d'un équilibre entre pression de sélection naturelle, pression humaine indirecte et adaptation aux contraintes locales. Ces populations conservaient une diversité génétique importante, résultant essentiellement des événements migratoires et de l'adaptation locale. Nombre d'entre elles peuvent retracer leur lignée jusqu'à des populations anciennes ayant subi peu de croisements avec d'autres populations (Boyko et al., 2009).

Durant la plupart de leur cohabitation avec les humains, les chiens furent élevés avant tout pour remplir une fonction utilitaire : chasse, garde, pastoralisme, pistage, attelage et autres tâches vitales pour la survie et les activités économiques. Dans chaque région, les éleveurs privilégièrent des traits comportementaux et de performance répondant à leur besoin, l'instinct de troupeau des chiens de berger ou bien la vigilance des chiens de garde, mais aussi les individus robustes, adaptés aux conditions climatiques locales (Worboys et al., 2018 ; Menor-Campos, 2024). Les caractéristiques géographiques

(e.g., montagnes, îles, étendues désertiques ou zones pastorales) renforcèrent cette spécialisation régionale, chaque communauté humaine façonnant son compagnon canin pour répondre aux défis environnementaux et à ses besoins particuliers (Lord et al., 2016).

Au cours de cette phase initiale, l'apparence fût certainement un sous-produit et non un objectif de la sélection. Les caractéristiques physiques (e.g., couleur du pelage, forme des oreilles, enroulement de la queue, etc.) émergèrent de manière fortuite, soit sous l'effet des contraintes fonctionnelles, soit par dérive génétique au sein de populations isolées et non par choix esthétiques délibérés (Lord et al., 2016). Ces premières races utilitaires constituèrent finalement le capital génétique à l'origine des races officiellement reconnues par la suite.

2 Émergence des races modernes et maladies génétiques

Le relâchement de la pression sélective, conjugué aux épisodes répétés de goulets d'étranglement génétiques et à la réduction de la taille effective des populations conduisit à une augmentation de la prévalence d'allèles non synonymes tant au niveau du génome nucléaire que mitochondrial (Björnerfeldt et al., 2006 ; Cruz et al., 2008). Cette diversité génétique constitua ainsi la matière première sur laquelle la sélection artificielle put façonner les races modernes.

À mesure que les sociétés humaines évoluèrent, leurs besoins se diversifièrent. Les races de chiens évoluèrent elles aussi vers d'autres tâches spécialisées, comme le gardiennage, la chasse ou le pastoralisme, mais aussi comme compagnons familiaux et ornementaux (Zhang et al., 2020). Ce tournant s'opéra particulièrement à l'époque victorienne avec la création des premiers clubs canins et concours où des standards de race furent définis pour valoriser des critères esthétiques tels que la couleur du pelage, la forme des oreilles ou la taille (Worboys et al., 2018). Cette période fut marquée d'une diversification phénotypique sans précédent : près de 400 races distinctes furent créées pour répondre aux goûts humains de nouveauté et de prestige plutôt qu'à des besoins fonctionnels (American Kennel Club, 1997 ; Ghirlanda et al., 2013). L'élevage canin passa ainsi d'une orientation utilitaire à une sélection de critères esthétiques et comportementaux (Morrill et al., 2022).

Afin de préserver l'apparence des races nouvellement codifiées, les clubs instaurèrent des livres généalogiques fermés, n'autorisant que la reproduction entre individus appartenant à la même race et enregistrés. Cette pratique a progressivement cloisonné les races dans des entités génétiquement isolées, réduisant leur diversité intrarace et empêchant le brassage génétique (Parker et al., 2004).

En mai 2025, la Fédération Cynologique Internationale (FCI) reconnaît officiellement 372 races réparties en 10 groupes différents (*FCI Breeds Nomenclature*, 2025).

Partant du goulet d'étranglement initial de la domestication, qui accrut la proportion d'hétérozygotie non synonyme chez le chien (Marsden et al., 2016), la formation des races imposa une série de goulets d'étranglement secondaires, spécifiques à chaque race. Ces événements démographiques successifs réduisirent encore davantage la diversité génétique intrarace, tout en accentuant la différenciation entre races. Les races anciennes, basées originellement sur leur fonction utilitaire, tel que les chiens de traîneaux (e.g., Husky, Malamute) ou les lévriers de chasse à vue (e.g., Afghan, d'Asie), furent ainsi progressivement affinés en morphotypes distincts par l'accumulation de modifications ciblées et, ponctuellement, par des croisements extérieurs (Parker et al., 2017). Les races canines se distinguèrent alors par leurs traits phénotypiques spécifiques, tels que la taille (ex. : Chihuahua vs Dogue Allemand), la texture et la couleur du pelage (ex. : Caniche, Dalmatien) ou encore des comportements particuliers comme la chasse à vue (Lévrier Afghan, Saluki) ou le gardiennage de troupeau (chiens de bergers) (Ostrander et al., 2017). Cette sélection spécifique à chaque race laissa une empreinte caractéristique sur le génome canin en remodelant profondément la structure génétique des populations canines. À l'échelle du génome, cela s'est traduit par un mosaïsme d'haplotypes de grande taille (Vaysse et al., 2011), entrecoupés de régions de forte hétérozygotie là où la pression de la sélection naturelle et sexuelle eu une pression moindre (Boyko, 2011).

L'essor rapide des races canines modernes, engendra non seulement une diversité morphologique, physiologique et comportementale spectaculaire, mais aussi des conséquences génétiques involontaires. L'augmentation des sites d'hétérozygotie non synonyme et la fixation de longs haplotypes favorisèrent l'accumulation de mutations délétères désormais ancrées dans de nombreuses races (Ostrander et al., 2017). Ce phénomène est distinct des effets génétiques associés à la domestication initiale du chien : il relève ici des pressions sélectives récentes imposées par l'élevage moderne, souvent orientées vers des critères esthétiques ou comportementaux exacerbés (Cruz et al., 2008 ; Tancredi & Cardinali, 2023). Marsden et al. en 2016 montrèrent en outre que les régions du génome soumises à une forte sélection artificielle concentrent une proportion accrue de gènes impliqués dans des maladies mendéliennes, illustrant clairement le compromis entre phénotype et santé (Marsden et al., 2016).

La fixation de mutations délétères entraîna l'émergence de nombreuses maladies génétiques dont la moitié serait spécifique à certaines races (Patterson, 2000 ; Nicholas & Tammen, 2025). Ces affections couvrent un large spectre : cancers, syndromes du développement, troubles comportementaux et affections psychiatriques (Menor-Campos, 2024). À titre d'exemple, la dysplasie de la hanche touche actuellement de nombreuses races de grandes tailles, avec une prévalence dépassant 50 % chez certaines (Genevois, 2020). Depuis plusieurs décennies, les chercheurs soulignent que cette sélection sur des critères principalement esthétiques se fait au détriment

du bien-être des individus (McGreevy & Nicholas, 1999 ; Cruz et al., 2008 ; Menor-Campos, 2024). Les races brachycéphales en sont une illustration frappante, souffrant de graves troubles respiratoires en raison de la conformation extrême de leur crâne.

3 Conclusion

La domestication du chien fut un processus complexe en deux phases, façonné à la fois par l'évolution biologique et par le contexte culturel et temporel. Au fil des générations, une sélection naturelle et anthropique favorisa les loups les moins peureux, les plus sociables et les mieux adaptés au comportement humain. Cette transformation progressive donna naissance au chien domestique, une lignée distincte de ses ancêtres sauvages aussi bien d'un point de vue génétique que comportemental. Lors des migrations humaines, les chiens suivirent, se diversifiant en populations locales fonctionnelles adaptées aux environnements et aux tâches locales (Perri et al., 2021). Cependant, au cours des 200 dernières années, le rythme et la nature de cette diversification canine changea radicalement. L'essor de l'élevage moderne, motivé principalement par des préférences esthétiques et d'aptitudes, conduisit à la création de centaines de races, plus récemment, sans réelles considérations pour la santé ou la diversité génétique.

Chaque race devint de plus en plus morphologiquement distincte conduisant à l'émergence de véritables isolats génétiques. Ce processus de spécialisation, souvent poussé à l'extrême dans certains cas (phénomène connu sous le nom d'hypertype), a été renforcé par des pratiques d'élevage strictes et l'usage généralisé de livres généalogiques fermés. Il en résulta une accumulation et une fixation d'allèles délétères, engendrant une augmentation notable de la prévalence de maladies héréditaires propres à certaines races. De ce fait, les races canines modernes peuvent être considérées comme de véritables isolats génétiques, aux profils uniques mais vulnérables sur le plan sanitaire.

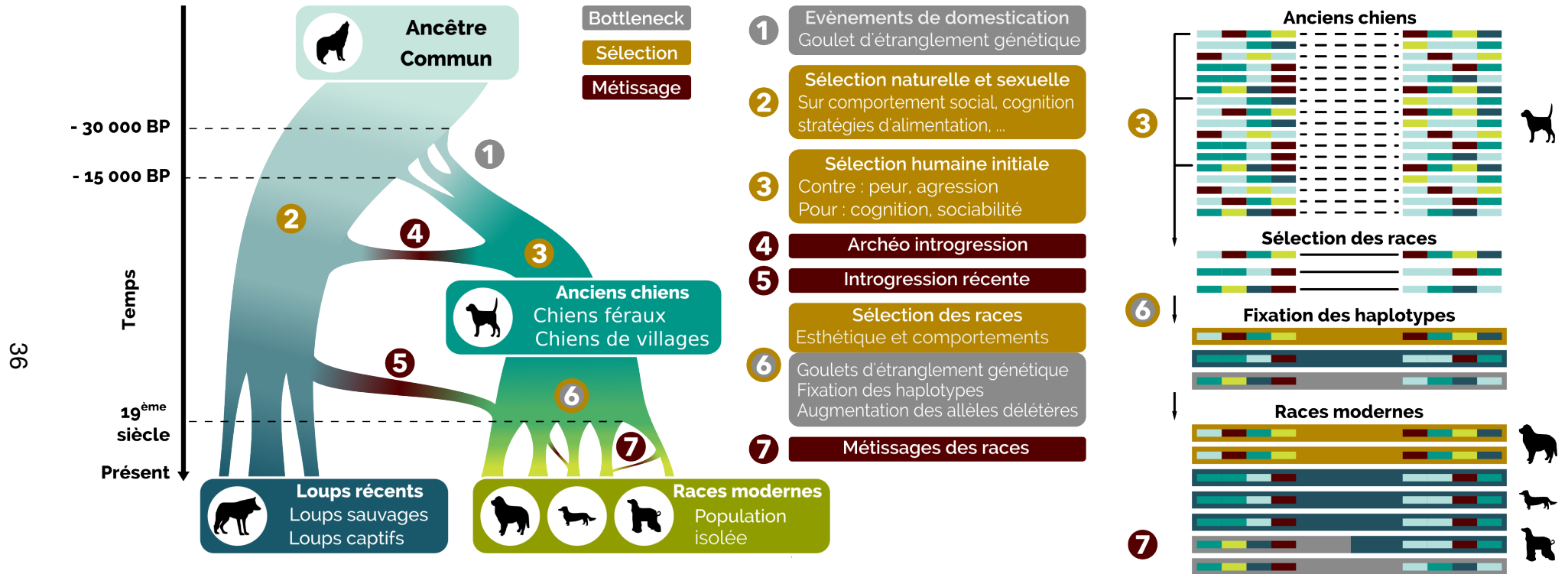


FIGURE 1 – Schéma de domestication du chien et conséquences génomiques de la création des races

À gauche sont représentées les principales étapes de la domestication du chien, ainsi que les différents goulets d'étranglement et événements d'introgression. À droite, le schéma illustre la fixation des haplotypes dans les races modernes, accompagnée d'une diminution de la diversité génétique au sein de chaque race. Schéma inspiré de (Lindblad-Toh et al., 2005 ; Range & Marshall-Pescini, 2022).

D Le génome canin

Le caryotype canin se compose de 38 autosomes acrocentriques et de 2 chromosomes sexuels, pour un génome totalisant environ 2,4 milliards de paires de bases. Cette organisation chromosomique a été décrite dès les années 1990 par les premières études de cytogénétique, puis progressivement enrichie par les efforts de cartographie génétique et de séquençage, notamment sous l'impulsion des équipes de E. A. Ostrander et F. Galibert, figures fondatrices de la génomique canine moderne. Bien que plus compact que celui de l'humain, le génome du chien présente une densité génique comparable. L'intérêt pour le chien en tant que modèle spontané en recherche biomédicale a émergé à la même époque, motivé par la forte similarité clinique de nombreuses affections humaines et canines et par la structure génétique particulière induite par la sélection artificielle. Dans ce contexte, l'équipe « Génétique du Chien » s'est engagée très tôt dans la construction de ressources génétiques de référence, bien avant l'ère du séquençage de nouvelle génération.

1 Carte de liaison génétique et carte d'hybrides d'irradiation

Avant l'ère du séquençage intégral, l'organisation du génome canin s'est appuyée sur la construction progressive de cartes de liaison génétique et de cartes d'hybrides d'irradiation (RH). Les premières cartes de liaison, élaborées à la fin des années 1990 à partir de pedigrees multigénérationnels et de marqueurs microsatellites (Mellersh & Ostrander, 1997 ; Neff et al., 1999), ont permis de regrouper les régions en groupes de liaison correspondant aux futurs chromosomes (Werner et al., 1999).

Parallèlement, des cartes d'hybrides d'irradiation ont été développées ont rapidement gagné en densité, intégrant des milliers de régions (Mellersh et al., 2000 ; Breen et al., 2001 ; Guyon et al., 2003 ; Breen et al., 2004). L'intégration de ces cartes RH avec des marqueurs de gènes connus, positionnés par l'hybridation *in situ* en fluorescence (FISH), a permis de constituer des cartes intégrées combinant génétique, cytogénétique et physique, facilitant ainsi la navigation dans le génome canin.

En parallèle, la génomique comparative, via des alignements de séquences entre gènes orthologues humains et canins, a révélé un degré élevé de synténie conservée entre les deux espèces (Hitte et al., 2005 ; Derrien et al., 2009) renforçant la valeur du chien comme modèle en génétique comparée.

Ces efforts combinés ont abouti à des cartes génétiques et RH intégrées de plus en plus précises, qui ont servi de base au positionnement des premiers contigs lors de l'assemblage des séquences du génome canin de référence *canFam1* (Lindblad-Toh et al., 2005). Des cartes complémentaires, fondées sur des panels de génotypage à haut débit, ont été développées par la suite pour raffiner encore la structure du génome (Wong et al., 2010 ; Auton et al., 2013).

2 Assemblage du génome

Avec l'essor des technologies de séquençage, les approches de cartographie ont progressivement cédé la place au séquençage direct du génome. Une première version du génome canin a été publiée avec une couverture faible (1×) au début des années 2000 (Kirkness et al., 2003). Ce travail a été suivi par un projet de séquençage complet à couverture intermédiaire (6×) d'un chien de race Boxer, une femelle nommée Tasha, publié par le Broad Institute sous le nom de *canFam1.0* et *canFam2.0* (Lindblad-Toh et al., 2005). Cet assemblage, fruit d'un large effort collaboratif soutenu par le NIH, a constitué la première référence canine largement diffusée.

Au fil des années, cet assemblage a été affiné à plusieurs reprises : *canFam2*, *canFam3*, puis *canFam3.1* (Hoepfner et al., 2014) et *canFam3.2* (Wucher et al., 2017). Il a récemment bénéficié d'un réassemblage en haute résolution basé sur les technologies à longues lectures (PacBio HiFi, 100×), produisant *canFam6* (Jagannathan et al., 2021). En parallèle, d'autres génomes de référence ont été produits pour refléter la diversité des races (voir figure 2) : *canFam4* (Berger Allemand ; Wang et al. (2021a)), *canFam5* (Dogue Allemand, Zoey ; Halo et al. (2021)), ainsi qu'un Labrador Retriever (Roslin Institute) et un Basenji (Université du Missouri). Plus récemment, plusieurs initiatives à grande échelle ont vu le jour : 722 génomes entier séquencés et analysés par Plassais et al. (2022), 2000 génomes de canidés séquencés à plus de 10× par le consortium international *Dog 10 000 genomes* (Meadows et al., 2023) et des centaines d'autres par DVDBC sous la coordination de T. Leeb (Université de Berne). Ainsi la constitution de génomes de référence avec des annotations de plus en plus fines et précises a permis de produire 12 génomes de 12 races d'intérêt majeur pour la recherche biomédicale mais aussi pour des projets de recherche s'intéressant au comportement canin ou à l'évolution des génomes.

canFam4 reste aujourd'hui l'assemblage de référence le plus utilisé dans les études génétiques canines, notamment en raison de la qualité de son annotation et de son intégration dans les grandes bases de données. Les principales caractéristiques de ce génome sont résumées dans le tableau 2.

3 Émergence de nouvelles approches en génomique canine

Actuellement, les avancées récentes des technologies génomiques incitent les chercheurs à produire un assemblage de référence complet de type télomère-à-télomère (T2T). L'utilisation des plateformes de séquençage longues lectures telles que *PacBio HiFi* et *Oxford Nanopore Technologie* (ONT) permettront de capturer l'intégralité de la structure chromosomique sans *gaps*, y compris les régions centromériques et télomériques qui restent encore aujourd'hui non résolues.



FIGURE 2 – Photos des différents chiens dont l'ADN a été séquencé et utilisé comme génome de référence

De gauche à droite : a) Tasha, un Boxer (canFam1, 2, 3.1 et 6); b) un Berger Allemand (canFam4); c) Zoey, un Dogue Allemand (canFam5); d) un Labrador Retriever du Roslin Institute; e) un Basenji de l'Université du Missouri. Photos extraites de <https://www.ensembl.org/info/about/species.html> .

Parallèlement, un changement de paradigme s'est amorcé avec le développement des pangénomes. Contrairement aux génomes de référence linéaires traditionnels, issus d'un seul individu, un pangénome intègre les variations génétiques de plusieurs individus dans une structure en graphe. Ce modèle reflète mieux la diversité populationnelle, permettant des analyses plus représentatives et réduisant le biais de référence du génome linéaire (Eizenga et al., 2020). Cependant, cette approche introduit de nouveaux défis et questions clés : combien d'individus faut-il intégrer au pangénome ? Quels variants doit-on conserver dans la structure en graphe ? Et comment intègre-t-on les pangénomes dans les outils existants pour l'alignement des séquences, la détection de variants ou les études d'association pangénomiques (GWAS) ?

À mesure que ces méthodologies mûrissent, ces assemblages promettent d'améliorer la précision de l'identification des variants et l'analyse structurale du génome à travers différentes races. Pour le génome canin, un consortium international, dont l'équipe « Génétique du Chien » fait partie, s'est constitué afin de construire ces pangénomes.

4 Caractéristiques du génome canin

Le génome canin présente une portion euchromatique qui couvre environ 2,4 gigabases (Gb), soit 150 mégabases (Mb) de moins que chez la souris et 600 Mb de moins que chez l'humain. Cette différence s'explique principalement par une proportion plus faible d'éléments répétitifs et plus particulièrement des éléments nucléaires dispersés courts (SINEs). La teneur en guanine-cytosine (GC) de 41,5 % est comparable à celle du génome humain et le nombre de gènes annotés et d'éléments non codants du génome, varie d'une version à l'autre, mais reste semblable aux chiffres déterminés chez l'humain.

Les premières estimations indiquent que le génome canin contient environ 334 Mb d'éléments répétés (34 % du génome), contre 609 Mb (46 %) chez l'humain et 954 Mb

(40 %) chez la souris. Le génome canin possède ainsi environ 26 000 copies reconnaissables de rétrotransposons, tandis que l'espèce humaine en compte plus de 183 000 (Lindblad-Toh et al., 2005). Cependant, des analyses plus récentes suggèrent une proportion en éléments répétitifs autour de 43 % répartie en trois catégories principales : les éléments nucléaires dispersés longs (LINEs : 504 Mb), courts (SINEs : 253 Mb) (*Short or Long Interspersed Nuclear Elements*) et les séquence terminale longue répétée (LTRs : 120 Mb) (Wang et al., 2021a).

Malgré une proportion d'éléments répétés plus faible que chez la souris ou l'homme, certains éléments conservent une structure potentiellement transposable. L'exemple le plus emblématique est celui des éléments *SINEC_Cf*, une famille de SINE spécifique aux canidés. Ces éléments, particulièrement abondants dans le génome du chien, ont connu des vagues de transposition relativement récentes au cours de l'évolution, dont certaines insertions sont polymorphes entre races (Lindblad-Toh et al., 2005). De telles insertions ont été liées à des phénotypes aussi divers que la narcolepsie, la robe *merle* ou l'atrophie progressive de la rétine (Kalla et al., 2020). Ces insertions, en modifiant l'expression ou l'épissage de gènes voisins, peuvent générer de la variabilité phénotypique. Ainsi, les *SINEC_Cf* représentent une source non négligeable de nouveaux polymorphismes pouvant être captés par la sélection artificielle et pourraient contribuer à la diversité morphologique et comportementale observée chez les races canines modernes (Wang & Kirkness, 2005).

Nom de l'assemblage	UU_Cfam_GSD_1.0, Assembly GCA_011100685.1
Nombre de chromosomes	39 (38 autosomes + XY)
Nombre de scaffold	2 198
Nombre total de bases	2 482 000 080 bp
Nombre total de bases sans lacunes	2 481 941 580 bp
Nombre de gènes codants	21 063
Nombre de petits ARN non codants	3 097
Nombre de longs ARN non codants	6 540
Nombre d'autres ARN non codants	22
Nombre de pseudogènes	676
Nombre total de transcrits géniques	57 308
Teneur en G/C	41,5 %

TABLE 2 – Caractéristique de l'assemblage du génome canin (*canFam4*)

5 Annotation du génome canin

L'annotation du génome constitue une étape essentielle dans l'interprétation de la variation génétique. En séquençant plusieurs individus et en alignant leurs génomes à un assemblage de référence, les chercheurs peuvent identifier un large éventail de variants génomiques.

On peut distinguer les familles suivantes :

- polymorphismes mononucléotidiques (SNPs ou SNVs) : substitutions d'un seul nucléotide par rapport au génome de référence ;
- insertions et délétions courtes (*InDels*) : séquences nucléotidiques insérées ou supprimées d'une taille inférieure à 50 bp ;
- variations de structure (SVs) : modifications de plus grande taille, telles que :
 - variations du nombre de copies (CNVs) : délétions, insertions ou duplications de séquence de 50 bp ou plus ;
 - inversions : segments du génome ayant été inversés dans leur orientation ;
 - réarrangements complexes : combinaisons d'insertions, de délétions, de duplications ou de translocations.

Une étude de référence menée par le consortium *Dog 10 000 genomes* a analysé les génomes de 2 075 canidés, incluant 321 races domestiques, ainsi que des chiens de village, des loups et des coyotes. Ils ont ainsi identifié plus de 30 millions de variants nucléotidiques simples (SNVs), 14 millions d'*InDels* et variants mixtes, ainsi que près de 150 000 variations de structures (délétions, duplications en tandem, insertions et inversions) (Meadows et al., 2023).

Le site *Ensembl* (www.ensembl.org), tout comme celui du *NCBI* (www.ncbi.nlm.nih.gov/), maintiennent un catalogue complet et régulièrement mis à jour de variants pour différentes espèces (Dyer et al., 2025). Il existe aussi des ressources dédiées au génome canin comme *DoGSD* (ngdc.cncb.ac.cn/dogsdv3/) (Bai et al., 2015). Une représentation graphique de l'évolution de l'annotation *Ensembl* à travers les différentes versions du génome canin est présentée en figure 3. D'autres ressources majeures incluent le *UCSC Genome Browser* (genome.ucsc.edu/), utilisé pour une visualisation interactive du génome et l'accès aux annotations comparatives, ainsi que *RefSeq* (*NCBI*) et l'Université d'Uppsala, qui fournissent également des jeux de données d'annotation génique, d'ARN et de variants.

Publications des annotations de gènes et des variations génomiques dans Ensembl

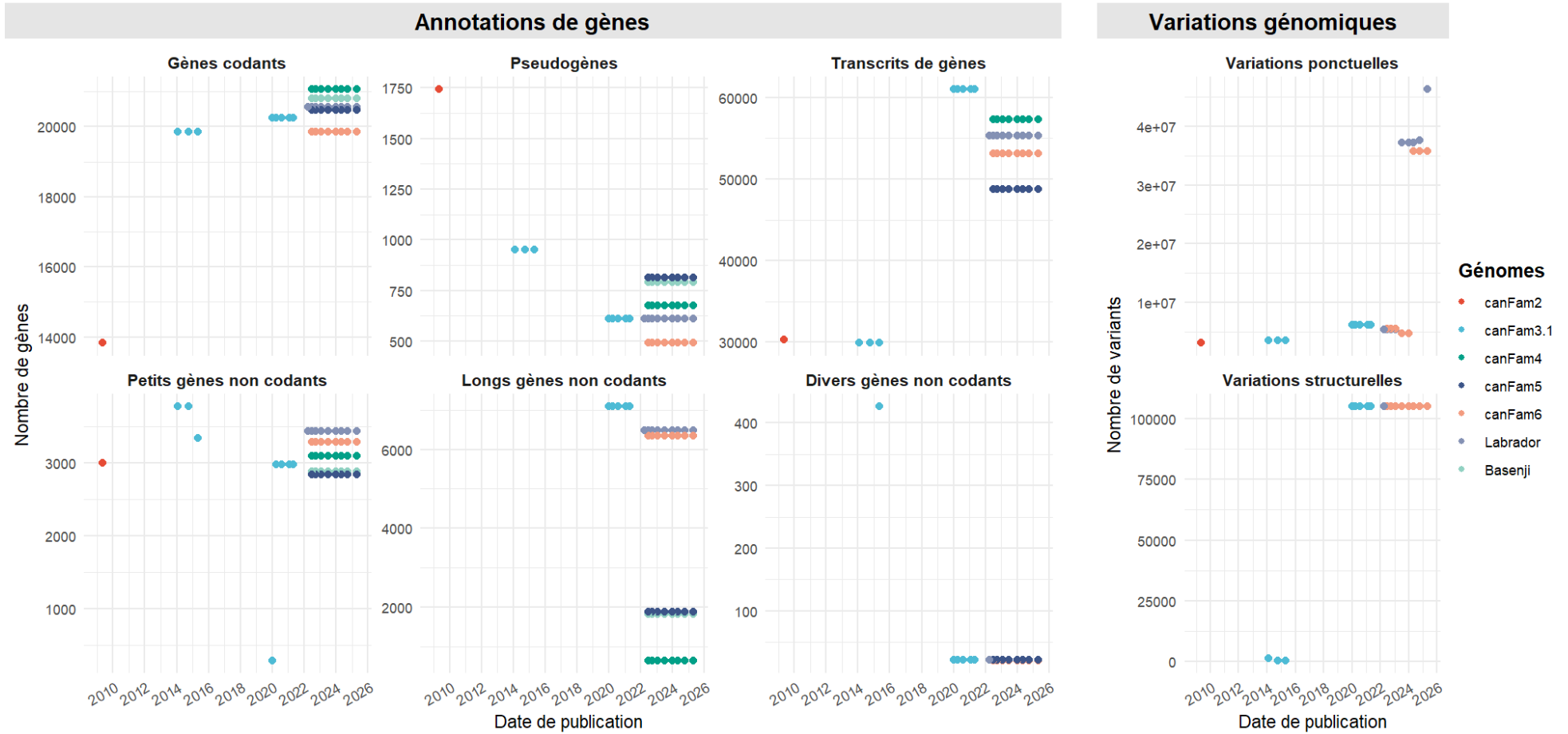


FIGURE 3 – Évolutions des annotations d'Ensembl de 2010 à 2025 pour les différentes versions d'assemblages du génome canin. Nombre de gènes codants, de petit, long et autres gènes non codants, pseudogènes, transcrits, variants courts et variants de structures annotés dans les différentes versions du génome canin dans la base de données Ensembl (https://www.ensembl.org/Canis_lupus_familiaris/genome_version/Info/Annotation).

6 Évolution et conservation génomique

Le génome canin comporte un nombre plus élevé d'autosomes que le génome humain : 38 contre 22. Cette différence se traduit par une organisation chromosomique plus fragmentée chez le chien, avec un plus grand nombre de chromosomes de plus petite taille comparativement à ceux de l'humain. La comparaison chromosomique révèle que les petits chromosomes canins (CFA) correspondent généralement à un seul chromosome humain (HSA) individuel, tandis que les grands CFA sont souvent des mosaïques composées de plusieurs segments de différents chromosomes humains. Malgré ces différences structurales, environ 94 % du génome canin se situe dans des régions de synténie conservée partagées avec les génomes humains ou murins (rat et souris) (Lindblad-Toh et al., 2005). Une représentation de cette synténie est illustrée en figure 4.

Au cours de l'évolution, la lignée canine a subi un plus grand nombre de réarrangements interchromosomiques que la lignée humaine, comme l'indique un nombre plus élevé de points de cassure chromosomique (100 chez le chien contre 83 chez l'humain). En revanche, le génome humain présente davantage de réarrangements intrachromosomiques (52 chez l'humain contre 33 chez le chien), ce qui suggère une réorganisation plus localisée au cours de l'évolution des primates. En revanche, l'organisation globale du génome canin est plus proche de l'humain que ne l'est celui de la souris, qui est encore plus fragmenté que celui du chien (Lindblad-Toh et al., 2005).

Environ 5,3 % des éléments génomiques du chien et de l'humain, situés dans un segment ancestral de 812 Mb également conservé chez la souris, sont maintenus par sélection négative, témoignant de leur importance fonctionnelle à travers les lignées mammaliennes. Ces éléments conservés, en particulier des séquences non codantes, montrent un regroupement dans environ 200 régions génomiques pauvres en gènes. Ces clusters sont fréquemment localisés à proximité de gènes essentiels à l'identité cellulaire, tels que des facteurs de transcription ou des récepteurs impliqués dans le guidage axonal, ce qui suggère des contraintes évolutives fortes dans les lignées mammaliennes (Lindblad-Toh et al., 2005).

Les analyses évolutives comparées montrent en outre que les groupes de gènes fonctionnellement apparentés tendent à suivre des schémas évolutifs parallèles chez l'humain et le chien. Cette observation appelle donc à la prudence dans l'interprétation des accélérations évolutives spécifiques à une lignée. Pour exemple, l'évolution des gènes liés au développement cérébral, décrit comme étant uniquement humaine présentent en réalité une évolution similaire chez le chien, ce qui indique que ces tendances pourraient refléter des adaptations mammaliennes générales, plutôt que des innovations proprement humaines (Lindblad-Toh et al., 2005).

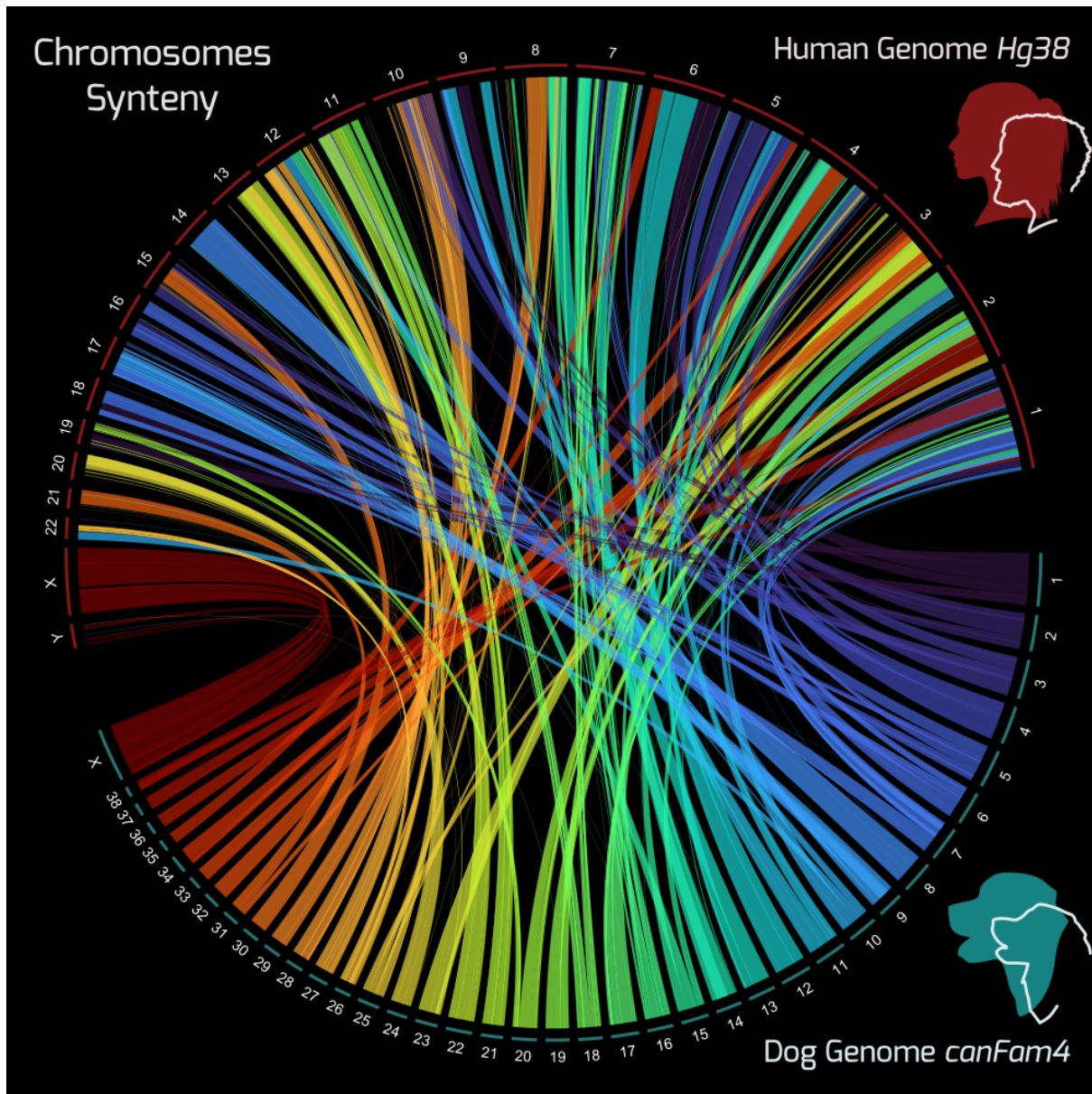


FIGURE 4 – Synténie (i.e., régions chromosomiques orthologues) entre le génome du chien (*canFam4*) et le génome humain (*hg38*).

Cette figure a été réalisée en utilisant un script R et le package *circlize*, à partir du fichier *.chain liftOver* mis à disposition par le site Genome Browser de l'Université de Californie UCSC (<https://hgdownload.soe.ucsc.edu/goldenPath/canFam4/liftOver/>). Cette figure a été primée lors du concours photo de l'IGDR 2024.

7 Déséquilibre de liaison (LD) et structure des haplotypes

Le génome canin fut profondément influencé par deux événements démographiques majeurs : un premier goulet d'étranglement survenu lors de la domestication au Pléistocène supérieur (126 000 BP - 11 700 BP), suivi d'un second, plus récent et plus intense, résultant de la sélection artificielle exercée lors de la formation des races (depuis ≈ 200 ans). Contrairement à l'humain, dont la population est largement panmictique, ces événements ont entraîné chez le chien la formation de blocs de déséquilibre de liaison (LD) exceptionnellement longs, souvent 10 à 100 fois plus étendus, ainsi qu'une forte réduction de la diversité des haplotypes au sein des races (Sutter et al., 2004).

Comme le montre les résultats de Lindblad-Toh et al. (2005), le LD intrarace peut s'étendre sur plusieurs mégabases (Mb), tandis que le LD entre races s'étend généralement sur quelques dizaines de kilobases (Kb). Cette organisation aboutit à des blocs d'haplotypes ancestraux d'environ 5 à 10 Kb et des blocs races-spécifique pouvant atteindre jusqu'à 100 Kb et comprenant généralement 3 à 5 allèles.

Cette structure génomique a un impact considérable sur la variation génétique. Parker et al. ont ainsi rapporté une homogénéité génétique importante au sein des races ($\approx 94,6$ %) et une hétérogénéité génétique extrêmement marquée entre races (27,5 %), en contraste frappant avec les valeurs observées chez l'humain (72,5 % de variation intrapopulation ; 5,4 % entre populations) (Parker et al., 2004).

Les analyses du génome de référence *canFam4* confirment cette structuration particulière, révélant un génome en mosaïque composé de longues régions quasi homozygotes et de segments fortement hétérozygotes, répartis de manière variable selon les races (Lindblad-Toh et al., 2005). Cette combinaison unique, longues régions en déséquilibre de liaison au sein des races, faible diversité des haplotypes et forte différenciation entre races, a des implications majeures pour les études d'association. En particulier, cela suggère que de nombreux allèles à risque peuvent être partagés entre plusieurs races tout en étant rares, voire absents, dans d'autres, ce qui souligne l'importance d'approches comparatives inter-races en génétique canine.

8 Conclusion

Le génome canin présente donc une combinaison unique de conservation évolutive avec l'humain et des caractéristiques structurales distinctes, façonnées par la domestication et la sélection artificielle. Ces particularités, notamment les longues régions en déséquilibre de liaison au sein des races et l'hétérogénéité génétique marquée entre les races, font du chien un modèle puissant pour les études génétiques des caractères complexes, avec des tailles d'échantillons relativement modestes et une densité de marqueurs réduite, ce qui ne serait pas envisageable chez l'humain (Ostrander & Wayne, 2005).

Ensemble, ces caractéristiques soulignent l'intérêt du chien en génomique translationnelle et son potentiel à fournir des informations pertinentes pour la santé vétérinaire et humaine.

APPROCHE *ONE HEALTH*

A Définition et principes clés

Le concept *One Health* ou « Une seule santé », est une approche intégrée et unificatrice apparue au début des années 2000. Il repose sur le constat que la santé humaine, la santé animale et celle des écosystèmes sont étroitement liées et interdépendantes (voir figure 5). Selon l'Organisation Mondiale de la Santé (OMS), son objectif est d'optimiser les ressources et de renforcer la coordination entre les différents secteurs concernés.

Cette approche implique de nombreuses disciplines traditionnellement cloisonnées. Elle favorise ainsi la collaboration entre la médecine vétérinaire, la médecine humaine, les sciences de l'environnement et les acteurs de la santé publique.

Initialement portée par une vision anthropocentrée et utilitariste, visant un bénéfice direct pour l'humain, elle tend aujourd'hui à évoluer vers une perspective plus holistique, prenant en compte le bien-être des autres espèces et la préservation des écosystèmes, même en l'absence d'un bénéfice immédiat pour l'humain.

B Justification scientifique

Selon l'OMS :

« Environ 60 % des maladies infectieuses émergentes notifiées dans le monde proviennent d'animaux, qu'ils soient sauvages ou domestiques. Plus de 30 nouveaux agents pathogènes humains ont été détectés au cours des 30 dernières années, dont 75 % sont d'origine animale. »

Ce phénomène s'explique notamment par l'intensification des échanges internationaux, la mondialisation et la perturbation croissante des écosystèmes par l'activité humaine.

Mais au-delà des maladies infectieuses, le paradigme *One Health* s'applique également aux maladies non infectieuses, notamment celles à déterminisme génétique complexe. En effet, les espèces animales partagent avec l'humain une longue histoire évolutive et des bases génomiques communes. Cette proximité phylogénétique se traduit par la conservation de nombreux mécanismes biologiques fondamentaux : voies de signalisation, développement embryonnaire, voies métaboliques, immunité, ainsi que des systèmes plus spécialisés, eux aussi très proches, comme par exemple, le développement osseux ou l'homéostasie articulaire.

Ainsi, des altérations dans ces mécanismes peuvent conduire à des dérèglements pathologiques homologues entre espèces. L'étude de ces maladies chez les animaux et en particulier chez le chien, constitue un levier pertinent pour mieux comprendre leur origine, leur expression et leur variabilité chez l'humain.

C Pertinence du chien dans une approche *One Health*

L'espèce canine occupe une place particulière dans le cadre du concept *One Health*, en raison de son histoire évolutive unique avec l'humain et aujourd'hui avec le partage étroit de notre niche écologique. À ce titre, les chiens sont exposés aux mêmes facteurs environnementaux que les humains : pollution de l'air, contaminants alimentaires, perturbateurs endocriniens, bruit, sédentarité ou encore obésité. Cette proximité environnementale et comportementale fait d'eux de véritables sentinelles de santé publique, mais aussi des modèles pertinents pour l'étude des maladies liées au mode de vie moderne.

Sur le plan anatomique et physiologique, le chien présente de fortes similarités avec l'humain, notamment en ce qui concerne le système musculosquelettique, nerveux et cardiovasculaire, malgré des différences évidentes en termes de posture et de locomotion. Ces ressemblances permettent l'application de techniques médicales comparables, tant en imagerie qu'en chirurgie ou en thérapeutique.

La médecine vétérinaire partage d'ailleurs de nombreuses méthodes diagnostiques et approches thérapeutiques avec la médecine humaine. L'essor de la médecine vétérinaire de spécialité et plus récemment de la médecine vétérinaire de précision, renforce encore cette convergence. Des approches croisées enrichissent ainsi les deux disciplines : par exemple, des innovations en oncologie vétérinaire sont aujourd'hui transférées à la cancérologie humaine et inversement.

De nombreuses études viennent illustrer concrètement cette valeur du modèle canin dans le cadre du concept *One Health*. Une étude réalisée en France a ainsi pu mettre en évidence une corrélation entre la distribution géographique des cas de lymphomes canins dits non-Hodgkiniens et la présence de facteurs de risques environnementaux tels que des décharges publiques ou des incinérateurs (Pastor et al., 2009). Une autre étude a quant à elle montré que les mêmes régions chromosomiques sont impliquées dans l'adaptation au froid, notamment à l'hypoxie, entre l'humain et le chien sur les hauts plateaux du Tibet (Gou et al., 2014).

Enfin, d'un point de vue éthique, le chien de compagnie peut être considéré comme un patient à part entière, soigné pour son propre bénéfice. Cela le distingue des modèles animaux expérimentaux classiques, chez lesquels la maladie est souvent induite.

Étudier des affections naturellement présentes chez le chien, comme la dysplasie coxo-fémorale, permet ainsi de réduire le recours à l'expérimentation animale, tout en fournissant des données biologiques et cliniques plus transposables.



FIGURE 5 – Concept « Une Seule Santé » (*One Health*)

Cette proposition de logo se distingue des précédents représentant le concept One Health en mettant l'humain au même niveau que l'environnement et les autres espèces et en ne le mettant pas au centre.

ETUDE DES MALADIES GÉNÉTIQUES

A Les maladies génétiques

Les maladies génétiques peuvent être classées selon le nombre de gènes impliqués et l'influence de l'environnement, formant un continuum allant des affections monogéniques aux maladies complexes, faisant intervenir plusieurs gènes et cofacteurs. Les maladies génétiques héréditaires présentent des altérations génétiques dans l'ADN germlinal qui vont donc pouvoir se transmettre aux générations suivantes.

Les maladies monogéniques autrement appelées « simples » ou « mendéliennes » résultent de mutations dans un seul gène. Leurs modes de transmission sont simplifiés, ils peuvent être autosomique dominant, récessif ou lié à un chromosome sexuel. Cependant l'expression d'un gène ou bien l'expression clinique de la maladie peut être modulée par l'interaction de plusieurs gènes. On parle alors de maladies oligogéniques et leur mode de transmission est alors moins prévisible. Les maladies polygéniques, quant à elles, résultent de l'effet cumulé de nombreux variants, chacun ayant un impact faible mais significatif sur le risque de développer la maladie. Leur étude nécessite des cohortes importantes de cas et de contrôles et des approches statistiques robustes comme les études d'association pangénomiques (GWAS).

Enfin, les maladies dites complexes résultent d'interactions multiples entre le génome et l'environnement, comme c'est le cas pour de nombreuses affections chroniques telles que le diabète, les allergies ou certains troubles psychiatriques, les cancers, etc.

Ces catégories ne sont pas étanches et illustrent un gradient de complexité croissante de ces maladies héréditaires, tant dans leur compréhension que dans leur modélisation génétique (Badano & Katsanis, 2002).

L'objectif des études génétiques est donc d'identifier les variants et loci (i.e. positions fixes dans le génome) associés aux maladies génétiques héréditaires afin de décrypter leurs mécanismes pathogéniques, d'évaluer le risque génétique individuel et de guider la sélection ou le développement de stratégies thérapeutiques ciblées. Cependant, cette quête se heurte à l'héritabilité souvent incomplète des traits, à l'effet modeste de nombreux variants et aux interactions gènes et environnement qui peut fausser les résultats des études génétiques.

B Obtention des données génomiques

L'étude des bases génétiques des maladies complexes repose sur l'accès à des données génomiques de qualité. Trois grandes approches sont aujourd'hui utilisées pour générer ces données, chacune avec ses avantages, ses limites et des implications distinctes pour les analyses qui s'ensuivent.

Le génotypage, à l'aide de puces SNP (polymorphismes mononucléotidiques), permet de génotyper plusieurs centaines de milliers à quelques millions de SNP préalablement sélectionnés dans le génome, sur autant d'individus que souhaité. Cette méthode est historiquement la plus utilisée en génétique des populations et dans les études d'association pangénomiques (GWAS). Cette méthode possède en effet des avantages certains : son faible coût et sa rapidité d'exécution la rendent idéale pour l'étude de grands effectifs. Cependant, en raison de l'ensemble prédéfini de variants étudiés, elle limite la découverte de nouveaux variants et présente un biais inhérent lié aux populations de référence sur lesquelles les SNP ont été identifiés (LaFramboise, 2009).

Le séquençage du génome entier à haute couverture ($\geq 20-30\times$) (*Whole Genome Sequence*, WGS) offre une exploration complète du génome incluant les SNP rares, les *InDels*, les CNV et les régions non codantes, celles répétées, etc. Il n'y a ainsi pas de biais de sélection et la qualité ainsi que la résolution des données sont extrêmement élevées. Cette approche permet de surmonter les limitations des GWAS traditionnels, notamment en identifiant des variants rares contribuant à l'hérabilité manquante et en facilitant la cartographie fine des loci associés. Cependant, cette technologie reste coûteuse, tant pour le séquençage lui-même que pour les ressources computationnelles nécessaires au traitement des données générées, ce qui est particulièrement contraignant pour l'étude des maladies complexes nécessitant de grandes cohortes (Wu et al., 2016).

Il est également possible de recourir au séquençage de l'exome entier (*Whole Exome Sequencing*, WES), une approche intermédiaire entre les puces SNP et le séquençage du génome entier. Le WES cible spécifiquement les régions codantes du génome, qui représentent environ 1 à 2 % de l'ADN. Cette méthode offre une couverture approfondie des exons, permettant une détection efficace des variants rares, notamment les mutations ponctuelles et les petites insertions / délétions, tout en générant un volume de données et un coût plus réduit que le séquençage complet du génome. Toutefois, cette technologie présente aussi des limitations : elle ne couvre pas les régions non codantes, peut manquer certains variants de structures complexes et ne garantit pas une couverture uniforme de tous les exons, en particulier ceux présentant une forte teneur en GC ou des séquences répétées (Seaby et al., 2016).

Le séquençage à faible profondeur ($0,5\times$ à $4\times$, dit *low-pass*) est une alternative à ces trois approches qui permet de couvrir l'ensemble du génome pour un coût modeste

rivalisant avec un génotypage sur puce SNP. Cependant la qualité des génotypes obtenus est très variable car le génome est lu de manière hétérogène et n'est couvert que par un faible nombre de lectures. Cet inconvénient peut cependant être compensé par l'utilisation de méthodes d'imputation statistique permettant d'inférer les génotypes manquants et d'évaluer la qualité de ceux observés à l'aide d'un panel de référence bien choisi et bien calibré. Cette approche est de plus en plus utilisée pour combiner faible coût et richesse des données. Il est important de noter que la qualité de l'imputation dépend de la fréquence de l'allèle mineur (MAF) des variants ainsi que de la taille du panel de référence, de la proximité génétique entre les individus du panel et de l'individu à imputer ainsi que de la couverture de séquençage de ce dernier. D'autres paramètres peuvent aussi entrer en compte comme la structure de la population et la taille des régions en déséquilibre de liaison. Cette approche relativement nouvelle est prometteuse afin d'améliorer la précision des analyses d'association pangénomiques et d'identifier les variants causaux ou de prédisposition, mais nécessite l'application d'imputation génomique (Li et al., 2021).

C Approches classiques d'analyse génétique

A partir des données génomiques acquises, plusieurs méthodes d'analyses existent afin d'identifier les gènes responsables de maladies héréditaires ou de phénotypes particuliers. Si l'analyse de liaison génétique a été la première utilisée avec des marqueurs de type microsatellites, l'analyse d'association pangénomique (GWAS) a rapidement été privilégiée grâce à l'émergence des puces SNP. Une comparaison schématique des deux méthodes est illustrée dans la figure 6. D'autres méthodes existent néanmoins et sont utilisées en fonction de leur complémentarité, forces et limites, selon les populations et les traits étudiés.

1 Études de liaison génétique (*Genetic Linkage Analysis*)

L'analyse de liaison génétique est l'une des premières méthodes développées pour identifier les régions chromosomiques associées à des phénotypes ou à des maladies héréditaires, en suivant la ségrégation des allèles dans la descendance. Historiquement mise en œuvre à partir de marqueurs microsatellites, elle repose sur l'étude de la co-ségrégation entre un phénotype donné et des allèles de marqueurs génétiques au sein d'une ou de plusieurs familles. Cette approche a pour but d'identifier les allèles qui sont transmis avec la maladie plus souvent qu'attendu par le hasard et dépend du taux de recombinaison lors de la méiose entre le marqueur et la mutation causale. Ainsi, plus un marqueur est physiquement proche de la mutation, plus la probabilité de recombinaison entre eux est faible et plus la mutation recherchée ségrégera avec un allèle du marqueur.

La mesure de la liaison génétique repose donc sur le taux de recombinaison, exprimé en centimorgan (cM), où 1 cM correspond à une probabilité de recombinaison de 1 %. Ce taux est influencé non seulement par la distance physique entre les loci, mais également par la composition génomique de la région chromosomique impliquée, incluant le contenu en GC, la densité de séquences répétées ou encore l'état de méthylation de l'ADN. Le degré de liaison est quantifié à l'aide du LOD score (*Logarithm Of Odds*), qui compare la probabilité d'observer les données familiales sous l'hypothèse de liaison génétique à une distance donnée (pour une valeur de taux de recombinaison $\theta \leq 0,5$) à celle attendue en l'absence de liaison ($\theta = 0,5$). Le taux de recombinaison θ (thêta) reflète la probabilité de recombinaison entre deux loci : $\theta = 0$ indique une liaison complète (aucune recombinaison), tandis que $\theta = 0,5$ correspond à une absence totale de liaison (ségrégation indépendante). Une valeur de LOD score maximale ≥ 3 indique une liaison au θ correspondant, tandis qu'un score ≤ -2 permet de rejeter cette hypothèse (Bailey-Wilson & Wilson, 2011).

Cette approche a été largement utilisée dans le modèle canin pour cartographier des maladies et des phénotypes « monogéniques », tels que la narcolepsie chez le Doberman (Mignot et al., 1994) ou le phénotype *merle*, une spécificité de couleur de robe grisonnante en taches, dans certaines races de bergers (Hédan et al., 2006). L'analyse de liaison génétique est également applicable à des maladies multifactorielles, bien que plus complexe à mettre en œuvre. Dans ce cadre, la pénétrance des génotypes joue un rôle essentiel : elle détermine la probabilité qu'un individu présentant un certain génotype exprime le phénotype pathologique. Lorsque la pénétrance est incomplète ou que des phénocopies (i.e., phénotypes similaires mais indépendants du génotype) existent, la relation entre génotype et phénotype devient incertaine, rendant les analyses plus difficiles à interpréter. Cette situation est fréquente dans les maladies complexes, où l'effet combiné de multiples variants génétiques et de facteurs environnementaux altère la correspondance génotype-phénotype. Plusieurs études ont tout de même utilisé cette approche dans les années 2000 pour cartographier les loci impliqués dans la dysplasie de la hanche, tout d'abord avec des panels de microsatellites (Chase et al., 2004 ; Todhunter et al., 2005 ; Marschall & Distl, 2007) puis avec des puces SNP (Zhu et al., 2012).

Bien que l'analyse de liaison génétique ait été supplantée dans certains contextes par les études d'association, elle demeure un outil puissant, notamment dans les familles avec une forte agrégation phénotypique. Elle conserve un intérêt particulier pour la détection de variants rares à effets forts et gagne à être utilisée en combinaison avec d'autres méthodes pour affiner les régions candidates et mieux comprendre la génétique des maladies complexes. Cette méthode connaît un regain d'intérêt, stimulé par les avancées des technologies de séquençage et les améliorations méthodologiques

récentes (Ott et al., 2015), mais elle implique une bonne connaissance des arbres généalogiques de l'espèce étudiée et du trait suivi.

2 Études d'association pangénomiques (GWAS)

Les études d'association pangénomiques (*Genome Wide Association Studies*, GWAS) sont des analyses statistiques explorant l'ensemble du génome sans hypothèse ni région chromosomique candidate *a priori*. Elles visent à identifier des associations entre un phénotype et les allèles de nombreux variants génétiques, généralement plusieurs millions de polymorphismes mononucléotidiques (SNPs). Les GWAS reposent sur l'hypothèse clé que les SNPs génotypés ou bien séquencés capturent, grâce au déséquilibre de liaison, des variants prédisposants ayant un effet mesurable sur le phénotype étudié. Les GWAS consistent en la comparaison des fréquences alléliques entre des groupes de cas (individus atteints) et de témoins (individus non atteints). Les tests statistiques, tels que le test du chi-deux ou les régressions logistiques univariés et multivariés, permettent ainsi d'évaluer la significativité des associations observées en prenant en compte ou non des covariables (Cardon & Bell, 2001).

Cependant, la multiplicité des tests statistiques réalisés lors d'une étude d'association pangénomique (souvent plusieurs centaines de milliers à plusieurs millions) impose l'application de corrections rigoureuses pour contrôler le taux de faux positifs. La méthode classique repose sur une correction de Bonferroni, d'où découle le seuil de significativité génome-entier de 5×10^{-8} , historiquement établi en supposant environ un million de tests indépendants. Des travaux récents suggèrent cependant qu'un seuil plus permissif de 5×10^{-7} peut être envisagé, notamment dans les études de grande taille, car il augmente la puissance de détection tout en n'entraînant qu'une modeste hausse du taux de faux positifs (Chen et al., 2021).

De ce fait, une cohorte importante de cas et de témoins est nécessaire pour détecter les effets modestes des différents variants associés à une maladie complexe. La stratification de population, c'est-à-dire la présence de sous-groupes génétiquement distincts au sein de l'échantillon, peut également introduire des biais et produire de faux positifs. Pour détecter cette stratification un filtre des variants ne respectant pas l'équilibre de Hardy Weinberg est souvent appliqué. Cet équilibre est un modèle théorique selon lequel les fréquences génotypiques d'un locus restent constantes de génération en génération, à condition que la population soit panmictique (accouplements aléatoires), de grande taille et qu'il n'y ait ni sélection, ni mutation, ni migration. Pour un locus biallélique (A et a) avec des fréquences alléliques p et q , les fréquences génotypiques attendues sont données par la loi $p^2 + 2pq + q^2 = 1$, soit p^2 pour AA , $2pq$ pour Aa et q^2 pour aa . La conformité des données observées à ce modèle est testée à l'aide d'un test du chi-deux (χ^2) comparant les fréquences génotypiques observées à celles attendues. Si la statistique χ^2 est inférieure au seuil critique (généralement 3,84 pour un

risque $\alpha = 5\%$ avec un degré de liberté), l'équilibre est supposé respecté. À l'inverse, une déviation significative peut signaler une stratification génétique ou d'autres biais (Turner et al., 2011).

Cependant, cet équilibre n'est pas toujours respecté, notamment pour les espèces domestiques, où la reproduction est contrôlée par l'humain, rompant ainsi le principe d'appariement aléatoire. Afin de contrôler la structure de population potentiellement présente, il est recommandé de reproduire les études dans des populations indépendantes et d'appliquer différentes méthodes de correction statistique. Parmi celles-ci, le contrôle génomique (*genomic control*) permet d'ajuster les statistiques de test à l'aide d'un facteur d'inflation calculé à partir de SNPs supposés neutres (Devlin & Roeder, 1999). Des approches bayésiennes ou de vraisemblance, comme les méthodes d'association structurées (*structured association*), modélisent de leur côté explicitement la structure génétique de la population afin de limiter les faux positifs liés à la stratification, mais sont complexes à mettre en place (Zhou et al., 2013). Enfin l'analyse en composantes principales (ACP) identifie quant à elle les principaux axes de variation génétique au sein de la population et les utilise comme covariables pour ajuster les analyses. Le nombre de dimensions à intégrer dans la régression linéaire varie en fonction de chaque étude et peu de recommandations existe à ce sujet (Zhao et al., 2018).

Malgré leurs apports majeurs, notamment dans l'identification de nombreux loci associés à des maladies complexes, les études d'association pangénomiques (GWAS) présentent plusieurs limites. Elles n'expliquent qu'une faible part de l'héritabilité manquante, ne permettent pas toujours d'identifier les variants causaux et détectent difficilement les interactions génétiques (épistasie) ou gène-environnement. Les résultats peuvent aussi être biaisés par une stratification de population cryptique ou refléter des associations synthétiques entre SNPs communs et variants rares. Enfin, leur valeur prédictive clinique reste souvent limitée. Néanmoins, les GWAS ont permis de révéler un grand nombre de loci impliqués dans des affections comme la dysplasie de la hanche, aussi bien chez l'humain (Harsanyi et al., 2020b ; Wen et al., 2022) que chez le chien (Pascual-Garrido et al., 2018 ; Mikkola, 2020). Ces résultats, obtenus grâce à une approche exploratoire agnostique, doivent cependant être confirmés par une autre étude de réplication (König, 2011) ou par des analyses fonctionnelles ciblées afin de valider les variants d'intérêt et comprendre leur rôle biologique réel.

a Particularités de l'espèce canine pour les études d'association

Le fort déséquilibre de liaison (LD) observé dans le génome canin renforce considérablement la puissance et l'efficacité des études d'association pangénomiques (GWAS) (Quignon et al., 2007). Ces régions, pouvant être de grandes tailles chez le chien, sont associées à la forte homozygotie des blocs haplotypiques et font de chaque race des « outils » de choix pour les analyses génétiques, à l'instar des populations humaines

isolées, comme les Mormons, les Islandais, certaines populations juives, etc. En effet, dans une population isolée, pour identifier les mutations responsables de maladies génétiques héréditaires, des régions génomiques avec des LD longs permettent d'utiliser un plus petit nombre de marqueurs et d'individus pour interroger le génome avec une puissance statistique satisfaisante (Karlsson et al., 2007).

Par exemple, 15 000 SNP répartis uniformément peuvent suffire pour détecter des allèles de risque ayant un facteur de risque entre 2 et 5, en utilisant seulement 100 cas et 100 témoins au sein d'une même race (Lindblad-Toh et al., 2005). De plus, les régions soumises à une sélection récente peuvent être identifiées par des signatures telles que des blocs de LD étendus et des réductions locales de l'hétérozygotie, caractéristiques de sélection positive.

Les études d'association commencent généralement par des analyses intrarace, où le LD est le plus étendu et l'homogénéité génétique la plus élevée. Cette approche est optimale pour la détection initiale des loci. En revanche, la comparaison inter-races permet une cartographie plus fine des régions associées, car les différences de structure de LD entre les races permettent de resserrer les régions candidates en exploitant les points de recombinaisons et la diversité des points de cassure. Cette approche est particulièrement puissante à la condition que les races sélectionnées partagent des causes génétiques communes pour la maladie étudiée.

Un point crucial dans les études génétiques chez le chien est que les hypothèses de l'équilibre de Hardy-Weinberg (HW) ne s'appliquent que rarement en raison de la sélection artificielle. Cet équilibre suppose un accouplement des individus de manière aléatoire, des générations non chevauchantes, une taille de population efficace importante et l'absence de pressions évolutives, des conditions qui ne reflètent pas les pratiques d'élevage chez les chiens domestiques (Short et al., 2007). La sélection artificielle, notamment via des pratiques comme l'effet « étalon populaire » (où un nombre restreint de mâles contribue de manière disproportionnée au pool génétique), réduit drastiquement la taille de population effective et augmente la dérive génétique. De plus, les générations chevauchantes, la sélection directionnelle intense pour des traits comme la couleur de robe, le comportement ou la morphologie, ainsi que la consanguinité marquée, modifient fortement la distribution des fréquences alléliques.

Le non-respect de l'équilibre de Hardy-Weinberg peut favoriser la détection, lors des GWAS, d'allèles récessifs délétères masqués par l'hétérozygotie. Cependant, la stratification de populations et la consanguinité peuvent entraîner des biais, où les différences de fréquence allélique entre cas et témoins reflètent la structure de race plutôt que de véritables associations, générant ainsi de faux positifs si ces facteurs ne sont pas correctement contrôlés.

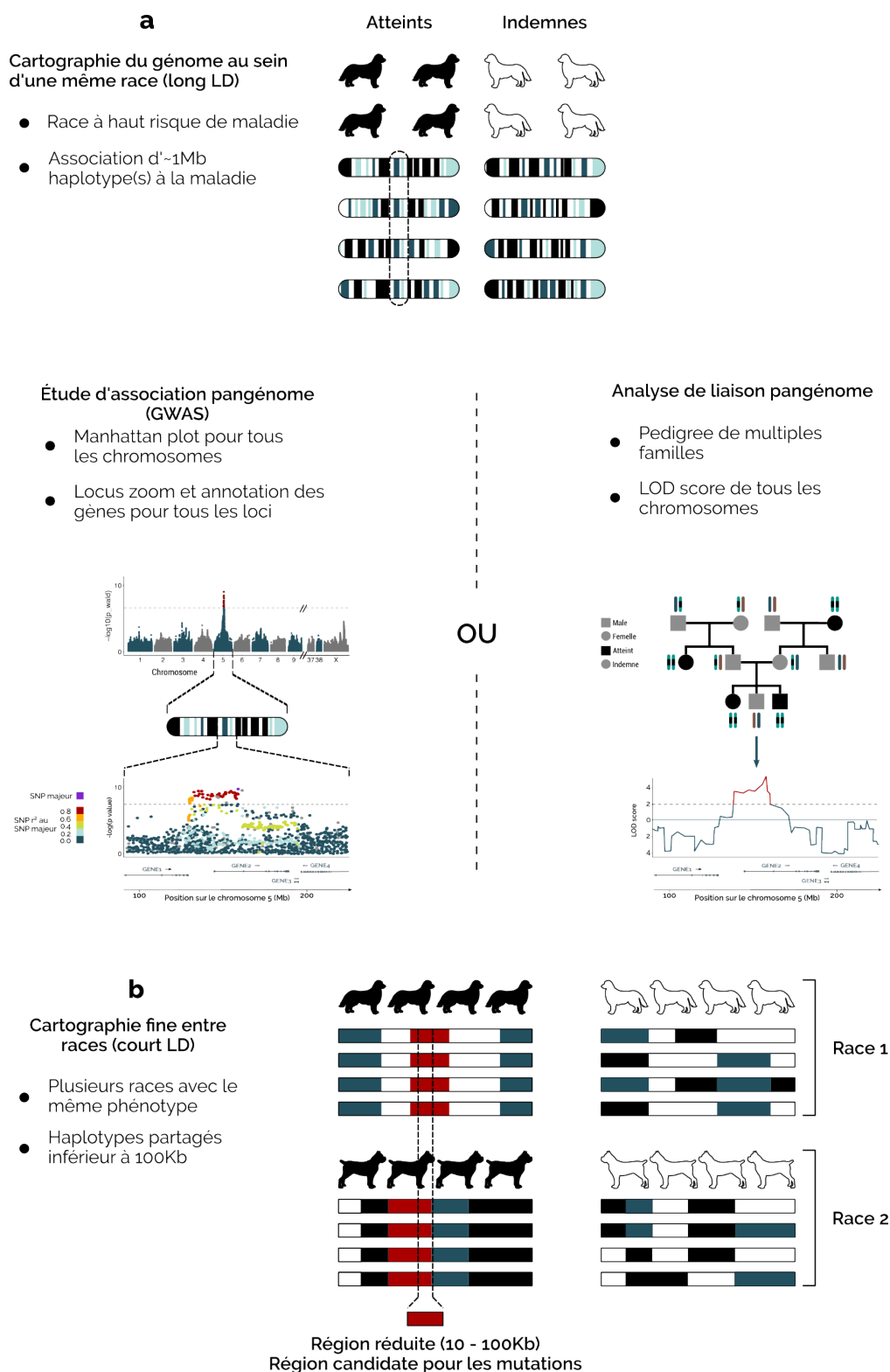


FIGURE 6 – Stratégie de cartographie des maladies en deux étapes

Une approche en deux étapes permet de tirer pleinement parti du déséquilibre de liaison (LD) étendu au sein des races, ainsi que des haplotypes ancestraux courts partagés entre les races. a) Lors de la première étape, une étude d'association (GWAS) ou une analyse de liaison pangénomique est réalisée séparément pour chaque race. Cela permet d'identifier de longs haplotypes, couvrant plusieurs mégabases, associés à la maladie. b) Dans la seconde étape, une cartographie fine est effectuée en exploitant les différences de points de recombinaison entre les différentes races. Inspiré de Karlsson & Lindblad-Toh (2008).

3 Études gènes candidats (*Candidate Gene Studies*)

Les premières études génétiques se sont concentrées sur des polymorphismes localisés dans des gènes dits « candidats », c'est-à-dire des gènes codant des protéines dont la fonction est supposée intervenir dans le processus physiopathologique de la maladie étudiée. Le choix de ces gènes repose sur plusieurs sources d'informations complémentaires : la connaissance des mécanismes physiopathologiques, les données issues de modèles animaux, l'analyse transcriptomique, ainsi que les recherches *in silico* dans la littérature et les bases de données génomiques. Une compréhension fine des mécanismes biologiques altérés est nécessaire afin de cibler les voies métaboliques potentiellement impliquées et d'en déduire une liste de gènes d'intérêt.

Deux approches peuvent être envisagées pour identifier des gènes candidats. La première est une approche directe, qui consiste à sélectionner des gènes déjà impliqués dans le phénotype étudié dans d'autres espèces, afin d'évaluer leur implication potentielle chez l'espèce cible. La seconde approche repose sur l'analyse des gènes situés dans des régions génomiques (loci) préalablement identifiées par des méthodes comme la liaison génétique ou les GWAS. On examine alors les gènes contenus dans ces régions en tenant compte de leur fonction biologique, de leur expression dans les tissus pertinents ou de leur rôle connu dans des maladies analogues.

Malgré son efficacité prouvée dans l'identification de gènes responsables de maladies monogéniques, la stratégie des gènes candidats souffre de limitations importantes. Elle n'est ni exhaustive ni exploratoire. De plus, le nombre élevé de gènes potentiellement candidats, qu'ils soient sélectionnés pour des raisons fonctionnelles ou positionnelles, peut engendrer un travail de criblage considérable, chronophage et coûteux. Avec l'essor des nouvelles technologies, notamment le séquençage à haut débit, d'autres approches, plus globales et moins biaisées, sont devenues plus performantes. Toutefois, la stratégie des gènes candidats reste un outil complémentaire précieux, en particulier lorsqu'elle est intégrée à une démarche multimodale combinant différentes sources d'information.

D Études de validation fonctionnelle

Après l'identification de loci associés à un phénotype par des approches génétiques exploratoires décrites précédemment, une étape cruciale consiste à affiner ces régions par séquençage ciblé ou analyse fine de l'haplotype afin d'identifier les mutations causales ou prédisposantes potentielles. Les variants initialement détectés ne sont le plus souvent que des marqueurs en déséquilibre de liaison avec ces mutations.

Il est donc essentiel de réaliser une validation fonctionnelle des variants présents dans les loci identifiés afin d'évaluer leur impact biologique. Cette étape permet de distinguer les variants pathogènes, des variants bénins ou simplement associés statistiquement. Pour répondre à ce défi, de nombreuses méthodologies ont été développées. On peut distinguer, la validation par méthode computationnelle reposant sur l'intégration d'annotations provenant de base de données de référence et la validation expérimentale ayant pour but l'observation directe de l'impact des variants observés.

1 Validation fonctionnelle *in silico*

En premier lieu les variants identifiés sont généralement annotés à l'aide d'outils tels que *VEP* (McLaren et al., 2016), *SIFT* (Ng & Henikoff, 2001) ou *ANNOVAR* (Wang et al., 2010), qui s'appuient sur des bases de données de référence comme *Gencode*, *RefSeq* ou *Ensembl* afin de prédire l'impact fonctionnel de ces variants génétiques (Pilalis et al., 2025). D'une manière générale, la conservation évolutive des variants ou de la région dans laquelle ils se situent est un élément important sur leur rôle fonctionnel.

Des méthodes d'enrichissement fonctionnel permettent ensuite de relier les variants associés à des traits complexes à des contextes cellulaires ou moléculaires spécifiques. Des outils comme *LDSC* (Bulik-Sullivan et al., 2015), *fGWAS* (Huang et al., 2017) ou *GARFIELD* (Iotchkova et al., 2019) exploitent des annotations fonctionnelles et profils d'expression pour identifier des enrichissements dans certains types cellulaires, tandis que *SNPsea* (Slowikowski et al., 2014) ou *CHEERS* (*TrynkaLab/CHEERS*, 2024) utilisent des données d'expression ou de chromatine pour identifier les types cellulaires impliquées. L'analyse de colocalisation, possible par le biais de logiciel comme *COLOC* (Giambartolomei et al., 2014) ou *eCAVIAR* (Hormozdiari et al., 2016), cherche à associer un même variant à la fois à un trait complexe et à un locus de caractères quantitatifs grâce à des bases de données tel que *GTEX*. Cela permet d'établir des liens plus directs entre les loci identifiés par les GWAS et les mécanismes moléculaires, tels que l'expression génique (données transcriptomiques). Ces approches restent observationnelles et statistiques, sans pouvoir prouver la causalité et dépendent souvent de données issues de tissus hétérogènes. Elles constituent néanmoins un cadre puissant pour interpréter les résultats d'association génétique (Cano-Gamez & Trynka, 2020).

Les scores de risque polygénique (PRS) estiment la prédisposition génétique aux maladies complexes en agrégeant les effets de nombreux variants génétiques, souvent issus des GWAS. Traditionnellement, ces scores sont calculés en pondérant les allèles de risque par leurs effets estimés et en les sommant. Cependant l'utilisation et l'interprétation de ces scores est critiquée en raison des différences génétiques entre les populations et de la compréhension fonctionnelle insuffisante des variants inclus. Pour améliorer la précision et la généralisation des PRS, des méthodes récentes intègrent des annotations fonctionnelles telles que l'accessibilité de la chromatine, l'expression génique ou la conservation évolutive (Jayasinghe et al., 2024).

2 Validation fonctionnelle expérimentale

L'approche expérimentale vise à confirmer la causalité des variants identifiés au cours des analyses génétiques. Il ne s'agit pas ici de dresser une liste exhaustive des manipulations possibles, car les stratégies expérimentales dépendent fortement de la nature du variant (e.g., codant, non codant, structurel, etc.), du type cellulaire concerné et du phénotype étudié. Nous présentons donc ici quelques principes généraux et exemples d'approches couramment utilisées. Ces approches restent néanmoins lourdes à mettre en œuvre et sont donc généralement réservées aux variants pour lesquels les données précédentes suggèrent fortement une implication fonctionnelle.

L'analyse comparative de l'expression génique entre tissus pathologiques et témoins constitue une méthode clé pour cette validation fonctionnelle. En combinant les données transcriptomiques avec des bases d'annotations telles que *Kyoto Encyclopedia of Genes and Genomes* (KEGG) ou *Gene Ontology* (GO), il est possible de détecter des voies métaboliques et des processus biologiques altérés. Cela permet de caractériser les dysrégulations moléculaires associées au phénotype étudié. Toutefois, les associations bio-informatiques doivent être interprétées avec prudence : chacune nécessite une validation critique, en confrontant les liens suggérés par les bases d'annotation à l'état réel des connaissances fonctionnelles issues de la littérature.

Une autre stratégie expérimentale repose sur les expériences de rescue fonctionnel, qui consistent à réintroduire la version sauvage d'un gène dans des cellules déficientes, issues de patients ou de modèles *knock-out/down* de modèles expérimentaux, afin d'observer la restauration d'un phénotype mesurable (activité enzymatique, respiration mitochondriale, etc.). Pour aller plus loin, l'édition génomique ciblée par technologie *CRISPR/Cas9* ou bien par mutagenèse dirigée permet d'introduire ou de corriger des variants dans des lignées cellulaires ou modèles animaux, suivie d'analyses transcriptomiques pour mesurer les conséquences fonctionnelles des modifications (Cano-Gamez & Trynka, 2020). Ces méthodes sont très lourdes à mettre en œuvre et demandent les expertises de laboratoires spécialistes.

E Conclusion

Les différentes approches d'analyse génétique, de l'analyse de liaison génétique aux analyses d'association génétiques pangénomiques (GWAS), permettent d'identifier des régions chromosomiques contenant des variants associés à un phénotype ou une maladie. Les premières études ont permis des avancées majeures, en particulier dans le contexte de maladies monogéniques ou au sein de familles fortement atteintes. En ce qui concerne les maladies complexes, les GWAS se sont imposées comme une approche de référence, notamment en raison de leur efficacité accrue dans des populations génétiquement structurées telles que les races canines.

Cependant, les variants identifiés sont souvent en déséquilibre de liaison avec le variant causal réel. En fonction de l'histoire génétique de la race ou population étudiée, cette région chromosomique (locus) sera plus ou moins grande (en moyenne quelques mégabases), et contiendra plusieurs gènes et éléments génomiques codants ou non, qui eux, contiendront le variant recherché. Il faut donc rester prudent dans le rôle de ces variants identifiés, qui sont très rarement les variants causaux ou prédisposants et qui peuvent aussi correspondre à des artefacts liés à la structure de population.

Dans ce contexte, la validation fonctionnelle constitue une étape clé pour affiner l'interprétation des résultats.

Elle peut inclure plusieurs niveaux d'analyse, de complexité croissante :

- *in silico* (annotation, colocalisation, enrichissement fonctionnel) ;
- *in vitro* (profil d'expression, tests cellulaires) ;
- *in vivo* (modèles animaux, édition génique).

La mise en œuvre d'expériences de validation fonctionnelle étant complexe et souvent difficilement réalisable, il est essentiel de valider autant que possible les loci et marqueurs identifiés grâce aux analyses statistiques les plus robustes, notamment, lorsque cela est possible, par la réplication des analyses génétiques sur une cohorte indépendante. Ces validations génétiques, statistiques et fonctionnelles représentent des étapes majeures pour confirmer l'implication des variants identifiés dans les traits étudiés et leur impact biologique.

Ainsi, c'est par l'intégration cohérente des approches statistiques, génétiques, biologiques et fonctionnelles que l'on peut passer de la détection d'associations génétiques entre des marqueurs d'un locus avec le trait étudié, à une compréhension mécanistique et biologique des bases génétiques d'un phénotype.

AVANTAGES DU MODÈLE CANIN DANS L'ÉTUDE DES MALADIES GÉNÉTIQUES COMPLEXES

A Aspects pratiques

Le chien domestique, en particulier les chiens de compagnie, présente de nombreux avantages pour l'étude des maladies génétiques complexes. Deuxième espèce la plus étudiée en pratique médicale après l'humain (Lindblad-Toh et al., 2005), il bénéficie d'une médecine vétérinaire en plein essor, structurée autour d'un réseau dense de spécialistes et d'une adaptation continue des pratiques, en parallèle de la médecine humaine (Mobasher, 2015 ; Sykes, 2024).

De surcroît, d'importantes ressources sont disponibles au niveau français et au niveau européen :

- des bases généalogiques exhaustives en France (la SCC LOF select (2025)) et dans les autres pays européens, des données génétiques multiples et récemment des données pangénomiques (Fernandes et al., 2016 ; Dyer et al., 2025) ;
- des collections biologiques (tissus, ADN) dans les pays européens ayant contribué au projet européen LUPA (2008-2012) (Lequarré et al., 2011) ;
- des Centres de Ressources Biologiques (CRB) ont été développée en Suisse, Finlande et Suède, en Grande-Bretagne et en France. Dès les années 2000, Catherine André à l'IGDR avait anticipé ce besoin, en créant le CRB Cani-DNA, qui comporte actuellement plus de 33 000 échantillons de sang et plus de 7 000 échantillons de tissus sur environ 300 races, représentant plus de 120 maladies (André et al., 2022).

Contrairement à l'humain, où la taille des familles est limitée et où la collecte d'informations généalogiques sur plusieurs générations est difficile, les chiens de race bénéficient d'une traçabilité généalogique précise grâce à des registres nationaux comme le Livre des Origines Français (LOF). Ces bases de données, aujourd'hui accessibles numériquement, contiennent des informations sur l'ascendance de milliers d'individus,

ainsi que sur le statut clinique pour de nombreuses maladies (e.g., dysplasie coxo-fémorale, atrophie progressive de la rétine, cataracte, hyperuricosurie) (LOF select, 2025).

La disponibilité de ces données généalogiques et de données phénotypiques, via les vétérinaires, permet d'estimer des paramètres génétiques tels que l'hérédité, les modes de transmission, leur éventuelle localisation autosomique ou liée au sexe, ainsi que de formuler les hypothèses de maladies, traits monogéniques ou plus complexes.

La richesse des informations généalogiques et phénotypiques disponibles chez le chien est étroitement liée à l'organisation de l'élevage canin en races distinctes. Chaque race constitue une population génétiquement isolée, fondée sur la sélection d'individus répondant à des critères morphologiques (e.g., la brachycéphalie, les variations de couleur de robe) ou comportementaux (e.g., aptitudes à la chasse ou au travail de troupeau), définis par des standards de race établis par des institutions cynophiles telles que la Société Centrale Canine (SCC) en France et les différents *kennels* en Europe ou l'AKC (*American Kennel Club*) aux États Unis.

Une autre caractéristique intéressante dans l'étude des maladies génétiques de l'espèce canine est le grand nombre de chiots par portée, avec une moyenne de 3,5 chiots pour les races de petite taille et de 7,1 pour celles de grande taille (Borge et al., 2011). Cela permet d'avoir accès plus facilement à un grand nombre d'individus par génération et facilite les études génétiques de liaison.

Enfin, l'espérance de vie des chiens est relativement courte, environ 10 ans pour les grandes races et jusqu'à 15 ans pour les petites. De ce fait des maladies liées à l'âge tel que les cancers, se manifestent plus précocement et sur un intervalle de temps plus restreint. Les traitements sont donc prodigués plus rapidement et par conséquent un essai clinique complet, qui demanderait entre cinq et quinze ans chez l'humain peut se dérouler entre un et trois ans chez le chien (Hansen & Khanna, 2004). Cette espérance de vie réduite offre également la possibilité d'observer plusieurs générations sur une période restreinte, facilitant ainsi les études longitudinales et l'analyse de la transmission héréditaire de caractères simples ou complexes.

B Aspects génomiques

1 Prédispositions par races

a Effets fondateurs et co-ségrégation d'allèles

La création des races canines repose sur une sélection intense et dirigée, rendue possible par l'utilisation de reproducteurs populaires et d'accouplements consanguins répétés. Cela se traduit par la contribution génétique disproportionnée de quelques individus, au patrimoine génétique de la race, favorisant l'émergence de lignées très homogènes et de dérives génétiques. Ces pratiques génèrent par la même occasion des effets fondateurs : une mutation délétère survenue chez un seul fondateur peut dès lors se diffuser largement, entraînant une forte prévalence de maladies héréditaires au sein de la population (Giger et al., 2006). En outre, les épisodes successifs de goulets d'étranglement lors de la formation des races induisent une réduction progressive de la diversité génétique intrarace.

En recherchant l'homogénéisation des phénotypes et des comportements au sein de chacune des races, l'humain a sélectionné les allèles impliqués ou responsables des traits désirés mais, ce faisant, a involontairement cosélectionné des allèles ou des combinaisons inadéquates d'allèles impliqués dans des maladies génétiques, soit par l'effet de ces allèles eux-mêmes, soit par effet de position (proches d'un allèle responsable d'un trait souhaité).

Un exemple emblématique est celui de la robe *merle* chez le Berger Australien, qui résulte d'une insertion d'une séquence répétée dans le gène *PMEL17* ou *SILV* (Hédan et al., 2006 ; Murphy et al., 2018). Cette insertion modifie la pigmentation et produit le motif de pelage recherché, mais elle peut également, lorsqu'elle est présente à l'état homozygote, entraîner des troubles sévères du développement, dont des affections oculaires et la surdité congénitale (Strain et al., 2009). Ce cas illustre comment la sélection de caractères esthétiques peut conduire à la fixation de mutations pathogènes, particulièrement dans un contexte de diversité génétique restreinte.

b Prévalence élevée des maladies génétiques

La structuration génétique des races canines se traduit pour chaque race par une prédisposition marquée à développer une ou plusieurs maladies génétiques. Ainsi, de nombreuses affections mendéliennes (dites « simples ») sont spécifiques d'une race ou d'un petit groupe de races. La mutation causale est alors identique chez tous les individus d'une ou de plusieurs races partageant un ancêtre commun, tandis que la même affection observée chez une autre race peut résulter d'une variation génétique différente.

Chaque année, la base de données *OMIA* enrichit son répertoire d'environ cinq à dix nouvelles affections, couvrant un large spectre allant des cancers et malformations cardiaques aux maladies ophtalmiques, neurologiques et orthopédiques (Ostrander et al., 2000 ; Patterson, 2000).

Ces spécificités raciales se manifestent par des affections emblématiques : chez le Border Collie, par exemple, on observe fréquemment surdité (De Risio et al., 2011), épilepsie (Hülsmeier et al., 2010) ou atrophie progressive de la rétine (Bunel et al., 2019), tandis que d'autres races, comme les retrievers et le Cane Corso, présentent respectivement 11 % et 60 % de prévalence de dysplasie de la hanche (Genevois et al., 2008 ; King, 2017). D'autres encore, comme les Boxer, les Bouvier Bernois et les retrievers seront très touchés par différents cancers, le Bouvier Bernois étant particulièrement atteint de sarcome histiocytaire. Ces taux, bien supérieurs aux fréquences observées chez l'humain (1/2 000 pour « fréquente », 1/500 pour « très fréquente »), facilitent le recrutement de sujets malades et l'obtention de prélèvements, mais rendent parfois plus délicate la constitution d'un groupe témoin indemne.

2 Mode de transmission simplifié ?

L'espèce canine se distingue également par un mode de transmission des maladies monogéniques majoritairement autosomique récessif, là où les affections monogéniques humaines relèvent le plus souvent de formes autosomiques dominantes ou liées au chromosome X. Cette singularité peut s'expliquer par les pratiques d'élevage : lorsqu'une maladie à expression dominante ou liée à l'X se manifeste précocement, les reproducteurs atteints sont rapidement écartés, réduisant fortement la propagation de ces allèles dans la population. À l'inverse, les allèles récessifs, souvent silencieux chez les hétérozygotes, échappent à cette sélection directe.

Or, les stratégies d'élevage, reposant fréquemment sur l'utilisation intensive de quelques individus champions et une forte consanguinité, favorisent la mise en homozygotie de ces mutations délétères. En conséquence, des variants rares mais pathogènes chez l'humain peuvent devenir visibles cliniquement chez le chien. Ce phénomène rend le chien particulièrement précieux en médecine comparée, car il permet de révéler des gènes impliqués dans des maladies génétiques humaines restées jusqu'alors inaccessibles par les approches classiques (Giger et al., 2006).

Cependant, la structure génomique des races, caractérisée par de longs blocs haplotypiques, peut masquer la complexité réelle des maladies héréditaires. Une affection qui semble suivre un mode de transmission monogénique simple, dit mendélienne, peut en fait résulter de la participation de plusieurs gènes. Lorsqu'un petit nombre de gènes, souvent deux à cinq, contribue de manière significative au phénotype, on parle de maladies oligogéniques, comme les cancers du sein chez les femmes porteuses

de mutations *BRCA1* et *BRCA2* (Badano & Katsanis, 2002) ou la dystrophie de la rétine chez le chien (Ripolles-Garcia et al., 2023). Dans les cas où des dizaines ou des centaines de variants à effet modeste s'additionnent pour déterminer la susceptibilité, on parle de maladies polygéniques ; c'est notamment le cas de beaucoup de cancers chez l'humain ou de maladies auto-immunes ou de développement, comme les dysplasies de la hanche chez l'humain et le chien (King, 2017 ; Harsanyi et al., 2020a ; Harsanyi et al., 2020b), où de multiples loci répartis sur le génome agissent ensemble, voire interagissent avec des facteurs environnementaux afin de déclencher l'affection. Cette complexité nécessite, lors des études d'association, non seulement de repérer les grands haplotypes, mais aussi de pouvoir modéliser les interactions géniques et gène-environnement afin de distinguer l'effet des variants majeurs de celui des nombreux variants mineurs.

3 Développer un test génétique de risque

Connaître la/les mutation(s) responsable(s) d'une maladie génétique chez le chien présente de multiples intérêts. En plus de proposer des gènes candidats pour les maladies correspondantes humaines, ces découvertes servent en premier lieu à aider la médecine vétérinaire et l'élevage. Le développement de tests génétiques à vue diagnostique et de sélection est l'un des objectifs des études génétiques canines. Le test diagnostique permet de confirmer la composante génétique chez un chien atteint et peut permettre d'orienter le traitement.

De plus, dans le cas d'une maladie monogénique récessive, le test permet d'identifier les chiens hétérozygotes pour la mutation et dans ce cas, éviter de marier deux chiens hétérozygotes (dits porteurs sains) qui produiront statistiquement $1/4$ de chiens atteints dans leurs portées et la moitié de chiens porteurs sains du variant. De plus, dans des maladies à déclaration tardive, détecter tôt (avant la reproduction) les chiens atteints de la maladie et donc homozygotes pour l'allèle muté, qui transmettront nécessairement un allèle muté à la génération suivante, est très utile pour l'élevage. Il faut toutefois faire attention de ne pas éliminer l'ensemble des chiens portant le variant à l'état hétérozygote, de la reproduction et ne sélectionner que les chiens homozygotes pour l'allèle de référence, ce qui risquerait de diminuer la diversité génétique et mener à l'apparition de nouvelles maladies.

C Maladies homologues entre humain et chien

Parmi les 977 phénotypes canins répertoriés dans la base de données *Online Mendelian Inheritance in Animals* (OMIA, 2025), 629 présentent des similarités cliniques notables avec des maladies humaines homologues. L'espèce canine se place ainsi loin devant les autres espèces domestiques en termes de correspondances identifiées : 355 sur 702 pour les bovins, 286 sur 454 pour les chats, 227 sur 406 chez les porcs, 166 sur 307 chez les chevaux, 145 sur 333 chez les moutons, 91 sur 272 chez les poules, 89 sur 149 pour les lapins et 60 sur 129 pour les chèvres.

Cette prévalence élevée de maladies homologues chez le chien s'explique en partie par la proximité génétique avec l'humain, mais aussi par une médecine vétérinaire plus proche de la médecine humaine et par le fait que nous partageons le même environnement, impactant similairement nos fonctions biologiques. Des études de génomique comparative ont en effet montré une forte conservation de la structure, de l'organisation et du contenu des génomes entre les deux espèces (Guyon et al., 2003 ; Lindblad-Toh et al., 2005 ; Derrien, 2007). À cela s'ajoute le partage d'un même environnement domestique, favorisant l'émergence de phénotypes pathologiques similaires, notamment dans les domaines de la cancérologie, de la neurologie ou de l'immunologie.

Le chien s'est ainsi imposé comme un modèle naturel ou dit « spontané », pertinent pour de nombreuses affections humaines, en raison de similitudes cliniques, mais aussi par l'évolution naturelle des maladies, les méthodes diagnostiques comparables et parfois même les options thérapeutiques partagées. Ces parallèles ont été observés dans des domaines aussi variés que les cancers (Gardner et al., 2016 ; Khanna et al., 2006), les troubles neurologiques tels que l'épilepsie (Löscher, 2022), les troubles du comportement et psychiatriques (Overall, 2000), les maladies liées au vieillissement (Mazzatenta et al., 2017), les affections cardiovasculaires (Camacho et al., 2016), oculaires (Bunel et al., 2019), dermatologiques (Plassais et al., 2015) ou encore les affections orthopédiques (Riser, 1975 ; Meeson et al., 2019 ; Binversie et al., 2022b).

Fait remarquable et d'ailleurs objectif des équipes travaillant avec le « modèle chien » dans le monde, certaines mutations identifiées initialement chez le chien ont ensuite permis de guider la découverte de variants causaux dans les mêmes gènes chez l'humain, illustrant la puissance du modèle canin en recherche translationnelle (Karlsson & Lindblad-Toh, 2008). C'est notamment le cas de la narcolepsie associée au gène *HCRTR-2* (Hypocrétine) ; de la rétinopathie pigmentaire impliquant le gène *PRCD* (Tsai et al., 2007) et beaucoup d'autres gènes (Bunel et al., 2019) ; l'ichtyose (Grall et al., 2012) ; ou encore du cancer rénal canin (RCND), qui a permis d'orienter les recherches sur des altérations comparables dans le cancer rénal nodulaire (Ostrander et al., 2008).

D Exemple d'une maladie « simple » : l'ichtyose du Golden Retriever

L'ichtyose du Golden Retriever est une maladie dermatologique d'accumulation de pellicules sur l'épiderme, caractérisée par un défaut de différenciation des kératinocytes et une desquamation excessive de la peau. Dès 2009, une analyse de pedigree avait suggéré un mode de transmission autosomique récessif et encouragé l'équipe « Génétique du Chien » à collecter des chiens atteints et des chiens indemnes afin de constituer des groupes cas - contrôles et des familles en vue d'analyses génétiques. Ces travaux ont permis à l'équipe de rechercher puis de caractériser la mutation responsable : une insertion-délétion dans l'exon huit du gène *PNPLA1* (*patatin-like phospholipase domain containing 1*), qui code une enzyme essentielle au métabolisme des lipides cutanés. Cette *InDel* induit un codon stop prématuré, entraînant la synthèse d'une protéine PNPLA1 tronquée et non fonctionnelle, induisant une anomalie de la différenciation des kératinocytes (Guaguere et al., 2009 ; Grall et al., 2012).

La découverte du gène *PNPLA1* comme gène causal chez le Golden Retriever, avec une mutation présente chez tous les chiens atteints, a rapidement permis à l'équipe de contacter une équipe spécialiste des ichtyoses en médecine humaine et d'identifier de fait, des mutations de ce même gène dans des familles de patients atteints d'ARCI (*Autosomal Recessive Congenital Ichthyosis*). Grall et al ont identifié deux mutations distinctes de *PNPLA1* chez les patients atteints de l'ichtyose récessive de deux familles distinctes (Grall et al., 2012). Depuis cette découverte, ce ne sont pas moins de 59 mutations de *PNPLA1* qui ont pu être mises en évidence dans ce gène pour de nouveaux patients atteints d'ARCI (Zeng et al., 2022). Par ailleurs, la fonction même de la protéine PNPLA1, jusqu'alors inconnue, a été clarifiée par ces travaux, révélant son rôle critique dans l'homéostasie de la desquamation, dans la barrière cutanée.

Sur le plan de la valorisation de ces résultats, la mutation canine a fait l'objet d'un dépôt de brevet (EP2501825A1) international aux noms du CNRS et de l'Université de Rennes (André et al., 2014) et a conduit au développement, dès 2010, d'un test génétique commercialisé sous licence par le laboratoire *Antagene*, permettant le dépistage des porteurs et la connaissance des chiens atteints (homozygotes mutés) et des chiens indemnes (homozygotes allèle sauvage) dans les programmes d'élevage.

Cet exemple montre la puissance du modèle naturel de chiens de races touchées par une maladie homologue à l'humain, permettant non seulement de développer un test génétique de diagnostic pour la médecine vétérinaire et de dépistage pour l'élevage et tout en même temps de démontrer l'implication ou la causalité de mutations dans la maladie homologue humaine, avec, là aussi, la mise en place d'un test génétique en routine via les panels de diagnostic des ichtyoses humaines.

LA DYSPLASIE COXO-FÉMORALE, UNE MALADIE « COMPLEXE »

Une revue détaillée de la littérature portant sur les aspects anatomiques, diagnostiques et génétiques de la dysplasie de la hanche est en cours de rédaction par notre équipe, en collaboration avec des vétérinaires, un médecin orthopédiste et des généticiens. Elle est présentée dans le chapitre I « Étude bibliographique comparative » des résultats. Les paragraphes suivants reprennent les éléments essentiels nécessaires à la compréhension de cette affection chez le chien, mais aussi chez l'humain.

La dysplasie de la hanche est une maladie complexe résultant de l'action conjointe de nombreux gènes participant au développement de cette articulation, ainsi que des interactions entre ces gènes et l'environnement. Elle est décrite comme un développement anormal de l'articulation sphéroïde synoviale de la hanche. La coaptation entre l'acétabulum, la cavité articulaire du bassin qui reçoit la tête fémorale, et cette dernière ne se faisant plus normalement, un profil arthrosique apparaît systématiquement avec l'âge chez le chien, entraînant douleur et troubles de la mobilité.

A Description anatomique

En tant qu'articulation synoviale sphéroïde, l'articulation coxo-fémorale est composée d'un grand nombre de tissus (voir figure 7). Elle fait ainsi intervenir deux os : le fémur et le bassin avec un point de contact entre la tête fémorale et l'acétabulum. Tous deux sont recouverts d'un cartilage, de manière quasi intégrale pour la tête fémorale et seulement sur la partie périphérique de l'acétabulum poursuivie par un bourrelet cotyloïdien situé sur le limbus acétabulaire. Ces structures sont alors recouvertes par une structure fibreuse de contention appelée capsule articulaire et baignent dans le liquide synovial sécrété par les synoviocytes de la membrane synoviale. Enfin les ligaments ilio-fémoral, pubo-fémoral et ischio-fémoral entourent la capsule et participent au maintien de l'articulation tandis que le ligament rond (*teres ligament*) possède un rôle de porte vaisseaux pour irriguer la tête fémorale chez les deux espèces qui est complété par un rôle de maintien mécanique chez le chien (Riser & Shirer, 1966). Enfin ce ne sont pas moins de 27 muscles chez l'humain qui permettent à la hanche de faire l'ensemble des mouvements sur les trois axes : flexion et extension sur le plan sagittal, adduction et abduction sur le plan frontal et enfin rotation interne et externe sur le plan horizontal.

B Impact de la bipédie

Le passage de la quadrupédie à la bipédie a profondément transformé l'anatomie de l'articulation de la hanche chez l'humain par rapport au chien, notamment par une réorientation du bassin, l'apparition d'un valgus du genou et l'alignement vertical des membres inférieurs, optimisant la biomécanique. Ces changements ont aussi modifié la musculature pelvienne, avec une réorganisation du trajet de l'ilio-psoas et une évolution des rôles des muscles glutéaux, favorisant la stabilité posturale et l'efficacité de la marche bipède. La bipédie a aussi modifié les allures humaines, réduisant le nombre d'appuis au sol et augmentant les contraintes sur l'articulation de la hanche. Cette évolution a entraîné une redistribution des charges, une réduction des surfaces articulaires en contact et une augmentation des forces exercées, influencées par la posture, la masse corporelle, la vitesse de déplacement ou la présence d'affections, avec des implications biomécaniques majeures. L'humain étant le seul primate strictement bipède, aucune autre espèce animale ne peut reproduire à l'identique l'ensemble des contraintes anatomiques et biomécaniques associées à cette posture. Malgré cela l'articulation de la hanche canine et humaine montre une grande similarité. Bien que leurs formes et les charges fonctionnelles diffèrent, les deux partagent une architecture comparable, incluant les mêmes structures osseuses, musculaires et un réseau vasculaire analogue (Pascual-Garrido et al., 2018).

C Pathogénèse et physiopathologie de la dysplasie de la hanche

En temps normal, la tête fémorale a une forme sphérique assurant une bonne congruence articulaire. En cas d'atteinte dysplasique, elle devient ellipsoïde, entraînant une mauvaise congruence et des mouvements arthrogènes. Cette affection s'accompagne d'une hypoplasie du toit acétabulaire, réduisant la surface de contact et ainsi la couverture antérieure de la tête fémorale et pouvant provoquer des douleurs inguinales et antérieures. Chez l'humain, le spectre d'altération s'étale ainsi de dysplasie, hanche luxable, subluxation, luxation réductible à luxation non-opérable.

La pathogénèse de ces anomalies du développement englobe de nombreux mécanismes survenant durant une fenêtre critique du développement et de la croissance. Il peut s'agir d'un retard dans l'innervation musculaire ou bien d'anomalies survenant lors du développement du bourgeon du membre. Ces altérations sont alors susceptibles de compromettre la forme, l'orientation et la profondeur de l'acétabulum et de la tête fémorale. Ce déséquilibre initial est à l'origine d'une mauvaise congruence articulaire et de remaniements secondaires (Brand, 2008).

Le rôle de la laxité de la hanche est un point de divergence notable entre les deux espèces. Chez le chien, elle est aujourd'hui reconnue comme le facteur déclencheur

principal. Elle apparaît précocement dans la croissance et persiste suffisamment longtemps pour permettre une subluxation articulaire, entraînant des contraintes mécaniques délétères et une cascade de remaniements dégénératifs. Chez l'humain, la laxité était historiquement considérée comme le facteur central de la DDH (Dysplasie Développementale de la Hanche). Toutefois, les connaissances actuelles la relèguent à un facteur contributif parmi d'autres, notamment chez les nouveau-nés où elle tend à se résoudre spontanément. La dysplasie serait plutôt liée à un défaut intrinsèque de développement de l'acétabulum et de la tête fémorale.

Dans les deux cas, la mauvaise congruence articulaire induit une cascade de déformations secondaires (voir figure 7) :

- hypoplasie et aplatissement de la tête fémorale ;
- pente acétabulaire accrue ;
- épaissement de la capsule articulaire et déformation du labrum ;
- accumulation de tissus intra-articulaires anormaux (pulvinar chez l'humain) ;
- altérations mécaniques accentuant la subluxation ou la luxation.

Hanche saine

Hanche atteinte

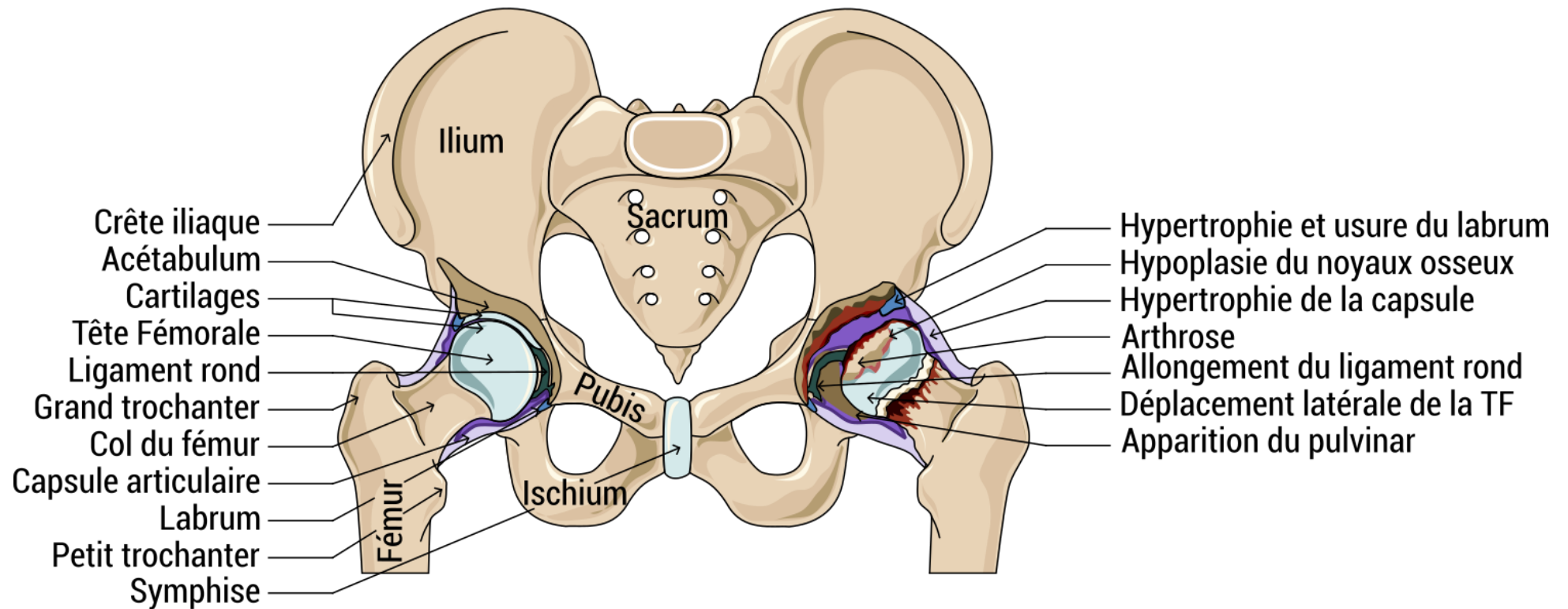


FIGURE 7 – Représentation anatomique d'une hanche humaine normale et d'une hanche atteinte de dysplasie de la hanche.

TF : Tête Fémorale.

D Similarités et différences entre la dysplasie de la hanche chez le chien et chez l'humain

Le lien avec l'arthrose est différent entre les deux espèces. Chez le chien, l'arthrose est une conséquence fréquente et bien établie de la DCF, apparaissant relativement tôt et étant directement corrélée à la sévérité de l'incongruence articulaire ; elle varie cependant en fonction des races. Chez l'humain en revanche, le lien entre dysplasie de hanche et arthrose est plus complexe et débattu. Bien que certaines formes de dysplasie soient associées à un risque accru d'arthrose à long terme, cette relation est moins systématique, notamment grâce à une prise en charge précoce dans de nombreux cas (Vaughan-Scott & Taylor, 1997 ; Runge et al., 2010).

Enfin, les deux espèces partagent une fenêtre critique post-natale durant laquelle la croissance neuromusculaire et squelettique doit être coordonnée. Chez le chien, cette période se situe entre la naissance et 60 jours, moment où la laxité peut compromettre la congruence articulaire. Chez l'humain, cette période est plus étendue, mais les premières semaines de vie restent cruciales, comme en témoignent les efforts de dépistage néonatal par les tests d'Ortolani et de Barlow, dans les premières semaines de la vie (Alexander, 1992 ; King, 2017).

Chez le chien comme chez l'humain le diagnostic s'établit sur la présence conjointe d'examens cliniques et d'imagerie. Les praticiens partagent des techniques similaires aussi bien sur l'examen physique (e.g., manœuvre de Barlow, signe d'Ortolani) que sur le dépistage par imagerie (e.g. équivalence de l'angle de Norberg Olsson chez le chien et de l'angle VCE chez l'humain). Cependant chez le chien la méthode de dépistage de référence est la radiographie à un an en vue ventro-dorsale avec hanche étendue ou bien en vue en contrainte avec un dispositif de distraction, tandis que chez l'humain c'est l'échographie à un mois qui fait référence bien que les coupes et mesures soient variables entre les pays. Chez les deux espèces, des scores composites ont été mis en place afin de classifier le degré de sévérité de l'atteinte. Chez le chien, la DCF est évaluée sur une échelle de A à E créée par la Fédération Cynologique Internationale (FCI), cette échelle prend en compte des mesures et indices radiographiques. Chez l'humain en revanche la méthode de Graff et ses variations, a été, et est encore, dans plusieurs pays, le diagnostic de référence pour la DDH. Cette classification prend en compte deux angles ainsi que des indices échographiques et est complété par les différents examens physiques réalisés lors de la consultation (Willemsen et al., 2021).

E Facteurs de risque environnementaux de la dysplasie de la hanche

Du fait de la complexité du développement de l'articulation et du nombre important de structures anatomiques impliquées, il n'est pas étonnant que le développement anormal de cette articulation soit influencé par un grand nombre de gènes du développement et par des facteurs environnementaux. Ainsi, la nutrition, le poids, l'exercice au jeune âge mais aussi la saison de naissance et le mode de parturition (i.e., présentation par le siège ou césarienne) ont un impact plus ou moins direct sur le développement de la dysplasie de la hanche chez l'humain et le chien (Pascual-Garrido et al., 2018).

Plusieurs études dans les années 1960 - 1980 ont évalué l'impact de l'œstradiol et plus récemment l'impact de la relaxine et de la stérilisation en fonction du sexe sur la pathogenèse de la DCF. Cependant les résultats divergent entre les études et seule la stérilisation trop précoce des chiots semble être confirmée pour le moment comme ayant un impact délétère sur les races de grandes tailles (Torres de la Riva et al., 2013).

F Aspects génétiques de la dysplasie de la hanche

Sur le plan génétique, de nombreuses études ont d'ores et déjà été réalisées aussi bien chez l'humain que chez le chien. L'héritabilité est prouvée depuis de nombreuses années aussi bien chez l'humain que chez le chien. La prévalence ainsi que l'héritabilité varient de manière très importante chez le chien, en fonction des races et des paramètres utilisés. La prévalence peut ainsi varier de 4 % chez les huskys à 60 % chez les Cane Corso tandis que l'héritabilité peut aller de 0,2 à 0,6 en fonction des races, du sexe et de la mesure étudiée (Genevois et al., 2008 ; Mikkola, 2020).

Malgré l'existence d'analyses génétiques dans les deux espèces, peu de loci associés à la DCF ont été confirmés de manière robuste par plusieurs études aussi bien chez l'humain que chez le chien. Cette faible reproductibilité n'est pas surprenante : les études ont été menées dans des contextes très hétérogènes, sur des populations de géographies et de structures génétiques différentes. Par ailleurs, le manque d'études de réplication à grande échelle sur des cohortes homogènes limite encore notre capacité à identifier des signaux communs robustes.

Il est toutefois possible de visualiser des loci communs entre humain et chien au niveau de la portion terminale du chromosome 2 et sur les chromosomes 19 et 20 humains, contenant les gènes *COL6A3*, *HOXB9* et *COL1A1*, *UQCC1* et *GDF5* respectivement. Enfin, chez le chien, plusieurs tests génétiques de risque ont été développés et mis en vente dans différentes races, puis tour à tour retirés du marché lorsqu'ils ont été invalidés par des études indépendantes. Le champ de recherche reste donc largement ouvert et constitue un véritable défi à la fois génétique et populationnel.

SCIENCE OUVERTE ET FAIREST

Ces difficultés à identifier des loci robustes, conjuguées à l'hétérogénéité des études et à la complexité génétique de la dysplasie de la hanche, soulignent la nécessité de disposer d'outils d'analyse bio-informatique transparents, reproductibles et adaptables. Dans le cadre de cette thèse, j'ai été amené à développer des outils informatiques permettant de visualiser et d'analyser efficacement les multiples données nécessaires à ce projet de recherche. Cette démarche m'a naturellement conduit à m'intéresser aux principes de la science ouverte, dans l'objectif de renforcer la fiabilité et la réutilisabilité des analyses produites.

"More than 70 % of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments."

C'est par ce constat que s'ouvre un article sur la crise de reproductibilité dans la recherche scientifique dans *Nature* (Baker, 2016). Cette problématique touche l'ensemble des disciplines, y compris particulièrement la bio-informatique. Malgré leurs publications dans des revues renommées, de nombreux outils bio-informatiques souffrent de défauts majeurs : accès restreint ou payant, erreurs d'installation, dépendances non documentées, instabilité liée aux virgules flottantes, absence de jeux de données de test, etc. Autant d'obstacles qui entravent leur utilisation et leur appropriation, freinent la progression scientifique et contribuent à une perte de confiance dans les résultats publiés.

Face à cette situation, plusieurs initiatives ont vu le jour ces dernières années, afin de promouvoir une recherche plus transparente, reproductible et accessible. Parmi celles-ci, les mouvements de la Science Ouverte (*Open Science*) et les principes FAIR (*Findable, Accessible, Interoperable, Reusable*) se sont imposés comme des piliers de bonnes pratiques pour la gestion, le partage et la valorisation des données et des outils scientifiques informatiques.

A L'Open Science

L'*Open Science* ou science ouverte, désigne l'ensemble des pratiques visant à rendre accessibles, de manière libre et gratuite, les données, les méthodes, les outils et les résultats issus de la recherche scientifique. Elle repose sur le principe directeur : « Aussi ouverte que possible, Aussi fermée que nécessaire. » soulignant la tension inhérente entre la transparence scientifique et les contraintes éthiques, juridiques ou stratégiques, ainsi que pour les chercheurs, l'adage bien connu "*publish or perish*"

L'ouverture des données constitue sans doute l'aspect le plus complexe de cette démarche. Leur collecte, souvent longue et coûteuse, représente un capital scientifique que les laboratoires peuvent hésiter à partager, notamment en contexte de compétition académique. Dans le domaine de la santé, les défis sont encore plus importants : les données de patients sont soumises au Règlement Général sur la Protection des Données (RGPD), qui impose des exigences strictes en matière de confidentialité. L'anonymisation ne suffit pas toujours à garantir la protection des individus, car le risque de réidentification doit être évalué à l'échelle de l'écosystème global des bases de données accessibles. À cela s'ajoutent des contraintes techniques, notamment la taille massive de certains jeux de données, en particulier en génomique, où les données issues des technologies de séquençage de nouvelle génération peuvent atteindre plusieurs téraoctets. Cette volumétrie soulève la question du périmètre des données à partager : faut-il fournir les données brutes, les données intermédiaires, les résultats finaux ? L'archivage de ces données soulève ainsi de nombreuses questions, tant sur le plan logistique que financier.

Des initiatives telles que ENA (*European Nucleotide Archive*), SRA (*Sequence Read Archive*) ou EGA (*European Genome-Phenome Archive*) permettent néanmoins de stocker et de partager ces données à grande échelle, à condition qu'elles soient accompagnées de métadonnées riches et normalisées, condition *sine qua non* de leur réutilisabilité.

La diffusion des méthodes d'analyse est généralement plus aisée. Des plateformes comme *GitLab* ou les dépôts institutionnels permettent de partager efficacement scripts, logiciels et pipelines. Toutefois, une certaine réticence demeure, notamment pour les scripts développés en interne, souvent considérés comme des avantages concurrentiels.

Enfin, les résultats scientifiques sont traditionnellement diffusés via des publications. Or, ces dernières restent souvent lacunaires en ce qui concerne les données sous-jacentes et l'accès libre aux articles (i.e., *open access*) n'est pas encore généralisé. Les coûts de publication en libre accès, ainsi que la perception d'un moindre prestige ou d'un facteur d'impact réduit, constituent des obstacles persistants à l'adoption massive de ces pratiques.

B Les principes FAIR

Les principes FAIR définissent un cadre de bonnes pratiques visant à optimiser la gestion et le partage des données, méthodes et outils en recherche scientifique. Ils sont particulièrement essentiels en bio-informatique, où ils s'appliquent à toutes les phases du cycle de vie d'un outil : de sa conception initiale, en passant par son développement continu, jusqu'à sa publication et son archivage (Wilkinson et al., 2016 ; Martín del Pico et al., 2024).

Les quatre piliers fondamentaux sont les suivants : *Findable*, *Accessible*, *Interoperable* et *Reusable*.

1 *Findable*

Les données, méthodes et scripts doivent être aisément localisables. Cela implique qu'ils soient systématiquement répertoriés et annotés avec des métadonnées complètes et normalisées. La recherche doit pouvoir s'effectuer aussi bien par des humains que par des machines, ce qui nécessite l'utilisation de langages contrôlés et de nomenclatures communes. Par exemple, l'anglais est privilégié comme langue de référence pour maximiser l'interopérabilité et des initiatives telles que l'ontologie *EDAM* (Black et al., 2022) facilitent une description normalisée et cohérente des données, formats, opérations et outils bio-informatiques. *EDAM* fournit un vocabulaire structuré permettant de mieux annoter les ressources, de favoriser leur réutilisation et de soutenir les principes FAIR.

2 *Accessible*

L'accès aux données doit être garanti sur le long terme, que ce soit en libre accès ou contrôlées. Des plateformes d'archivage telles que *GitLab*, *Software Heritage* ou *Zenodo* fournissent des solutions pérennes pour la conservation et la diffusion des ressources numériques.

3 *Interoperable*

Ce principe est avec le suivant l'un des plus complexes à mettre en œuvre. Concernant les données, il vise à garantir que leurs formats soient compatibles entre divers outils et systèmes. En génétique, des standards tels que VCF (*Variant Call Format*) ou SAM (*Sequence Alignment Map*) sont largement utilisés, avec différentes versions et variantes (e.g., compression, multi-échantillons, indexation, etc.). Les formats doivent aussi être ouverts et éviter les solutions propriétaires ou binaires fermées, afin de limiter les incompatibilités.

Du côté des logiciels, l'interopérabilité implique leur capacité à fonctionner sur différents environnements matériels et systèmes d'exploitation, malgré les disparités (e.g., entre

architectures *Intel* et *ARM64*). L'usage de conteneurs logiciels (comme *Docker* ou *Singularity*), bien que complexe à la mise en place, constitue une solution efficace pour « encapsuler » l'ensemble des dépendances et garantir un environnement d'exécution contrôlé et reproductible.

4 Reusable

L'objectif ultime des principes FAIR est de favoriser la réutilisation des ressources scientifiques par d'autres chercheurs. Cela nécessite des métadonnées exhaustives, une description rigoureuse des protocoles expérimentaux, des versions logicielles, des paramètres utilisés et du matériel employé. La crise de la reproductibilité est en partie liée à un défaut de cette documentation. Celle-ci est en effet chronophage à mettre en place et doit être régulièrement mise à jour afin de rester synchrone avec l'état d'avancement des outils. Enfin, l'attribution de licences claires et adaptées est essentielle pour définir les conditions d'utilisation et de redistribution des données et outils.

C L'extension FAIREST

L'extension FAIREST complète les principes FAIR en y ajoutant trois dimensions clés particulièrement pertinentes dans le contexte de la bio-informatique. Différentes variantes existent mais se recoupent (d'Aquin et al., 2023).

1 Engagement & Ethique / Equitable

Au-delà de la simple accessibilité des produits de la recherche, ce principe met l'accent sur la facilité d'utilisation et l'adhésion des utilisateurs. Les outils doivent être conçus non seulement pour être accessibles, mais aussi pour être faciles à utiliser. Ils doivent s'adapter aux méthodes de travail et aux habitudes de ceux qui les utilisent. Enfin, ces outils doivent être faciles à maintenir par les développeurs et ces derniers doivent s'engager sur le long-terme sur cette maintenance.

Ce volet met également en lumière la question de l'équité dans les données. Il est essentiel que les jeux de données incluent des populations diverses, humaines comme animales, pour éviter les biais systémiques dans les résultats ou les prédictions. Dans de nombreux outils bio-informatiques actuels, certains paramètres sont encore codés de manière rigide (i.e., *hard-coded*, c'est-à-dire fixés dans le code sans possibilité d'adaptation), ce qui limite leur usage dans des contextes biologiques différents de ceux pour lesquels ils ont été conçus. Favoriser l'inclusivité, la flexibilité et l'adaptabilité des outils et données est ainsi fondamental pour garantir une science ouverte réellement équitable, durable et éthique.

2 Scalable

La capacité d'un outil à gérer efficacement une augmentation substantielle des volumes de données, sans dégradation excessive de ses performances, constitue un enjeu central en bio-informatique. En effet, selon la complexité algorithmique sous-jacente, l'accroissement de la taille des jeux de données peut entraîner une croissance linéaire, quadratique, voire exponentielle des temps de calcul, compromettant l'utilisabilité des outils à grande échelle. Il est donc essentiel d'intégrer, dès la phase de conception, des stratégies de mise à l'échelle (i.e., scalabilité) et d'optimisation algorithmique, notamment en matière de parallélisation, de gestion mémoire ou de compression des données. Cette anticipation permet de garantir la robustesse et la durabilité des outils face à l'explosion des données en sciences omiques.

3 Social

Les produits de la recherche sont le fruit d'un processus collaboratif, mobilisant une diversité d'acteurs, d'expertises et d'interactions humaines. Il est donc essentiel que les dépôts et plateformes de partage reflètent cette dimension sociale en valorisant les contributions individuelles et collectives. Cela implique de documenter clairement les rôles des contributeurs (e.g., auteurs, développeurs, relecteurs, curateurs de données, etc.) et de favoriser des modes de collaboration ouverts, transparents et traçables. L'usage de plateformes collaboratives telles que *GitLab*, *GitHub*, *Zulip* ou *Slack*, associé à des systèmes de gestion de versions et à des standards de traçabilité des contributions (e.g., via le standard *CRedit*), permet ainsi de garantir une reconnaissance claire, équitable et durable du travail de chacun.

4 Tested & Trust

Le principe *Tested* souligne l'importance de la robustesse et de la fiabilité des outils et données scientifiques. Pour garantir ces qualités, les logiciels doivent être soumis à des tests unitaires, à des validations croisées et à des protocoles d'intégration continue. Ces mécanismes permettent de détecter précocement les erreurs, de stabiliser les versions et de s'assurer de la reproductibilité des résultats, notamment par le paramétrage explicite des générateurs de nombres aléatoires (e.g., gestion des *seeds*). Cela garantit que les analyses peuvent être refaites dans des conditions identiques, condition indispensable à la rigueur scientifique.

Le concept *Trust*, complémentaire, met l'accent sur la confiance des utilisateurs dans les dépôts et outils mis à disposition. Cette confiance repose sur plusieurs piliers : la qualité démontrée des outils via validation indépendante ou relecture par les pairs, la traçabilité des versions, la transparence des protocoles et le respect des bonnes

pratiques en science ouverte. La pérennité des données joue également un rôle crucial : il est indispensable que les dépôts mettent en œuvre des politiques robustes de conservation à long terme, incluant des sauvegardes régulières, des formats durables et des identifiants persistants (e.g., DOI).

En parallèle, des mécanismes de gouvernance clairs doivent encadrer l'accès aux données, notamment lorsque celles-ci comportent des informations sensibles (e.g., génétiques, médicales, propriétaires). Cela implique de définir des modalités d'accès différenciées, avec la possibilité d'imposer des embargos temporaires, de restreindre certains usages ou de conditionner l'accès à des engagements éthiques (e.g., via des comités d'accès aux données). Ces garde-fous assurent que la diffusion des données reste compatible avec le respect de la confidentialité, de la propriété intellectuelle et des normes éthiques en vigueur.

D Inconvénients

La science ouverte et la mise en œuvre des principes FAIREST présentent plusieurs défis et limites. La gestion et l'annotation rigoureuse des données demandent un investissement important en temps, compétences et ressources, ce qui peut représenter un frein pour certains laboratoires, en particulier ceux disposant de moyens limités. De plus, l'ouverture des données soulève des enjeux éthiques et légaux, notamment en ce qui concerne la protection des données personnelles, la confidentialité des données de patients et le respect des consentements (RGPD). La complexité technique liée à la standardisation et à l'interopérabilité des formats peut rendre leur adoption difficile, surtout lorsque les outils ou infrastructures existants ne sont pas compatibles ou nécessitent des adaptations coûteuses. Enfin, la pression pour publier rapidement des données ouvertes peut parfois se traduire par une qualité variable des métadonnées et des jeux de données, compromettant la fiabilité et la réutilisabilité des ressources. Ces obstacles nécessitent une organisation et un soutien institutionnel forts pour garantir un équilibre entre ouverture, qualité, sécurité et respect des cadres réglementaires.

E Avantages

Les avantages de la science ouverte et des principes FAIREST sont multiples et essentiels pour la recherche moderne. En facilitant une gestion rigoureuse et standardisée des jeux de données, souvent volumineux et complexes, ces approches améliorent considérablement leur accessibilité, leur traçabilité et leur exploitation. Un bon référencement et une annotation précise permettent non seulement de gagner du temps dans la recherche et l'utilisation des données, mais aussi d'assurer une meilleure transparence et reproductibilité des analyses scientifiques. En favorisant la réutilisation des ressources entre différentes études, elles réduisent les redondances, évitent les erreurs

déjà identifiées, maximisent l'impact des travaux réalisés et permet ainsi de gagner du temps et de l'argent. Ces principes encouragent aussi la collaboration, l'innovation et la diffusion rapide des connaissances, contribuant ainsi à accélérer les progrès scientifiques tout en renforçant leur fiabilité.

F Les communautés *Bioconductor* et *nf-core*

Les travaux présentés dans la suite de cette thèse ont été publiés au sein de deux communautés : *Bioconductor* pour le package *Pedixplorer* et *nf-core* pour le pipeline *nf-core/phaseimpute*. Ces deux communautés (voir logos en figure 8) sont des exemples phares d'initiatives de science ouverte et FAIR qui favorisent le développement, le partage et la maintenance d'outils bio-informatiques robustes, reproductibles et bien documentés.

1 *Bioconductor*

Bioconductor est une plateforme *open-source* et *open-development* dédié à l'analyse des données biologiques et particulièrement celles liés à la génomique. Fondé en 2001, elle repose sur le langage statistique *R* et regroupe plus de 3 600 *packages* développés par une communauté scientifique internationale en génomique. Son objectif est de fournir un accès large à des méthodes statistiques et graphiques puissantes pour l'analyse de données génomiques, tout en facilitant l'intégration de métadonnées biologiques issues de bases de données telles que *PubMed* ou *Entrez Genes*. *Bioconductor* encourage la création de package extensibles et interopérables. Ainsi le paradigme orienté objet de *R* (Team, 2025) est mis en avant par une structuration importante et commune entre les *packages* par le biais de classes *S4*. *Bioconductor* met de même l'accent sur la reproductibilité et une documentation de qualité. Chaque package inclut ainsi des vignettes détaillant ses fonctionnalités et usages, souvent accompagnées de tutoriels pratiques. Chaque package est soumis à une revue détaillée et fait l'objet de tests rigoureux, incluant un processus d'évaluation automatisé combinant les outils *rcmdcheck* et *BiocCheck* pour garantir la compatibilité et la robustesse sur différentes plateformes. Le projet promeut également une approche de développement ouvert, invitant les utilisateurs à devenir contributeurs en développant des *packages* ou en améliorant la documentation (Huber et al., 2015).

2 *nf-core*

La communauté *nf-core* incarne une approche collaborative, ouverte et rigoureusement standardisée du développement de pipelines d'analyse en bio-informatique. Construits autour du moteur de workflow *Nextflow*, les pipelines *nf-core* sont conçus pour être portables, interopérables et reproductibles, indépendamment de l'infrastructure de calcul (e.g., cluster de calcul, *cloud*, local). Pour cela, ils s'appuient sur l'usage systématique de conteneurs logiciels (e.g., *Docker*, *Singularity*, etc.) qui encapsulent l'environnement d'exécution, assurant ainsi une parfaite maîtrise des dépendances logicielles.

La communauté *nf-core* a mis en place des lignes directrices strictes couvrant toutes les étapes du cycle de vie d'un pipeline : génération, développement, documentation, tests automatisés, révision collaborative et maintenance via des outils de synchronisation intégrés. Tous les pipelines suivent une structure commune dérivée d'un modèle (i.e., *pipelines template*), assurant une homogénéité et une montée en compétence facilitée pour les nouveaux contributeurs. Initialement focalisée sur les analyses de données génomiques, *nf-core* s'est récemment diversifiée vers d'autres champs tels que la protéomique, l'écologie, l'imagerie ou même certaines applications en sciences sociales et économiques. La force de cette initiative repose sur une communauté mondiale active, regroupant plus de 10 000 utilisateurs et développeurs sur *Slack*, favorisant un échange fluide, un soutien rapide et une innovation collective continue (Ewels et al., 2020).

Enfin, la communauté *nf-core* joue un rôle central dans l'écosystème bio-informatique plus large, en contribuant activement au développement de *Nextflow* et à la création de recettes de paquets *Bioconda*, assurant une distribution logicielle harmonisée et pérenne.



FIGURE 8 – Logos des communautés *Bioconductor* et *nf-core*

Deuxième partie

AVANT-PROPOS



HISTORIQUE DU PROJET

Le projet présenté dans les pages suivantes a été initié en 2008 dans le cadre d'un partenariat entre l'Association les Chiens Guides d'Aveugles de l'Ouest (ACGAO) et l'équipe « Génétique du Chien » de l'Institut de Génétique et Développement de Rennes (IGDR) sous la direction de Catherine André. Cette collaboration est née d'un besoin concret exprimé par l'ACGAO : la dysplasie coxo-fémorale (DCF) touchait un nombre important de chiens en formation, entraînant leur réforme et occasionnant ainsi un coût humain et financier important pour l'association. A cette époque, au laboratoire, l'équipe avait fini les grandes étapes de cartographie du génome et avait commencé à collecter des prélèvements de chiens atteints de différentes maladies génétiques pour en étudier les bases génétiques. C'est lors d'une conférence de l'Association Française des Vétérinaires pour Animaux de Compagnie (AFVAC) en 2006, que le Dr Emmanuel Voisin a contacté Catherine André, pour réfléchir à la question de la DCF, qui représentait un enjeu majeur pour l'ensemble des écoles de chiens guides, en France comme à l'international.

Fondée en 1975, l'ACGAO est une association reconnue pour son engagement en faveur des personnes déficientes visuelles. Elle dispose aujourd'hui de deux centres de formation : l'un à Angers, site historique et l'autre à Pont-Scorff, ouvert en 2002. L'association offre de nombreux services pour les personnes atteintes de déficiences visuelles, allant de l'apprentissage de la locomotion avec une canne blanche à la formation et à la remise de chiens-guides ou de canne blanche électronique. Chaque année, de nombreux chiens sont ainsi formés pour devenir des auxiliaires de vie indispensables. Les chiots destinés à cette formation proviennent de trois élevages en France : le Centre d'Etude, de Sélection et d'Elevage de Chiens guides pour Aveugles et autres Handicapés (CESECAH), la Maison du Chiot interne à l'ACGAO et le Centre d'élevage Jacques Bouniol de l'École des chiens guides de Paris. Les races concernées incluent des Golden et Labrador Retriever, croisés Labrador x Golden, Caniche et Berger Allemand. De leur naissance jusqu'à l'âge de deux mois, les chiots sont socialisés à l'humain et soumis à divers exercices d'éveil, de jeux et de manipulations. Une première évaluation comportementale se déroule entre 8 et 10 semaines, accompagnée d'un test génétique permettant d'évaluer le risque de développer certaines maladies héréditaires propres à chaque race (e.g., ichtyose chez les Golden Retriever, hyperuricosurie chez les Berger Allemand, dysplasie squelettique chez les Labrador Retriever, etc.).

De deux mois à un an, les chiots sont placés en famille d'accueil afin d'acquérir les bases d'obéissance, de sociabilité et de stabilité émotionnelle nécessaires à leur futur rôle. Ils sont régulièrement suivis et évalués par un éducateur canin. À partir de l'âge d'un an, les chiens jugés aptes entament une formation intensive pour devenir chiens-guides. Ils y apprennent, entre autres, à éviter les obstacles, traverser les routes en sécurité ou utiliser les transports en commun. Les chiens ayant réussi cette formation sont ensuite remis à une personne malvoyante qu'ils accompagneront jusqu'à l'âge de 10 ans, avant une retraite bien méritée.

Cependant, ce parcours est loin d'être linéaire : un nombre non négligeable de chiens doivent être réformés en cours de formation, pour des raisons comportementales ou de santé. Historiquement, la dysplasie coxo-fémorale représentait l'une des causes majeures de réforme. En effet un chien atteint développera à terme de l'arthrose, provoquant boiterie et douleurs invalidantes. Un chien atteint ne peut pas exercer durablement le rôle de chien-guide.

Aujourd'hui encore, la DCF reste une problématique importante dans les programmes de sélection des chiens-guides. C'est dans ce contexte que s'inscrit le projet de recherche de cette thèse : rechercher les facteurs génétiques impliqués dans la DCF ou influençant sa sévérité permettrait de mieux dépister les sujets à risque, d'affiner les critères de sélection et d'orienter les stratégies de reproduction. À terme, cela permettrait de réduire la prévalence de la maladie au sein des élevages de chiens-guides, en assurant un meilleur bien-être animal et une gestion plus efficiente des ressources investies dans leur formation.

DIAGNOSTIC DE LA DCF

La dysplasie coxo-fémorale (DCF) est un vice rédhibitoire chez le chien, nécessitant l'exclusion des individus atteints des programmes de reproduction. C'est pourquoi, au sein de l'ACGAO, un diagnostic de certitude par radiographie est systématiquement réalisé à l'âge d'un an pour tous les chiens en formation. Soucieuse d'améliorer la détection précoce de cette affection, l'association avait également mis en place des diagnostics complémentaires à 4 et 6 mois, notamment par la mesure de l'indice de distraction, afin d'évaluer la laxité articulaire avant même l'apparition des signes cliniques. Afin d'éviter les disparités de lecture des radiographies, depuis une dizaine d'années, l'ensemble des radiographies est systématiquement relu par le Pr vétérinaire Jean-Pierre Genevois, expert reconnu en orthopédie vétérinaire et lecteur officiel pour de nombreuses races.

Dans le cadre du partenariat entre l'ACGAO et l'équipe « Génétique du Chien » de l'IGDR, une première campagne de collecte avait été lancée en 2008, visant à recueillir des échantillons biologiques, des données cliniques et généalogiques sur les chiens formés à l'ACGAO depuis 2003. Le Dr vétérinaire Noémie Foyart, embauchée en CDD de 2008 à 2010, a réalisé cette première phase de prélèvements, poursuivie jusqu'en 2013. Le Dr Vétérinaire Anne-Sophie Guillory a poursuivi le travail de collecte, avec d'autres vétérinaires recrutés par l'équipe. Des thèses vétérinaires (i.e., Élise Krier, Clémence Quellen) sont venues compléter les fichiers de données cliniques et poursuivre la collecte des prélèvements. Cette étape de collecte a toujours été suivie à l'ACGAO par Laurence Raguin, qui poursuit encore maintenant le suivi et la mise à jour des fichiers des chiens guides. Le Vétérinaire Dr Thomas Chapalain a également réalisé les prélèvements sanguins de tous les chiens de l'école d'Angers. L'effort de collecte a d'abord ciblé les Golden Retriever et les Retriever du Labrador, principales races utilisées à l'école d'Angers, avant d'être étendu aux autres écoles partenaires, au CESECAH, ainsi qu'à d'autres races particulièrement touchées par la DCF et dont l'équipe disposait de l'ADN (Bouvier Bernois, Dogue de Bordeaux, Cane Corso, voir figure 9). Au total, entre 2008 et 2016, 331 chiens issus de l'ACGAO ont été prélevés et leur ADN stocké au Centre de Ressources Biologiques (CRB) Cani-DNA (<https://igdr.univ-rennes.fr/crb-cani-dna>). Dans le même temps, 295 autres chiens ont été collectés au CESECAH et au centre d'éducation de l'école Nord de France des chiens-guides d'aveugles (Roncq).



FIGURE 9 – Races collectées au sein du projet génétique Dysplasie Coxo-Fémorale

Pour chaque chien, de nombreuses informations ont été recueillies : identité, caractéristiques morphologiques, données de santé à différents âges, contexte familial, etc. Cette première cohorte a été génotypée sur puce à ADN Illumina 170 000 SNP et une étude génétique préliminaire de ces données a permis d'identifier quelques loci, dont un potentiellement associé à la DCF chez le Dogue de Bordeaux.

Après une interruption de la collecte en 2015, faute de financement, le projet a repris en 2018 avec le recrutement du Dr Maud Rimbault et grâce à un financement de la Fondation pour la Recherche Médicale. En 2019, Catherine André a pris contact avec la Fondation VISIO, reconnue d'utilité publique, créée en 2011 à l'initiative de l'ACGAO, dans le cadre de son pôle de recherche vétérinaire canine dédiée à la santé des chiens d'assistance. Ce pôle créé en 2016 est unique en France. Acteur majeur dans le domaine de la déficience visuelle en France, la Fondation s'illustre également par son engagement dans la recherche médicale clinique en neuro-ophtalmologie et ophtalmopédiatrie, ainsi que dans le développement de solutions numériques et technologiques d'assistance aux personnes aveugles ou malvoyantes. Elle œuvre également pour l'accessibilité à la culture, à l'éducation et au sport.

Ainsi, grâce au financement du programme qu'elle alloue depuis 2019, le projet a pu être relancé et les travaux de recherche ont pu se poursuivre avec mon embauche en février 2020, en CDD à mi-temps. La collecte a repris, là où elle s'était arrêtée 5 ans plus tôt et j'ai commencé un vaste chantier de remise en forme de tous les fichiers, réalisés par les différents vétérinaires ayant participé à cette collecte depuis 10 ans. J'ai tout d'abord constitué un fichier unique rassemblant près de 2 000 chiens, vérifié et complété toutes les données, cliniques, généalogiques et d'identification. Ces données m'ont permis de commencer à explorer la généalogie de ces chiens.

Afin de poursuivre le travail engagé, j'ai reçu un financement spécifique de la Fondation du 1^{er} octobre 2021 au 30 septembre 2024. J'ai ensuite obtenu un financement pour une 4^{ème} année de thèse, par la Fondation pour la Recherche Médicale (FRM) ainsi qu'un nouveau soutien financier par la Fondation VISIO. Tout ce travail mené depuis février 2020 a permis de constituer une collecte à la fois rétrospective et prospective de tous les chiens nés à l'ACGAO depuis 2003. Pour compléter les prélèvements, le Dr vétérinaire Lauranne Machou a été recrutée en CDD dans le cadre de ce projet afin d'assurer le suivi des échantillons et la coordination avec les vétérinaires partenaires.

COLLABORATIONS EXTÉRIEURES

Ce projet mobilise également plusieurs clubs de race afin de recruter des chiens atteints de dysplasie coxo-fémorale (i.e., notés D ou E selon la classification de la Fédération Cynologique Internationale (FCI)), ainsi que des chiens indemnes (i.e., notés FCI A). Le Retriever Club a été le premier contacté, suivi de l'Association Française des Bouviers Suisses pour la collecte de Bouvier Bernois, puis de la Société des Amateurs des Dogues de Bordeaux. La communication menée auprès de ces clubs a permis de sensibiliser de nombreux éleveurs et propriétaires, facilitant ainsi le recrutement de chiens des différentes races.

Afin d'enrichir les données cliniques des individus déjà inclus dans le CRB Cani-DNA, une collaboration étroite a été établie avec plusieurs lecteurs officiels spécialisés dans l'interprétation des radiographies de dysplasie coxo-fémorale. Les Drs vétérinaires Jean-Pierre Genevois, Gilles Chaudieu, Didier Fontaine, Didier Fau et Francis Legeard ont ainsi activement participé au partage de leurs archives et données de lecture, permettant de consolider le diagnostic des chiens collectés (i.e., notes de A à E selon la FCI, angles de Norberg Olson, ainsi que les mesures de laxité quand celles-ci sont disponibles). Cette démarche a permis de renforcer la fiabilité des données existantes et d'harmoniser les informations relatives aux cas et aux témoins dans la base de données Cani-DNA, avec le suivi particulièrement attentif et expert du Pr Jean-Pierre Genevois.

Par ailleurs, plusieurs collaborations ont été établies avec des Centres Hospitaliers Vétérinaires, les Ecoles Nationales Vétérinaires et des cliniques spécialisées en orthopédie. Nous avons aussi pu compter sur la collaboration de longue date avec la société *Antagene*. Ce laboratoire d'analyses génétiques pour les animaux de compagnie et la faune sauvage, basé en périphérie de Lyon, nous a permis d'obtenir un grand nombre de prélèvements de chiens, dont nous n'avons que les données cliniques. Ces partenariats ont été réactivés pour la collecte de tissus (e.g., cartilage, os, ligament) issus de chiens atteints et indemnes, pour des études transcriptomiques comparatives destinées à identifier des différences de profil d'expression génique associées à la DCF.

Enfin, une collaboration de longue date, mise en place par Catherine André, avec le Pr Claude Ferec au CHU/laboratoire INSERM de Brest, sur la comparaison entre les dysplasies de la hanche chez le chien et chez l'humain (cohorte de personnes Bigouden), a été réactivée. En effet, l'équipe de Brest, maintenant dirigée par le Dr

Emmanuelle Génin (UMR INSERM 1072) est spécialisée dans l'étude des bases génétiques des maladies humaines et leur projet sur la DDH a été repris ces dernières années. Plusieurs réunions d'échange ont été organisées autour des aspects cliniques, anatomiques et génétiques de cette affection dans les deux espèces et un projet de revue de la littérature comparative humain/chien a été engagé pour concrétiser cette synergie interdisciplinaire.

HYPOTHÈSES ET OBJECTIFS

Ce projet de thèse repose sur deux hypothèses principales. Premièrement, comme beaucoup d'autres maladies génétiques, les bases génétiques de la dysplasie coxo-fémorale (DCF) chez le chien pourraient présenter des similarités avec celles de la dysplasie développementale de la Hanche (DDH) chez l'humain. Cette hypothèse s'inscrit dans une logique comparative, fondée sur les mécanismes biologiques (biologie du développement et génétique) conservés entre espèces.

Deuxièmement, le suivi rigoureux, à la fois médical, comportemental et généalogique des chiens guides constitue un levier méthodologique unique pour l'étude d'une affection aussi complexe que la DCF. Ce contrôle des facteurs non génétiques permet de réduire considérablement la variabilité environnementale, souvent un obstacle majeur à la détection de signaux génétiques fiables dans les maladies multifactorielles.

Par ailleurs, l'inclusion de plusieurs races canines dans cette étude représente un atout stratégique. En effet, nous (et d'autres auteurs) faisons l'hypothèse que parmi les nombreuses régions génomiques impliquées dans la DCF et les gènes correspondants, certaines régions seront communes à plusieurs races et d'autres seront uniques ou seulement partagées par un petit nombre de races phylogénétiquement proches. Le fait de travailler avec plusieurs races nous permettrait ainsi d'identifier les combinaisons de loci réparties dans ces différentes races. Ainsi, en tirant parti des cassures haplotypiques et des régions en déséquilibre de liaison spécifiques à chaque race de chiens, il sera possible d'affiner (réduire) les régions génomiques associées à la DCF entre ces races. Cette approche comparative inter-races augmente ainsi la résolution des analyses génétiques et notre capacité de repérer des loci pertinents pour la suite du projet.

Les objectifs de cette thèse étaient donc multiples.

Le premier consistait à constituer une base de données intégrée, combinant les informations d'identification, cliniques, généalogiques et génétiques des chiens guides nés depuis 2003, ainsi que d'individus issus d'autres races, atteints et indemnes de DCF, référencés au CRB Cani-DNA. La constitution de cette ressource (de plus de 3 500 prélèvements) représente une étape essentielle, préalable à l'identification de loci chromosomiques et d'altérations génétiques associés à un risque accru de développer ou de transmettre la DCF.

Cette ressource essentielle a ensuite servi de fondement au deuxième objectif, qui consistait à rechercher les régions chromosomiques associées à la DCF dans 5 races de chiens (i.e., Golden et Labrador Retriever, Dogue de Bordeaux, Cane Corso et Bouvier Bernois), les valider et si possible les répliquer et rechercher les gènes ou régions génomiques en lien avec cette affection, comme des gènes du développement des membres, de la différenciation osseuse et articulaire, de facteurs intervenants dans la genèse du cartilage, du processus arthrosique, etc.

Enfin, le troisième objectif consistait à utiliser les résultats obtenus pour développer un test génétique de risque, destiné aux écoles de chiens guides, aux éleveurs et aux propriétaires afin de réduire l'incidence de la DCF dans les races prédisposées, via une sélection plus éclairée.

Enfin, ce travail de recherche participera à l'approfondissement des connaissances sur les mécanismes biologiques à l'origine de ce trouble du développement articulaire : les résultats pourront également éclairer la compréhension des dysplasies développementales humaines, dans une perspective de médecine comparée.

TRAVAUX RÉALISÉS

Les travaux menés au cours de cette thèse se sont articulés autour de ces trois axes, faisant appel à des compétences et expertises complémentaires, allant de la collecte de ressources biologiques à l'analyse bio-informatique via des collaborations interdisciplinaires (voir déroulement en figure 10).

A Collecte de prélèvements et des informations associées

La première étape a consisté à poursuivre, réorganiser et pérenniser la collecte d'échantillons sanguins (voir protocole en figure 11) au sein de l'ACGAO mais aussi au sein des différents clubs de races partenaires. Ce travail a nécessité une mobilisation soutenue incluant la communication auprès des éleveurs, la tenue de réunions d'information, ainsi que des déplacements pour aller à la rencontre des différents acteurs du monde cynophile. En parallèle, des démarches ont été entreprises pour impliquer les Centres Hospitaliers Vétérinaires (CHV), les Ecoles Nationales Vétérinaires et les cliniques spécialisées en orthopédie, dans le but de collecter aussi des tissus articulaires. Malgré l'intérêt manifeste pour le projet, cette partie de la collecte s'est heurtée à des contraintes opérationnelles fortes, en raison notamment de la surcharge chronique de ces structures. L'expérience nous montre qu'au-delà des rencontres, de la communication (e.g., plaquette en figure 12) et des protocoles mis en place, la présence d'un vétérinaire est nécessaire afin de stimuler sans cesse la collecte et bien gérer le circuit de prélèvements.

En revanche, la collecte des données cliniques s'est révélée particulièrement fructueuse. Elle a permis de compléter de nombreux dossiers d'individus pour lesquels des prélèvements étaient déjà disponibles au sein du CRB Cani-DNA mais dont le statut phénotypique (i.e., diagnostic FCI de DCF) manquait. De plus, la collaboration avec la société *Antagene* a permis de d'obtenir des échantillons d'ADN pour des individus dont nous avons déjà le statut dysplasique. Cette recherche croisée avec tous nos partenaires a permis de considérablement étoffer la cohorte de chiens atteints et indemnes pour les analyses génétiques.

Un article vétérinaire dans la revue *Le point Vétérinaire* a été rédigé dès 2022 : « Faire avancer la recherche sur la génétique de la dysplasie coxo-fémorale du chien » Machou Lauranne, Le Nézet Louis et Quignon Pascale. *La semaine vétérinaire*, n°1928, pages 14-15, janvier 2022 (Machou et al., 2022) (cf. Article 1).

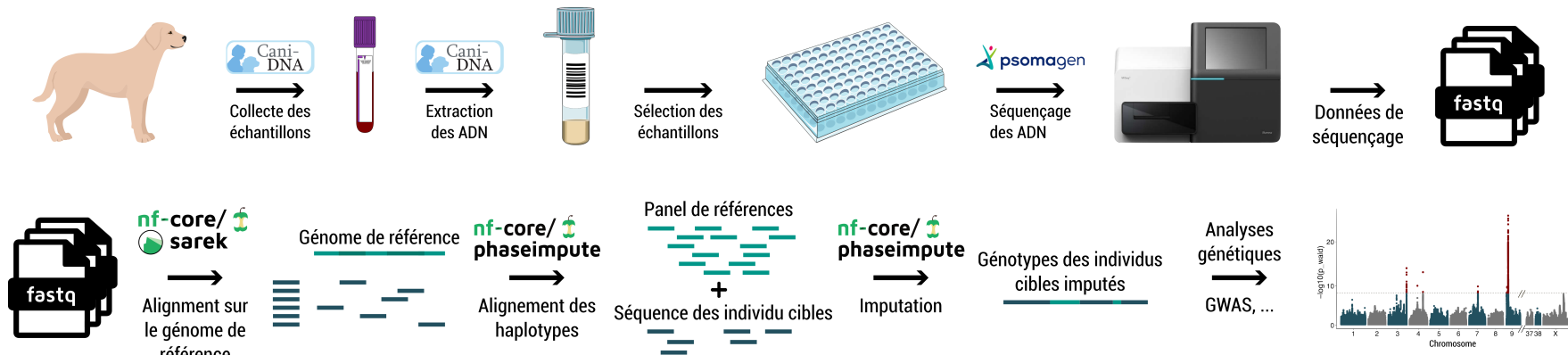


FIGURE 10 – Déroulement du projet de thèse

Schéma récapitulatif des différentes étapes du projet de thèse, depuis la collecte des échantillons et des données cliniques, jusqu'à la génération des données de séquençages (première ligne) suivi des analyses bio-informatique et statistiques (deuxième ligne).



Équipe « Génétique du chien »
 Institut de Génétique et Développement de Rennes
 UMR 6290 CNRS – Université de Rennes 1
 2 avenue du Pr Léon Bernard 35 043 Rennes Cedex
 Tél : 02 23 23 45 09 Mail : cani-dna@univ-rennes1.fr
<http://dog-genetics.genouest.org>



ETUDE GENETIQUE DE LA DYSPLASIE COXO-FEMORALE PROTOCOLE DE PRELEVEMENT SANGUIN

L'équipe "Génétique du chien" de l'Institut de Génétique et Développement de Rennes a entrepris un travail de recherche sur les causes génétiques de la dysplasie coxo-fémorale (DCF) chez le chien, en collaboration avec le Pr Jean-Pierre Genevois. Ce projet de recherche est réalisé avec la collaboration et le soutien financier de l'ACGAO et de la fondation VISIO.

Dans ce contexte, nous collectons des prélèvements sanguins de chiens de races prédisposées à la dysplasie coxo-fémorale, parmi lesquelles les races utilisées comme chiens guides. Nous avons besoin de nombreux prélèvements sanguins de chiens aux phénotypes extrêmes : stade **A** et stades **D** et **E** selon la nomenclature de la FCI (les chiens de stades B et C peuvent également être collectés mais seront moins informatifs).

Protocole pour le prélèvement sanguin :

- **prélever 3 à 5 ml de sang dans un tube EDTA K3 (tube à bouchon violet).**
 Bien mélanger le tube par retournement afin d'éviter la coagulation du sang.
- **indiquer sur le tube : le nom, le sexe, la race et la date du prélèvement.**
- **conserver le tube au réfrigérateur (4° C).**
- **envoyer le prélèvement par la poste dès que possible par Lettre suivie ci-jointe ou Colissimo (hors jours fériés et week-end) à température ambiante.**

Ce prélèvement ne doit engager aucun frais supplémentaire pour le propriétaire : soit le vétérinaire consent à la gratuité de l'acte de prélèvement dans le cadre de la recherche, soit le vétérinaire transmet une facture à l'ordre du CNRS UMR 6290. Le **remboursement par virement bancaire** du CNRS ne s'effectuera que sur présentation d'un RIB et du numéro de SIRET de la clinique.

Les prélèvements qui nous sont envoyés entreront anonymement dans le cadre d'une étude globale. Il n'y aura donc pas de résultat individuel.

Nous vous remercions sincèrement pour votre collaboration à cette étape de recrutement indispensable à la réussite de nos recherches. N'hésitez pas à nous contacter ou à consulter notre site pour tous renseignements complémentaires.

Dr Catherine André - Dr Pascale Quignon

Contact CRB Cani-DNA :

- tél : 02 23 23 45 09
- mail : cani-dna@univ-rennes1.fr

FIGURE 11 – Protocole de prélèvement sanguin transmis aux différents collaborateurs et accompagnants chaque kit de prélèvement.



UNIVERSITÉ DE
RENNES 1

Vétérinaires, Particuliers, Éleveurs,



Participez à notre programme de recherche sur la Dysplasie Coxo-Fémorale.

L'équipe "Génétique du chien" recherche des chiens atteints ou suspectés de dysplasie de la hanche :



- notés D ou E selon la nomenclature FCI
- races de chiens guides (Labrador Retriever, Golden Retriever...)
- autres races (Bouvier Bernois, Dogue de Bordeaux, Cane Corso...)
- tous âges, avec ou sans pedigree



Contactez le Centre de Ressources Biologiques Cani-DNA

 cani-dna@univ-rennes1.fr
  02 23 23 45 09

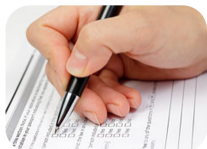
Nous vous ferons parvenir le matériel nécessaire à un prélèvement sanguin et son envoi.
L'acte de prélèvement n'engagera aucun frais supplémentaire pour le propriétaire.



Kit de prélèvement avec
enveloppe pré-affranchie



Prélèvement sanguin
sur anticoagulant (EDTA)



Questionnaire clinique
à compléter en consultation



Prélèvement à renvoyer par
lettre UN3373 (fournie)

Nous vous remercions par avance pour votre contribution à ce programme.

Équipe "Génétique du Chien" – CRB Cani-DNA
 Institut de Génétique et Développement de Rennes
 UMR 6290 – CNRS / Université de Rennes 1
 2, avenue du Pr Léon Bernard - 35 043 Rennes cedex
<https://igdr.univ-rennes1.fr/equipe-genetique-du-chien>

Avec la collaboration et le soutien financier de l'ACGAO
 (Association les Chiens Guides d'Aveugles de l'Ouest)
 et de la Fondation VISIO
<http://www.chiens-guides-ouest.org>
<http://www.fondation-visio.org>




FIGURE 12 – Plaquette de communication destinée aux éleveurs et propriétaires de chiens atteints de DCF.

B Comparaison de la dysplasie de la hanche chez le chien et l'humain

Un important travail de synthèse bibliographique a été mené en collaboration avec des experts vétérinaires et médecins orthopédistes, dans le cadre d'un partenariat avec l'INSERM UMR 1078 de Brest et le Pr Genevois, vétérinaire spécialiste en orthopédie, de l'Ecole Nationale Vétérinaire de Lyon. Ce travail a pour objectif la rédaction d'un état de l'art sur la dysplasie de la hanche, intégrant les aspects anatomiques, cliniques et génétiques chez l'humain et chez le chien. En parallèle, j'ai créé une suite d'outils sous le langage statistique *R* afin de comparer les loci identifiés dans la littérature chez l'humain comme chez le chien. Cela permet ainsi d'examiner les convergences éventuelles avec la DDH chez l'humain.

L'article est désormais rédigé en français et nécessite désormais d'être traduit et condensé en anglais : "*Hip dysplasia: a One Health approach between humans and dogs*" / « Dysplasie de la hanche : une approche *One Health* entre chiens et humains. » Le Nézet Louis, Genevois Jean-Pierre, Machou Lauranne, Le Douarin Niels, Caron Alexandre, Lefèvre Christian, Gaucher François, Scotet Virginie, Rouault Karen, Quignon Pascale, Génin Emmanuel, Férec Claude et André Catherine (cf. Article 2).

C Développement d'outils bio-informatiques

1 Préparation des données et constitution d'un fichier unique

Un effort particulier a été porté sur l'harmonisation des données recueillies. J'ai ainsi créé un script en langage *Python* afin d'automatiser la concaténation et l'intégration de nombreux fichiers *Excel* hétérogènes, permettant de consolider toutes les informations en une base de données structurée unique, contenant plus de 3 500 chiens.

2 Développement d'un package *R Bioconductor* de construction de pedigree : *Pedixplorer*

Parallèlement, j'ai développé un package *R* de construction d'arbres généalogiques, nommé *Pedixplorer* au sein de la communauté *Bioconductor*. Cet outil permet de générer automatiquement des arbres généalogiques à partir de données de parenté sous format tabulaire. Cet outil permet non seulement une visualisation claire et interactive des pedigrees, mais aussi une interrogation fine des données généalogiques. Il a été spécifiquement pensé pour faciliter l'exploration et la sélection de familles d'intérêt en vue d'analyses de liaison génétique, en identifiant par exemple des lignées comportant un nombre important d'individus atteints ou un échantillon d'ADN est disponible.

Cet outil a été publié dans la revue *Bioinformatics* : “*Pedixplorer: a Bioconductor package to streamline pedigree design and visualization*” Le Nézet Louis, Sinnwell Jason, Letko Anna, André Catherine et Quignon Pascale. *Bioinformatics* Volume 41, Issue 6, juin 2025, btaf329. 10.1093/bioinformatics/btaf329 (Le Nézet et al., 2025) (cf. Article 3).

3 Développement d'un pipeline d'imputation : *nf-core/phaseimpute*

J'ai également travaillé, dès le début de ma 2^{ème} année de thèse, à l'imputation des données génétiques issues des données de séquençage 1×. En effet, une collaboration s'est organisée avec le Dr Antony Herzig (Laboratoire GGB, INSERM, de Brest) pour mettre en place une suite logicielle permettant l'imputation de données issues de séquençage *low-pass*. Notre premier jeu de données de séquençage *low-pass* a initialement été confiée à la société *Gencove* (États-Unis). Cependant nous avons souhaité mieux maîtriser cette analyse, en raison de l'impossibilité de la société *Gencove* d'exploiter la dernière version du génome de référence canin, *canFam4*. J'ai donc développé un pipeline bio-informatique dédié à cette tâche. Ce pipeline, conçu au sein de la communauté *nf-core*, permet l'alignement des haplotypes et l'imputation des données génétiques par le biais de plusieurs outils, que j'ai agencé pour leur bon fonctionnement. Désormais publié, ce pipeline a permis l'imputation de l'ensemble des données génomiques du projet sur la version la plus récente du génome de référence (i.e., *canFam4*), à partir des données de séquence de génomes de chiens les plus abouties : celles du panel de référence du consortium *Dog 10 000 genomes*.

Le pipeline *nf-core/phaseimpute* a fait l'objet d'un *proceeding* et d'une présentation au congrès JOBIM : “*PhaseImpute an nf-core pipeline for genetic imputation*” Le Nézet Louis, communauté *nf-core*, Quignon Pascale et André Catherine. Article de colloque présenté aux Journées Ouvertes en Biologie, Informatiques et Mathématiques juin 2024, Toulouse.

Une version plus complète de cet article est en cours de rédaction et sera soumis à *Bioinformatics* : “*nf-core/phaseimpute a pipeline for genetic imputation*” Le Nézet Louis, Trigila Anabella, communauté *nf-core*, Derrien Thomas, André Catherine, Herzig Anthony et Quignon Pascale (cf. Article 4).

D Analyses d'association pangénomiques

Partant de la base de données consolidée, un ensemble de plus de 900 individus appartenant à 5 races a été sélectionné. Pour chacun, j'ai extrait l'ADN, puis quantifié leur concentration par spectrophotométrie (i.e., *Qubit*). Parmi eux, plus de 700 chiens ont été sélectionnés pour un séquençage de type *low-Pass Whole Genome Sequencing* à une couverture moyenne de 1×, réalisé par la société *Psomagen* (États-Unis). J'ai ensuite analysé les données brutes, réalisé les contrôles qualité (i.e., *QC*), puis imputé les données de séquences ainsi générées à l'aide du génome canin de référence *can-Fam4*, grâce à *nf-core/phaseimpute*. J'ai ensuite utilisé ces données de génomiques pour réaliser les analyses dites d'association pangénomiques (GWAS), afin d'identifier les régions chromosomiques (loci) associées avec la maladie.

Un article décrivant les résultats des GWAS sur les 5 races de chiens étudiées est en cours de rédaction en parallèle de la thèse et sera soumis à une revue de génétique : *"Identification of loci predisposing to Canine Hip Dysplasia in four dog breeds using imputed genotypes from low-pass sequencing"* Le Nézet Louis, Machou Lauranne, Hédan Benoit, Guyon Richard, André Catherine et Quignon Pascale (cf. Article 5).

ARTICLE 1

« Faire avancer la recherche sur la génétique de la dysplasie coxo-fémorale du chien »

AUTEURS : Lauranne Machou, Louis Le Nézet, Pascale
Quignon

STATUT : Publié

JOURNAL : *La semaine vétérinaire*, n°1928, 21 janvier 2022

PRATIQUE CANINE FÉLINE NAC

ORTHOPÉDIE

FAIRE AVANCER LA RECHERCHE SUR LA GÉNÉTIQUE DE LA **DYSPLASIE COXO-FÉMORALE DU CHIEN**

L'équipe « Génétique du chien » (IGDR, CNRS-Université de Rennes 1) recherche les bases génétiques de maladies homologues entre l'humain et le chien avec pour objectif un bénéfice mutuel pour les médecines vétérinaire et humaine. **PAR LAURANNE MACHOU, LOUIS LE NÉZET, PASCALE QUIGNON¹**

La dysplasie coxo-fémorale (DCF) chez le chien correspond à la dysplasie développementale de la hanche chez l'humain. Cette anomalie du développement, consécutive à une laxité excessive de l'articulation, se traduit par une mauvaise coaptation et une subluxation de la tête fémorale par rapport à l'acétabulum. Elle détermine à plus ou moins long terme une déformation des profils articulaires et l'installation d'un processus arthrosique. Cette maladie génétique affecte la mobilité, le bien-être, et conduit à la réforme précoce de nombreux chiens de travail, dont les chiens guides. Elle touche particulièrement les races de grande taille et reste présente dans de nombreuses races malgré les efforts de sélection appliqués dans les élevages.

Suite aux coûts engendrés par les réformes, la déception des propriétaires et la souffrance des chiens, l'Association les Chiens guides d'aveugles de l'Ouest (ACGAO) et la fondation Visio se sont engagées depuis plusieurs années auprès de l'équipe du CNRS-Université de Rennes 1, avec la collaboration du Pr. Jean-Pierre Genevois, pour mener un projet de recherche sur les causes génétiques de cette affection.

Un mode de transmission génétique complexe

Si cette maladie est liée à des facteurs environnementaux (nutrition, exercice), son origine génétique est maintenant avérée : le mode de transmission est complexe, faisant intervenir plusieurs gènes, avec des combinaisons différentes en fonction des races. À ce jour, les gènes et les mécanismes impliqués dans la DCF sont encore peu connus chez le

chien comme chez l'humain, et le chien en représente un bon modèle spontané. En médecine vétérinaire, même si des tests génétiques ont déjà été commercialisés, ils ne sont pas encore aboutis et de fait, peu utilisés.

Le but de l'étude en cours est double. Il s'agit tout d'abord de rechercher les *loci* (régions chromosomiques) conférant un risque de développer et de transmettre la DCF dans les races de chien guide et d'autres races. Les applications à long terme sont, chez le chien, le développement de tests génétiques de risque permettant de diminuer l'incidence de la maladie dans la race, tout en maintenant une diversité génétique suffisante. Dans un deuxième temps, les connaissances acquises chez le chien seront transférées à la médecine humaine. En effet, ces recherches peuvent aider à l'identification de nouveaux marqueurs, nouveaux gènes et mécanismes moléculaires, moins difficiles à identifier chez le chien, ouvrant ainsi la voie à des dépistages génétiques voire à de nouvelles thérapies chez le chien comme chez l'homme.

Recherche des bases génétiques

La méthode de recherche des bases génétiques de la DCF repose dans un premier temps sur la comparaison des génomes de chiens présentant des phénotypes extrêmes : chiens diagnostiqués A *versus* chiens diagnostiqués D ou E, selon la classification de la Fédération cynologique internationale (FCI). Cette méthode nécessite la collecte de prélèvements sanguins ainsi que des statuts dysplasiques de chiens A et de chiens D/E. Le diagnostic retenu est effectué par radiographie à l'âge d'un an.

Les ADN issus des prélèvements sanguins sont ensuite génotypés pour des centaines de milliers de marqueurs répartis sur tout le génome. L'objectif des analyses génétiques est d'identifier des marqueurs dont les allèles sont plus fréquemment retrouvés chez les chiens

Classification (FCI)	Chiens déjà prélevés		Chiens restant à prélever	
	A	D ou E	A	D ou E
Races				
Retriever du Labrador	256	53	np*	50
Golden retriever	63	24	40	80
Berger allemand	52	30	50	70
Bouvier bernois	221	33	np*	70
Cane corso	19	21	80	80
Dogue de Bordeaux	134	75	np*	30
Total	745	236	170	380

Nombre de prélèvements obtenus dans le Centre de ressources biologiques (CRB) Cani-DNA au 1^{er} septembre et nombre de prélèvements restant à effectuer.
(np* = prélèvement non-prioritaire)

© Legofin-Stock



Le but de l'étude menée par l'équipe « Génétique du chien » du CNRS-Université de Rennes est de rechercher les *loci* conférant un risque de développer et de transmettre la dysplasie coxo-fémorale dans les races de chien guide et d'autres races. Les applications à long terme sont, chez le chien, le développement de tests génétiques de risque permettant de diminuer l'incidence de la maladie dans la race, tout en maintenant une diversité génétique suffisante. Dans un deuxième temps, les connaissances acquises chez le chien seront transférées à la médecine humaine.

atteints (D/E) que chez les chiens indemnes (A). Ces marqueurs permettront le développement de tests génétiques de risque et, à terme, la mise en évidence des altérations génétiques impliquées dans la prédisposition à la DCF.

En parallèle, l'analyse des statuts dysplasiques sera réalisée par famille de chiens afin d'étudier la ségrégation de la DCF dans les pedigrees disponibles, avec, le cas échéant, le séquençage du génome des chiens les plus informatifs. Cela permettra de mettre en évidence des altérations génétiques communes à une famille et affiner ainsi la compréhension de la maladie.

Les praticiens acteurs de la recherche

Ce travail est rendu possible par l'implication de vétérinaires traitants qui procèdent aux prélèvements sanguins, avec le consentement du propriétaire. Ces prélèvements sanguins sont effectués lors de la radiographie ou lors de la visite annuelle si le statut du chien est connu. En outre, au cours d'une chirurgie orthopédique, les vétérinaires sont invités à prélever cartilage, capsule articulaire ou liquide synovial. Ces prélèvements seront alors utilisés pour étudier l'expression des gènes dans ces tissus spécifiques. À ce jour, les races de chiens étudiées sont les chiens guides (labrador retriever, golden retriever, berger allemand...) mais aussi les races d'intérêt suivantes : bouvier bernois, dogue de Bordeaux, cane corso. Afin d'obtenir une puissance statistique suffisante pour les analyses génétiques, il est nécessaire de collecter les prélèvements de plus de 100 chiens A et plus de 100 chiens D ou E par race étudiée. Le Centre de ressources biologiques (CRB) Cani-DNA (CNRS-Université de Rennes, rattaché à l'équipe) a déjà collecté les prélèvements sanguins de 745 chiens indemnes (A) et 236 chiens dysplasiques (D/E) dans les six races. Ainsi, l'effort de collecte est à prioriser pour les chiens D ou E dans ces six races, pour pouvoir procéder aux études génétiques (voir tableau). Les vétérinaires qui souhaitent s'impliquer dans ce projet peuvent prélever les chiens A ou les chiens D ou E des races concernées et ayant fait

l'objet d'un dépistage officiel par radiographie de la dysplasie. Le sexe et l'âge sont indifférents, mais doivent être indiqués. Les prélèvements sanguins (5 ml) sont à réaliser sur un tube EDTA, les prélèvements tissulaires sur un tube RNA-later et, si possible aussi, sur un tube de formol. Un kit de prélèvements, le protocole détaillé ainsi que les documents à compléter seront envoyés sur demande aux vétérinaires volontaires. Chaque prélèvement devra être accompagné du résultat de la radiographie (compte-rendu de lecture et/ou images radiographiques), du questionnaire clinique complété, du consentement du propriétaire signé, ainsi que, si le chien est LOF, de son pedigree. L'acheminement des prélèvements et des documents associés vers le laboratoire sera organisé par le CRB Cani-DNA.

Les vétérinaires désireux de participer à ce projet peuvent s'adresser au Centre de ressources biologiques Cani-DNA de Rennes, par mail à l'adresse cani-dna@univ-rennes1.fr ou par téléphone au 02 23 23 45 09.

Plus d'informations sur le site Internet : <https://dog-genetics.genouest.org/dysplasie-de-la-hanche/>

Respectueux hommage à Francis Legeard, qui avait participé à ce travail depuis de nombreuses années. ●



Retrouvez la bibliographie de cet article sur www.bit.ly/34ZmkFe

1. Université de Rennes 1, CNRS, IGDR (Institut de génétique et développement de Rennes), UMR 6290, Équipe « Génétique du chien », 35000 Rennes

Troisième partie

RÉSULTATS



ÉTUDE BIBLIOGRAPHIQUE

COMPARATIVE

Dans le cadre de cette thèse, une revue de la littérature a été entreprise afin d'explorer les parallèles entre la Dysplasie Coxo-Fémorale (DCF) chez le chien et la Dysplasie Développementale de la Hanche (DDH) chez l'humain. Cette étude comparative vise à identifier les similarités et les différences entre ces deux affections homologues, tant sur le plan anatomique que physiopathologique.

Cette revue débute par une comparaison détaillée de l'articulation coxo-fémorale chez l'humain et le chien, en soulignant les mécanismes physiopathologiques spécifiques à chaque espèce, de même que les approches diagnostiques employées respectivement en médecine vétérinaire et en médecine humaine. Les facteurs environnementaux susceptibles d'influencer l'expression clinique de ces affections sont également abordés.

Enfin, un volet génétique compare des loci associés à la DCF et à la DDH. Pour ce faire, j'ai développé une suite de scripts en langage *R*, reposant sur une série d'étapes de *lift-over*. En effet, les études génétiques portant sur la dysplasie de la hanche, chez l'humain comme chez le chien, couvrent une longue période durant laquelle les technologies d'acquisition des données génomiques ont considérablement évolué (e.g., STR, puces à ADN, séquençage de génome complet, etc.). Les génomes de référence eux-mêmes ont été régulièrement mis à jour. Il est donc nécessaire de standardiser l'ensemble des résultats sur une même version du génome pour pouvoir comparer les loci identifiés. Cette conversion d'une version de génome à une autre, et d'une espèce à l'autre, a été réalisé par l'outil *liftOver* (Bioconductor Package, 2025 ; *ucscGenome-Browser/Kent*, 2025), qui utilise des fichiers *.chain* pour traduire les régions chromosomiques. Ainsi, les données issues des différentes études ont toutes été converties vers la version *canFam4* pour le chien et vers la version *hg38* pour l'humain. Les régions génomiques identifiées chez l'humain ont ensuite été projetées sur le génome canin, révélant plusieurs zones de correspondance (i.e., synténie) entre les deux espèces.

Cette revue a pour objectif de nourrir une réflexion comparative entre médecine vétérinaire et médecine humaine, pour ouvrir la voie à une meilleure compréhension des mécanismes physiopathologiques et génétiques sous-jacents à ces affections articulaires et ainsi identifier, dans le cours de ce travail, les loci et gènes ou régions génomiques les plus en lien avec la physiopathologie de cette affection.

ARTICLE 2

« La dysplasie de la hanche : une approche *One Health* entre humain et chien »

AUTEURS : Louis Le Nézet, Jean-Pierre Genevois, Lauranne Machou, Niels Le Douarin, Alexandre Caron, Christian Lefèvre, François Gaucher, Virginie Scotet, Karen Rouault, Pascale Quignon, Emmanuel Génin, Claude Férec, Catherine André

STATUT : En préparation

JOURNAL : A définir

La dysplasie de la hanche : une approche *One Health* entre humain et chien

Louis Le Nézet^{1,*}, Jean-Pierre Genevois², Lauranne Machou¹, Niels Le Douarin¹, Alexandre Caron³, Christian Lefèvre, François Gaucher, Virginie Scotet⁶, Karen Rouault⁶, Anthony Herzig⁶, Pascale Quignon¹, Emmanuel Génin⁶, Claude Férec⁶, Catherine André¹,

¹ UMR-6290, IGDR (Institute Genetics and Development of Rennes), University of Rennes, CNRS, Rennes, France

² Unité de Chirurgie, Campus Vétérinaire de Lyon VetAgro Sup, Marcy l'Etoile, France

³ Clinique vétérinaire Trio Vet, Rennes, France

⁶ Inserm UMR-1078, CHRU Brest, University of Brest, Brest, France

* Corresponding author. louislenezet@gmail.com

1 Objectif

Cette revue a pour objectif de décrire et comparer les dysplasies développementales de la hanche (DDH), chez l'humain et le chien, au niveau anatomique, tissulaire voire cellulaire, afin d'avoir une vue mécanique et fonctionnelle de cette articulation complexe pour mieux estimer la pertinence des gènes candidats impliqués dans cette maladie du développement. Un état de l'art des données de génétiques publiées à ce jour dans les 2 espèces permet de comparer les loci identifiés et les gènes candidats afin de voir lesquels seraient partagés entre l'humain et le chien ou spécifiques à chaque espèce, voire à chaque race de chiens ou groupe ethnique humain, afin d'éclairer les causes génétiques de cette affection pour l'humain et le chien.

2 Description

Les DDH sont des anomalies du développement de l'articulation de la hanche qui se traduisent par une mauvaise coaptation ou une mauvaise congruence entre la tête fémorale et l'acétabulum. Les altérations portent sur les constituants articulaires dont le développement est retardé, incomplet ou irrégulier. Elle peut se traduire à l'âge adulte par des défauts anatomiques plus ou moins importants, facteurs d'arthrose secondaire. Anciennement appelé Luxation Congénitale de la Hanche chez l'humain, les DDH ont changé de nom, afin de refléter l'absence possible de cette affection à la naissance (i.e., non congénitale) ainsi que le *continuum* d'anomalies représenté par les termes suivants : la dysplasie acétabulaire, la subluxation, la hanche luxable, la hanche luxée. Chez le chien le défaut de coaptation serait dû à une trop grande laxité de l'articulation. Il était classique de distinguer les dysplasies acétabulaires (i.e., défaut de couverture de la tête fémorale par l'acétabulum) et les dysplasies fémorales (i.e., modifications

angulaires entre le fût fémoral et l'ensemble tête-col, conduisant à décrire des hyperantéversions et des *coxo valga*). Cependant cette distinction n'est plus faite par les vétérinaires.

Chez l'humain, il est aussi nécessaire de distinguer les différentes dysplasies congénitales. Certaines sont d'origine tératogène : luxation dite tératologique de hanche et d'autres dues à une hausse des contraintes mécaniques lors de la grossesse : syndrome positionnel du nouveau-né (i.e., syndrome du bébé moulé), oligoamnios, présentation par le siège. Ces dernières causes augmentent les contraintes sur une ou les deux hanches ou favorisent un positionnement en adduction.

L'origine de la DDH d'un individu va donc aussi dépendre de sa période de développement durant la gestation :

- la luxation tératologique durant la période embryonnaire (0 à 10 semaines) ;
- les dysplasies développementales d'origine génétiques durant la période dite de mobilité et de croissance (10 à 25 semaines) ;
- les dysplasies développementales d'origine mécanique durant la période dite de relative immobilité (25 semaines à la naissance).

Du fait de la très grande diversité des affections rencontrées et de la confusion possible des différentes catégories, un très grand nombre de noms ont été donnés à ces anomalies du développement.

Nous retiendrons les termes suivants :

- chez le chien : Dysplasie Coxo-Fémorale (DCF) est utilisé en français et plus simplement dysplasie de la hanche *Canine Hip Dysplasia* en anglais ;
- chez l'humain : Maladies Luxantes de la Hanche (MLH) est retrouvé en français mais le terme *Developmental Dysplasia of the Hip* en anglais pouvant se traduire par Dysplasie Développementale de la Hanche (DDH) sera préféré et utilisé ici.

3 Anatomie de la hanche chez l'humain et le chien

3.1 Développement d'une articulation synoviale

La formation de l'articulation de la hanche se passe en plusieurs étapes voire figure 1. Tout d'abord le mésenchyme se condense et se chondrifie, puis, au niveau de l'articulation, les cellules s'aplatissent et forment l'interzone. Cette dernière va réguler, par différentes molécules, la croissance ainsi que la différenciation des tissus environnants et ainsi contrôler la formation de l'articulation. Une cavitation va se produire à partir de cette interzone, séparer la structure en deux éléments et créer ainsi la cavité synoviale. Au fur et à mesure que la cavité s'agrandit, l'interzone se disperse, la membrane synoviale et le cartilage articulaire se forment par différenciation des tissus (Archer et al., 2003 ; Khan et al., 2007).

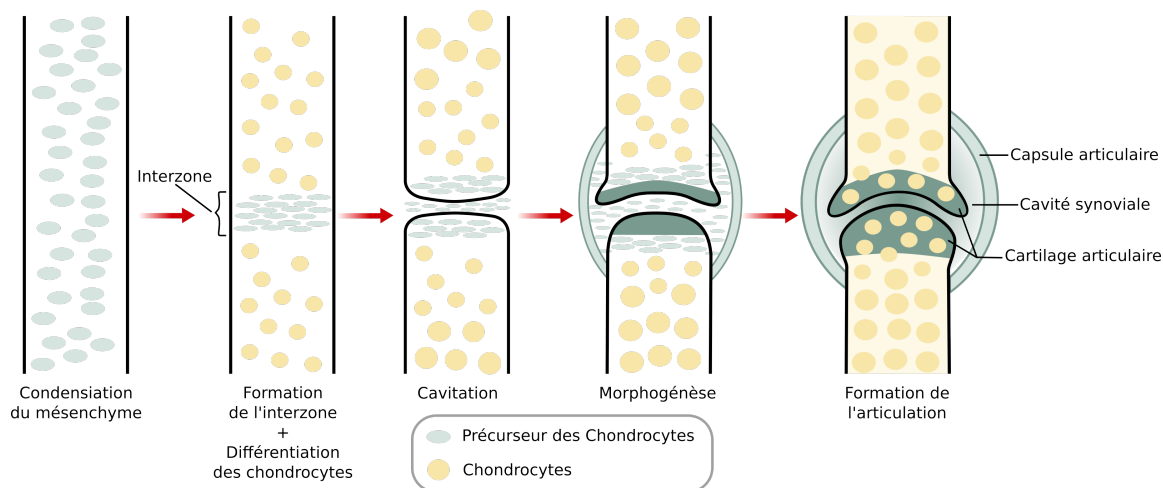


FIGURE 1 - Grandes étapes de la formation d'une articulation synoviale

*

Schéma inspiré de Khan et al. (2007) et Pacifici et al. (2005).

3.2 Adaptation, régénération et vascularisation

L'os est un tissu dynamique en perpétuel réarrangement par l'action d'ostéoblastes et d'ostéoclastes, permettant ainsi de réparer les dommages et de s'adapter aux efforts mécaniques (Robling et al., 2006). Ces efforts mécaniques modifient le développement du squelette en favorisant la chondroblastie du mésenchyme (Hall, 2005) ou en modifiant, chez les chondrocytes, la synthèse de protéoglycanes et de collagène de type 2, composants importants du cartilage (Tanimoto et al., 2010). L'os est un tissu vascularisé, afin d'alimenter ses cellules. Dans sa partie interne, l'os spongieux est constitué de moelle osseuse rouge et jaune, organisées en trabécules et petites cavités, responsables de la production des cellules sanguines et du maintien de l'homéostasie. Cette vascularisation est similaire chez l'humain et le chien (Eitel et al., 1981).

3.3 Composition et densité osseuse

Si l'on s'intéresse à la composition, la densité et la qualité des os de différentes espèces animales communément utilisées en recherche orthopédique (i.e., chien, cochon, vache, mouton, poule et rat), le chien est le modèle se rapprochant le plus de l'humain : sur le plan de la composition chimique de l'os, aussi bien cortical que trabéculaire et plus particulièrement concernant sa composition en cendres, eau, molécules organiques et inorganiques, hydroxyproline, protéine totales et *IGF1* (Gong et al., 1964 ; Aerssens et al., 1998). De plus, le chien possède à l'âge adulte plus d'os spongieux se rapprochant ainsi plus de l'humain (Eitel et al., 1981). Cependant l'humain se distingue par une densité osseuse plus faible (178 mg/cm^3) que le chien (340 mg/cm^3) ainsi qu'une résistance au stress plus faible (Aerssens et al., 1998). De plus, la régénération osseuse, bien que difficile à mesurer, semble différente entre l'humain et le chien (Kimmel & Jee, 1982 ; Bloebaum et al., 1993 ; Pearce et al., 2007).

3.4 Description de l'articulation coxo-fémorale chez l'humain

La structure de l'articulation coxo-fémorale est complexe et fait intervenir de nombreuses structures, tissus et types cellulaires différents (voir figure 2). Il sera fait ici un bref résumé de cette complexité afin de pouvoir identifier les possibles causes de la dysplasie ainsi que d'établir les bases de comparaison entre humain et chien.

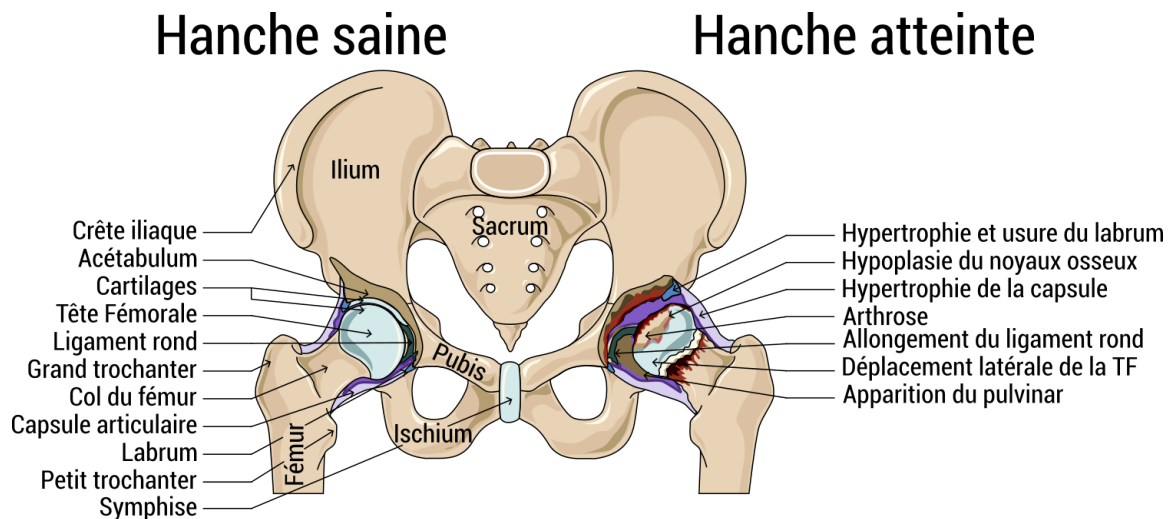


FIGURE 2 - Comparaison schématique d'une hanche saine et d'une hanche atteinte par la dysplasie de la hanche.

3.4.1 Structures osseuses

L'articulation de la hanche est une articulation sphéroïde composée de deux os : le fémur et l'os coxal. La jonction de ces deux structures se fait entre la tête fémorale et l'acétabulum. L'os coxal appartient au « bassin » (ou ceinture pelvienne). Cet ensemble osseux est constitué du sacrum, du coccyx et des deux os coxaux (os iliaques).

L'acétabulum, également appelé cotyle (du grec : κοτύλη « écuelle ») est une demi-sphère creuse. Son bord postérieur est rectiligne et son bord antérieur est échancré et forme une ligne brisée. Son orientation est inclinée vers le bas et en antéversion. Les « cornes » de l'acétabulum sont mobiles et s'ouvrent lors de la marche. Le ligament transverse freine cette mobilité. L'arrière-fond (i.e., zone du toit fixe) de l'acétabulum absorbe l'énergie cinétique.

Le fémur est le segment proximal du membre inférieur, comprenant la diaphyse (i.e., partie centrale du fémur) ainsi que deux épiphyses dont la tête fémorale. Il est l'os le plus long du corps humain. L'extrémité supérieure du fémur (i.e., la plus proche de l'os coxal) est constituée d'une saillie articulaire, la tête fémorale, posée sur un segment cylindrique, le col du fémur, ainsi que deux éminences osseuses servant à l'insertion de muscles : le grand et le petit trochanter. La tête fémorale est constituée d'une portion de sphère, inclinée vers l'intérieur (i.e., médial), vers le haut et en antéversion (i.e., vers l'avant).

3.4.2 Mouvements possibles

L'articulation de la hanche est une articulation synoviale de type cotyloïde autrement appelée sphéroïde et plus particulièrement énarthrose (i.e., la partie concave englobe l'équateur de la partie convexe). Elle possède trois axes directeur (i.e., transverse, sagittal, longitudinal) et trois degrés de liberté.

Elle permet ainsi les mouvements, décrits en figure 3, de :

- flexion (120°) et d'extension ($5-30^\circ$) : respectivement réduction et augmentation de l'angle entre le fémur et le plan de l'acétabulum ;
- adduction ($20-30^\circ$) et abduction ($45-50^\circ$) : respectivement réduction et augmentation de la distance entre le fémur et le plan sagittal ;
- rotation interne (médiale) et externe (latérale) 45° .

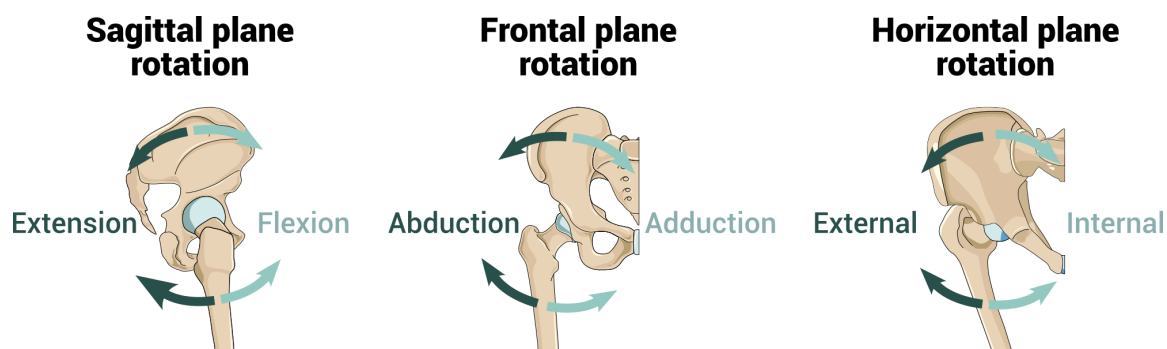


FIGURE 3 - Mouvements possibles de rotation de l'articulation coxo-fémorale chez l'humain

3.4.3 Muscles et cartilages de l'articulation

L'ensemble des mouvements décrits plus hauts est rendu possible par la présence de 22 muscles s'articulant autour de la hanche (voir figure S1). Des cartilages au niveau de la zone de contact entre la tête fémorale et l'acétabulum permettent des mouvements sans friction grâce au liquide synovial. Le cartilage de la tête fémorale la recouvre en quasi-totalité, en dehors d'une légère dépression ovoïde, appelée la fossette du fémur, qui est le point d'insertion du ligament de la tête du fémur. L'acétabulum est recouvert de cartilage uniquement en périphérie, sous la forme d'un croissant ouvert en bas, nommée surface lunaire de l'acétabulum, qui correspond à la zone d'appui de la tête fémorale. Cette surface est poursuivie par le labrum acétabulaire (i.e., bourrelet cotyloïdien), situé sur le rebord saillant externe de l'acétabulum (i.e., limbus), qui forme un anneau et qui étend la surface de contact entre la tête fémorale et l'acétabulum. Ce surplus de cartilage est également retrouvé chez les chiens de grande taille. La capsule articulaire est une structure fibreuse de contention qui recouvre l'ensemble de l'articulation en s'insérant à distance des cartilages mentionnés ci-dessus.

3.4.4 Ligaments

Trois ligaments sont présents autour de la capsule articulaire pour renforcer son rôle de contention :

- le ligament ilio-fémoral : en position antérieure, il unit l'épine iliaque antérieure et inférieure au petit et au grand trochanter ;
- le ligament pubo-fémoral : en position antérieure et inférieure, il relie la branche supérieure du pubis à la ligne intertrochantérique ;
- le ligament ischio-fémoral : en position postérieure, il relie la partie supérieure de la tubérosité ischiatique à la crête intertrochantérique.

La tête fémorale est de plus maintenue par le labrum acétabulaire, qui forme le ligament transverse de l'acétabulum, au niveau de l'incisure de l'acétabulum. Enfin un ligament intra-articulaire nommé ligament de la tête fémorale ou ligament rond, relie la fossette de la tête fémorale aux cornes de la surface lunaire ainsi qu'au ligament transverse. Le rôle du ligament rond diffère chez l'humain et chez le chien. Chez le chien, il exerce un rôle de soutien mécanique à la naissance et de porte vaisseaux pour l'alimentation de la croissance de la tête fémorale, il joue un rôle d'amortisseur des pressions chez l'animal adulte (Riser & Shirer, 1966). Chez l'humain son rôle est essentiellement vasculaire.

3.4.5 Liquide synovial

L'ensemble de l'articulation est lubrifié par le liquide synovial, contenu à l'intérieur de la cavité articulaire. Ce dernier est sécrété par les synoviocytes, les cellules de la membrane synoviale et sa grande viscosité (due à la présence d'acide hyaluronique) contribue au glissement harmonieux (i.e., sans friction) des surfaces articulaires.

Ces éléments anatomiques sont majeurs dans la compréhension des mécanismes articulaires mécaniques et cellulaires fonctionnels, en jeu dans cette affection développementale qu'est la dysplasie de la hanche. Ainsi, les cellules productrices / régulatrices des cartilages de la tête fémorale et de l'acétabulum, de la capsule articulaire, des ligaments et de la membrane synoviale (les synoviocytes, qui baignent dans le liquide synovial), vont être important à étudier. Il est donc important de collecter ces éléments, afin d'étudier l'expression de leurs gènes respectifs et leur différentiel d'expression dans les tissus de chiens atteints *versus* ceux de chiens indemnes.

3.5 Description de l'articulation coxo-fémorale chez l'humain

D'un point de vue anatomique, l'articulation de la hanche chez l'humain et chez le chien sont extrêmement similaires. Bien que leur forme et leur charge d'utilisation soit différentes, l'articulation de la hanche est composée des mêmes os, des mêmes muscles et d'un réseau sanguin similaire chez les 2 espèces (Pascual-Garrido et al., 2018).

4 Développement des dysplasies de la hanche

4.1 Hypothèses du développement anormal de la hanche

La tête fémorale est, en temps normal, assimilable à une sphère (i.e., un seul foyer). Il y a alors une bonne congruence entre les deux surfaces lors des différents mouvements. Lorsque la hanche est dysplasique, la tête fémorale est assimilable à un ellipsoïde (i.e., deux foyers). De fait, lors des mouvements cela entraîne une mauvaise congruence et est arthrogène.

De plus, la dysplasie se caractérise, chez l'humain, par une hypoplasie du toit, réduisant encore davantage la surface de contact déjà réduite. L'angle VCA (i.e., angle de couverture cotyloïdienne antérieure) est plus petit et le rebord antérieur de l'acétabulum couvre peu la tête fémorale. Des douleurs inguinales et antérieures peuvent alors se faire sentir.

Plusieurs théories ont été avancées sur la pathogenèse de la luxation congénitale de la hanche et de la dysplasie développementale. Celles-ci se divisent sur le responsable du défaut de développement : échec germinale du développement du contrefort postéro-supérieur de l'ilium, développement d'un acétabulum plat ou bien encore laxité excessive. Ces différentes hypothèses ont suscité de nombreux débats parmi les scientifiques (Henrigson et al., 1966 ; Brand, 2008).

Le rôle de la laxité a évolué au fil du temps. Au début du XX^{ème} siècle, il était considéré comme impossible de diagnostiquer la maladie avant la marche (Lorenz, 1920), mais des avancées dans le dépistage précoce (Ortolani, 1948 ; Barlow, 1962 ; Henrigson et al., 1966) ont montré que la laxité pouvait être considérée comme primitive et un élément fondamental au développement de la maladie. Elle est désormais considérée comme une des composantes de la dysplasie, impliquant des facteurs mécaniques *in-utero*, des influences *post-partum* et des prédispositions génétiques. L'explication de la genèse de la DDH se dirige désormais vers un défaut développemental. Les déformations de la tête fémorale et de l'acétabulum seraient réciproques et secondaires à des anomalies de développement, à la fois liés à la génétique et à l'environnement.

Le développement anormal de la hanche serait ainsi le suivant :

« Par le biais d'un défaut de développement, l'acétabulum ne s'est pas approfondi et la tête et le col du fémur se sont antéversés. L'antéversion tend à tourner la tête vers l'avant, déplaçant la sphère cartilagineuse latéralement, de sorte que le labrum glénoïdal et l'acétabulum couvrent moins la tête que d'ordinaire. La mauvaise adaptation de la tête et de l'acétabulum se poursuit de manière dynamique et nécessite des modifications de la croissance, ce qui altère le modèle mosaïque intrinsèque. Ces changements de croissance se manifestent par une subluxation ou une dysplasie acétabulaire par une poussée vers le haut du labrum glénoïdien, un élargissement de la cavité et un élargissement de la tête. Si la tête dépasse complètement le bord du labrum glénoïdal, il s'agit d'une véritable luxation. Le labrum glénoïdien, non obstrué, revient par son élasticité à sa position correcte, entre la tête et l'acétabulum. La tête de la hanche luxée n'est plus stimulée par la croissance excessive observée dans la subluxation, de sorte qu'elle reste petite et ronde, bien qu'elle s'aplatisse sur le côté si elle reste en contact avec l'ilion. »

— Brand (2008)

4.2 La dysplasie développementale la hanche chez l'humain

4.2.1 Prévalence

La prévalence de la DDH chez l'humain varie considérablement en fonction des populations étudiées. Cette prévalence évolue de 0-0,1/1000 sub-Sahara et Hong Kong (Skirving & Scadden, 1979 ; Hoaglund et al., 1981), 0,5/1000 aux USA (0,6/1000 pour caucasiens, 0,2/1000 africains, 0,8 hispanique, 0,4 indiens) (Loder & Shafer, 2015), 6,7/1000 Israël (Stein-Zamir et al., 2008), 28/1000 Budapest (Czeizel et al., 1974), 67/1000 Navajo (Coleman, 1968). Ceci s'expliquerait en partie par la coutume d'emmailloter le nourrisson en abduction ou en adduction (Schwend et al., 2014 ; Kotlarsky et al., 2015 ; Loder & Shafer, 2015).

En Angleterre, cette prévalence ne semble pas avoir beaucoup évolué depuis le moyen-âge (Mitchell & Redfern, 2007). Cependant elle peut varier du simple au double en fonction des indices radiographiques utilisé (Jacobsen & Sonne-Holm, 2005) ainsi que du stade considéré 1-5/1000 pour la luxation et 25-50/1000 pour la subluxation et la dysplasie (Kotlarsky et al., 2015).

4.2.2 Développement

4.2.2.1 Normal

Le développement de l'articulation de la hanche chez l'humain est un processus complexe qui débute dès la 5^{ème} semaine de gestation à partir des cellules mésenchymateuses. Au cours des semaines suivantes, l'articulation prend forme, la masse de cellules cartilagineuses se sépare afin de former la tête fémorale et l'acétabulum vers la 7^{ème} et la 8^{ème} semaine. Ce processus est en partie orchestré par l'action des muscles entourant l'articulation, soulignant l'importance de la maturation simultanée du système neuromusculaire. L'articulation atteint sa formation complète vers la 11^{ème} semaine,

mais la tête fémorale reste entièrement cartilagineuse jusqu'au 4^{ème} mois, moment où le noyau d'ossification apparaît. Pendant cette croissance, le diamètre de la tête fémorale et de l'acétabulum croît plus rapidement que la profondeur de l'acétabulum, résultant en une augmentation progressive de l'instabilité de l'articulation. Cette instabilité serait compensée par le rôle mécanique du ligament rond pendant la vie fœtale et la première année de vie (Walker & Goldsmith, 1981).

Une rotation progressive de l'ensemble de l'articulation se produit, alors même que celle-ci n'est encore qu'un bourgeon mésenchymateux, partant d'une position latérale en abduction et pivotant vers une position parallèle à l'axe long du tronc (i.e., antéverson). Ce mouvement est complété par une rotation interne d'environ 90° (i.e., torsion), afin de permettre à la rotule et à la jambe d'être tournées vers l'avant. Facilité par l'action musculaire, ce mouvement se produit au cours du 3^{ème} mois et se poursuit après la naissance. Un timing précis et symétrique entre le développement de la tête fémorale et de l'acétabulum est crucial pour cette rotation et une formation adéquate (Brand, 2008).

Pour que l'articulation se développe normalement il est donc nécessaire de considérer le développement de l'articulation sur ses 3 dimensions spatiales ainsi que sur sa dimension temporelle. Un mauvais positionnement, une mauvaise orientation ou un défaut de temporalité peuvent résulter en un mauvais développement de l'articulation. Un développement normal de l'articulation coxo-fémorale nécessite ainsi qu'une tête fémorale soit stable et réduite de manière concentrique dans l'acétabulum et que ces deux éléments soient sphériques.

4.2.2.2 Anormal

Le développement anormal de l'articulation de la hanche chez l'humain peut conduire à divers phénotypes pathologiques. La dysplasie développementale de la hanche représente un spectre d'altérations comprenant, par ordre de gravité : dysplasie, hanche luxable, subluxation, luxation réductible, luxation non-opérable. Bien que la plupart des instabilités soient visibles à la naissance, certaines peuvent apparaître plus tard au cours de la première année de vie.

Il est toutefois possible de noter les différentes observations suivantes lors d'un développement anormal (voir figure 2) (Aronsson et al., 1994 ; Brand, 2008 ; Kayes & Didelot, 2012) :

- la tête fémorale (TF) ainsi que l'acétabulum diminuent de sphéricité, la tête fémorale s'aplatit (i.e., hypoplasie du noyau osseux de l'épiphyse de la tête fémorale) et l'acétabulum est moins profond. La pente du toit de l'acétabulum augmente ;
- la capsule articulaire s'hypertrophie tout comme le labrum et le ligament rond qui, de plus, s'allonge ;

- du tissu fibreux et adipeux (i.e., pulvinar) s'accumule dans l'espace proche de la paroi médiale de l'acétabulum ;
- par usure, le labrum peut alors se replier, s'émousser et il se déplace de plus vers le haut, couvrant certes moins la TF mais restant toujours au-dessus ;
- la TF se déplace latéralement, augmentant la distance entre l'épiphyse fémorale supérieure et le plancher acétabulaire. En réponse, les adducteurs se contractent, tout comme l'ilio-psoas, ce qui induit une constriction de la partie antérieure médiale de la capsule articulaire en forme de sablier. De même le ligament acétabulaire transverse se contracte.

4.3 La dysplasie coxo-fémorale chez le chien

4.3.1 Prévalence

Malgré une prévalence très variable chez le chien (e.g., 71 % chez les Bulldog, 10 % chez les chiens de Rhodésie à crête dorsale) (Genevois et al., 2008 ; Baldinger et al., 2020 ; OFA, 2023), les manifestations cliniques, anatomiques, radiologiques sont très similaires (Willemssen et al., 2021).

La pression de sélection opérée par les éleveurs et les clubs de races a permis de réduire considérablement l'incidence de la maladie dans certaines populations de chiens Epagneul Picard, Chow-Chow, Cane Corso, Rottweiler, chien de la Serra da Estrela, mais pour d'autres comme le Husky la situation s'est détériorée (Willis, 1997 ; Ginja et al., 2009b ; Genevois et al., 2008 ; Baldinger et al., 2020).

Tout comme chez l'humain, la prévalence dépend des scores ainsi que des catégories considérées comme atteintes (e.g., selon la FCI : D+E ou C+D+E) (Ohlerth et al., 2001 ; Baldinger et al., 2020). De plus certaines base de données peuvent être biaisées par la soumission plus probable des résultats des chiens indemnes que des chiens atteints (Paster et al., 2005). La prévalence réelle des races concernées pourrait donc être plus élevée.

Les races de grandes tailles sont plus affectées (Comhaire & Snaps, 2008 ; Roberts & McGreevy, 2010) mais toutes celles de grandes tailles ne sont pas forcément fortement atteintes (e.g., Hovawart, Husky ou Lévrier).

4.3.2 Développement

4.3.2.1 Normal

La cavité acétabulaire se met en place à la zone de jonction entre l'ilium, l'ischium et le pubis qui présentent un cartilage de croissance « triradié » (i.e., appartenant aux 3 éléments). Ce cartilage se ferme (i.e., s'ossifie) vers l'âge de 6 mois. Parallèlement, vers 16 semaines, on observe l'apparition de l'os coxae quartum, sous l'aspect d'un élément

osseux triangulaire, positionné latéralement au rebord acétabulaire dorsal avec lequel il fusionne vers 21 semaines pour compléter la cavité acétabulaire.

Pour ce qui est du fémur, à la naissance, la plus grande partie de la diaphyse fémorale est ossifiée alors que la partie proximale de l'os est cartilagineuse. Ce sont le ligament rond et la capsule articulaire qui assument la congruence articulaire pendant le premier mois de vie. Le cartilage de croissance proximal, qui participe à la croissance en longueur de l'os, est présent dès la naissance alors que le noyau d'ossification de la tête fémorale n'apparaît qu'entre la 2^{ème} et la 3^{ème} semaine. Le développement de la tête fémorale s'effectue par ossification endochondrale, cette croissance s'achevant vers 6 mois environ et un peu plus tard dans les races géantes. Le cartilage de croissance proximal s'ossifie vers 11 mois environ et un peu plus tard dans les races géantes.

C'est l'interaction entre la tête fémorale et l'acetabulum, par l'interaction des pressions respectives résultant d'un contact permanent, qui assure le développement harmonieux de l'articulation et la bonne congruence de cette dernière.

4.3.2.2 Anormal

Tout comme chez l'humain la description de la pathogénèse de la dysplasie de la hanche à bien évolué. Tout d'abord considéré comme un défaut de l'équilibre entre le développement des muscles et du squelette après la naissance (Alexander, 1992), ce serait désormais la laxité la cause primitive de son développement (Smith et al., 1990 ; Morgan, 1992 ; Lust et al., 1993 ; Ginja et al., 2008a). Cette laxité apparaît tardivement mais reste présente pendant toute la croissance de l'articulation (Riser & Shirer, 1966). Il est considéré que le chien naît avec des hanches normales (Alexander, 1992 ; Smith, 1998). La période critique pour le développement d'une hanche normale se situe donc entre la naissance et 60 jours de vie. Durant cette période, le système neuromusculaire est immature et les tissus conjonctifs sont malléables. Cependant cette élasticité est normalement limitée. L'excès de laxité permet la subluxation de l'articulation et augmente le stress induit par le déplacement du chiot pour l'allaitement et la marche. La congruence des surfaces articulaires diminue alors par l'augmentation des frottements entre celles-ci, usant le cartilage et déformant l'articulation aussi bien du côté fémoral qu'acétabulaire. Une dégénérescence progressive des tissus articulaires se mettrait ainsi en place avec formation d'ostéophytes périarticulaires (Brass, 1989 ; Alexander, 1992 ; Remedios & Fries, 1995 ; King, 2017).

Les différentes modifications anatomiques engendrées par la DCF sont les suivants selon (Alexander, 1992 ; Remedios & Fries, 1995) :

- le labrum s'émousse sur sa partie dorsale ;
- le ligament rond s'hypertrophie et est tiré ventralement à partir de son attachement au dôme dorsal de la fosse acétabulaire ;

- la partie dorsale du cartilage de la TF s'érode au contact du labrum ;
- la capsule articulaire s'épaissit ;
- la surface de contact entre la TF et l'acetabulum se réduit ;
- des micros fractures apparaissent sur le cartilage qui se déforme tant qu'il existe un contact entre la TF et l'acetabulum ;
- des ostéophytes se développent, aggravant l'érosion du cartilage ;
- retard de fermeture du cartilage de croissance proximal du fémur ainsi que de l'ossification de la tête fémorale.

La laxité serait permise par un défaut des différents stabilisateurs de l'articulation : la musculature fessière, le ligament rond, la capsule articulaire et le rebord acétabulaire dorsal. Le liquide synovial jouerait lui-aussi un rôle important dans la stabilisation. Sa surabondance est observé chez les chiens atteints par la DCF et ayant une plus forte laxité, mais il n'est pas certains que cela soit la cause et non la conséquence de la laxité (King, 2017). De plus le lien direct entre laxité et développement anormal de la hanche est bien établi pour les extrêmes, cependant il est difficile de déterminer ce qu'est une laxité « normale » ainsi que les conséquences d'une laxité « moyenne » (Remedios & Fries, 1995).

La dysplasie coxo-fémorale est également favorisée par la subluxation, mais la cause de celle-ci n'est pas encore bien établie. Différentes raisons non nécessairement exclusives ont été avancées. Ainsi un port de charge trop important associé à un excès de laxité déplacerait la tête fémorale latéralement, favorisant une subluxation. Cependant lorsque le poids est porté sur une hanche, de nombreux muscles la contraignent en position. D'autres chercheurs ont donc suggéré que ce serait durant la phase de balancement de la patte que le recrutement de muscles stabilisateurs ferait défaut en raison d'un trop grand volume de liquide synovial et donc d'une hypo-stimulation des mécanorécepteurs de la capsule. La patte serait alors en position subluxée lors de l'appui suivant. D'autres facteurs ont aussi été proposés comme contributeur à la DCF, comme une ossification retardée de la tête fémorale ainsi qu'une augmentation de la pente acétabulaire. Cependant pour ces deux derniers, il est difficile d'estimer s'ils sont la cause ou la conséquence d'un développement anormal de l'articulation (King, 2017). Il serait ainsi possible d'empêcher le développement de la maladie en maintenant la congruence de l'articulation jusqu'à ce que les abducteurs et les tissus de support soit suffisamment matures pour maintenir la tête fémorale dans l'acetabulum, aux alentours de 6 mois (Alexander, 1992).

Selon Frost (1989), cette structuration progressive de l'articulation est influencée par les trois éléments suivants :

- la génétique induisant l'architecture de l'articulation ainsi que des muscles et nerfs périphériques ;
- les forces externes induites par l'activité physique du chiot ;
- l'ossification du cartilage.

Le déplacement de la tête fémorale chez le chien est principalement cranio-dorsal, mais il peut parfois être postérieur lors d'un traumatisme.

4.4 Arthrose

L'arthrose est une maladie dégénérative des articulations caractérisée par une usure du cartilage articulaire et une formation de tissus osseux à la marge des surfaces articulaires. Elle est connue sous différents noms : en anglais *mechanical osteoarthritis*, *degenerative joint disease* ou encore *osteoarthritis*. Elle se traduit par des douleurs, une raideur, une hydarthrose, des crépitations et une diminution de la mobilité de l'articulation concernée. L'arthrose peut être aggravée par un processus inflammatoire s'intensifiant avec les contraintes mécaniques sur les articulations (Vaughan-Scott & Taylor, 1997 ; Musumeci et al., 2015). C'est une boucle de rétroaction négative à la fois mécanique, structurale et moléculaire. Ainsi la surcharge mécanique augmente la dégradation du cartilage, augmentant la laxité ainsi que la synovite, diluant l'acide hyaluronique du liquide synovial. Celui-ci perd alors en viscosité, absorbant de moins en moins la charge mécanique et augmentant les contraintes sur le cartilage. De plus, la surcharge mécanique stimule l'expression de médiateurs inflammatoires qui vont favoriser la production d'enzymes dégradant les différents tissus de l'articulation (voir figure 2) (Boehme & Rolaufts, 2018 ; Meeson et al., 2019 ; Tomé et al., 2023). Une fois la cascade inflammatoire débutée, l'arthrose ne peut être guérie. Les traitements actuels consistent en une lutte contre la douleur et une amélioration des signes cliniques (Vaughan-Scott & Taylor, 1997). Bien que se développant progressivement avec l'âge, il est possible d'observer de l'arthrose dès l'âge de 7 à 8 mois chez le chien et vers 20 ans chez l'humain.

L'arthrose peut être classifiée selon les catégories suivantes (Meeson et al., 2019) :

- secondaire à une instabilité acquise chez le juvénile (DCF + DDH) ;
- secondaire à une instabilité acquise chez l'adulte (e.g., rupture du ligament croisé antérieur) ;
- défaut de développement vasculaire : Legg Perthes Calvé ;
- défaut de développement endochondral : ostéochondrite disséquante / Maladie de König ;

- environnementale : obésité ou traumatisme ;
- maladie inflammatoire : polyarthrite rhumatoïde, syndrome de Reiter, etc.

Chez le chien, il y a très peu d'arthrose spontanée (i.e., primaire) tout comme chez l'humain selon les nouvelles théories (Vaughan-Scott & Taylor, 1997 ; Hinton et al., 2002 ; Ganz et al., 2008). La dysplasie de la hanche est une des causes majeures d'apparition d'arthrose coxo-fémorale aussi bien chez l'humain (Jacobsen & Sonne-Holm, 2005 ; Reijman et al., 2005 ; Jacobsen et al., 2009) que chez le chien (Runge et al., 2010). Cela explique l'interconnexion dans son diagnostic avec la présence d'arthrose dans certains scores composites utilisés pour mesurer la sévérité de la dysplasie (Corley et al., 1997). Cependant il n'y a aucune relation directe entre l'arthrose et la sévérité de la dysplasie ni même entre les signes cliniques de l'arthrose et la sévérité radiographique de l'arthrose, chez le chien comme chez l'humain (Tönnis, 1976). De plus le développement de l'arthrose varie en fonction des races (Smith et al., 2001).

4.5 Conclusion

En résumé, la prévalence de la DDH chez l'humain et de la DCF chez le chien varie considérablement selon les populations. Chez l'humain et le chien, la dysplasie de la hanche partage des mécanismes de développement similaires, impliquant des facteurs génétiques et environnementaux. La pathogénèse de ces anomalies du développement englobe de nombreux mécanismes dont des anomalies dans le développement du bourgeon du membre pendant une période critique entraînant des défauts structurels, un possible retard dans l'innervation musculaire, la laxité articulaire, un rôle de l'arthrose précoce, etc. La grande difficulté de cette maladies complexe du développement est que causes et conséquences s'interconnectent.

5 Diagnostic

Chez l'humain comme chez le chien, le diagnostic de la dysplasie de la hanche s'établit sur la présence conjointe d'un panel d'indicateurs cliniques et d'imagerie.

5.1 Examen clinique

5.1.1 Chez l'humain

L'examen physique chez l'humain consiste en des observations anatomiques ainsi qu'un examen dynamique fait de la naissance à 3 mois. Des contraintes sont appliquées sur l'articulation afin de tester sa stabilité.

Cette examen comprend (Kotlarsky et al., 2015) :

- le volant d'abduction ;
- l'asymétrie des plis, qui est remis en question quant à sa prédictivité (Kang et al., 2019) ;
- le piston ou ressaut de réintégration (Barlow, 1962 ; Ortolani, 1948) ;
- une différence de longueur de jambe (test de Galeazzi ou d'Alli) ;
- démarche Trendelenburg.

Cet examen clinique est complété dans le cas de signe dysplasique ou d'antécédents familiaux par un examen échographique. Celui-ci est complexe, plus particulièrement la méthode de Graph, cette technique étant opérateur dépendant. Il est donc nécessaire de comparer l'échographie et la clinique. En fonction des caractéristiques de la DDH celle-ci serait plus ou moins compliquée à dépister. Il est plus évident de dépister à l'examen clinique une luxation unilatérale raide qu'une bilatérale souple chez le tout petit (Vitale & Skaggs, 2001).

L'âge au diagnostic est important, en effet le signe d'Ortolani et de Barlow ne sont plus prédictifs à partir de 3 mois de vie (Kotlarsky et al., 2015).

5.1.2 Chez le chien

Chez le chien, les signes d'appels de présence de dysplasie apparaissent tardivement. L'examen clinique doit se faire à différents moments : au repos, pendant une activité et après afin de mettre en évidence une potentielle boiterie.

De même il est possible d'observer de la boiterie à l'effort, une amyotrophie de la hanche, une réduction de l'amplitude de mouvement. La douleur durant l'extension ou la flexion se caractérise par un refus de trot (allure qui sollicite chaque membre postérieur individuellement), une course en saut de lapin, ou encore par une difficulté à monter des escaliers ou dans une voiture, etc.

Il faut cependant noter qu'un individu présentant une dysplasie prononcée peut tout de même présenter une démarche normale et ne pas présenter de signes cliniques (Barr et al., 1987).

Les tests dynamiques développés dans un premier temps en médecine humaine ont été adaptés en médecine vétérinaire. On retrouve ainsi le test d'Ortolani et celui de Barlow. Certains tests physiques ont cependant été spécialement conçus pour le chien comme celui de Bardens (Ginja et al., 2010).

5.2 Imagerie

L'imagerie vient en appui de l'examen clinique, elle permet de visualiser et de quantifier finement la sévérité de l'atteinte de l'articulation.

5.2.1 Radiographie et Tomodensitométrie

5.2.1.1 Chez le chien

L'ensemble des différentes mesures faites chez le chien sont représentées en figure 4.

5.2.1.1.1 Vue ventro-dorsale standard avec hanche étendue (A)

Le chien est placé en décubitus dorsal sur la table de radiologie, les membres postérieurs étendus parallèlement l'un par rapport à l'autre et les grassettes en rotation interne afin de mettre les rotules au zénith (Ginja et al., 2008b). Cette vue permet de détecter facilement un début d'arthrose chez le chien mais elle manque de sensibilité dans la détection de la laxité coxo-fémorale (Smith et al., 1990 ; Broeckx et al., 2018). Bien que non physiologique chez le chien, elle permet d'évaluer la couverture crâniale et le degré de couverture dorsal de la tête fémorale. Cette vue ventro-dorsale est actuellement le standard international dans le cadre du dépistage de la dysplasie de la hanche chez le chien.

Plusieurs mesures sont faites sur cette vue :

- **angle de Norberg-Olson (A1)** : angle obtenu par une ligne joignant les centres des 2 têtes fémorales et la ligne passant par le centre de la tête fémorale et le point de rencontre entre le rebord acétabulaire crânial et le rebord acétabulaire dorsal du même côté. Pour une hanche normale, l'angle a une valeur supérieure à 105° (Comhaire et al., 2009).
- **angle de couverture crâniale (A2)** : dénommé *center-edge angle* en anglais, il s'agit de l'angle formé par deux demi-droites partant du centre de la tête fémorale, l'une parallèle à l'axe sagittal du bassin et l'autre passant par le rebord crânial de l'acétabulum (Mostafa et al., 2022).
- **pente crâniale du toit acétabulaire (A3)** : angle mesurant la pente obtenue entre une droite reliant la limite médiale et la limite latérale de la zone de condensation

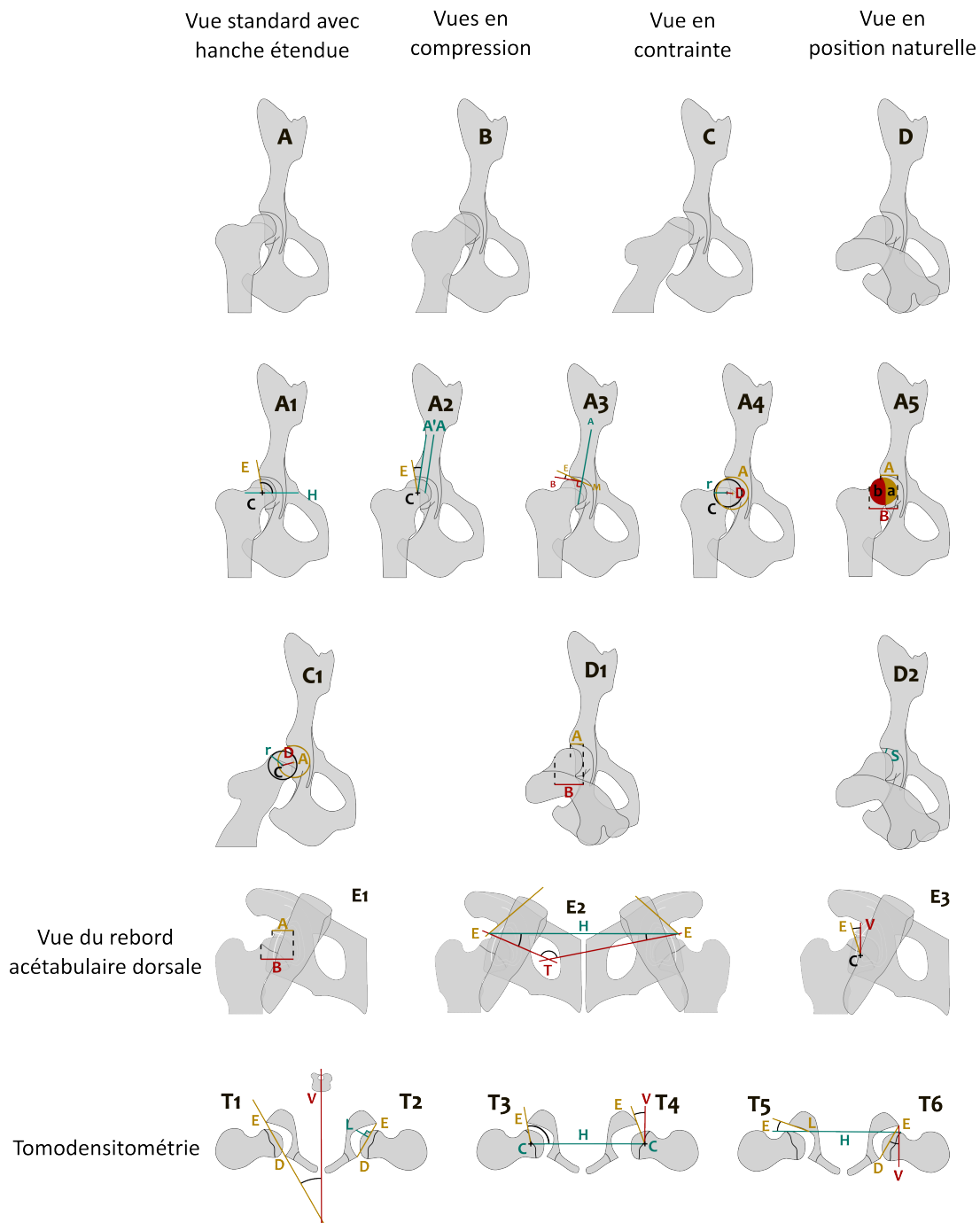


FIGURE 4 - Représentations graphiques des différentes mesures d'imagerie faites chez le chien

(A) Radiographie de la hanche d'un chien en position ventrodorsale avec les différentes mesures utilisées pour évaluer la dysplasie coxo-fémorale : angle de Norberg-Olsson (1), angle de couverture crâniale (2), pente crâniale du toit acétabulaire (3), indice de distraction (4), pourcentage de couverture dorsale par distance (A/B) ou par aire (a/b) (5). (B) Radiographie de la hanche d'un chien en position ventrodorsale avec compression médial des hanches. (C) Radiographie de la hanche d'un chien en position ventrodorsale avec abduction latérale des hanches avec distracteur : indice de distraction (1). (D) Radiographie de la hanche d'un chien en position debout avec le faisceau perpendiculaire au plan du bassin : pourcentage de couverture dorsale par distance (A/B) (1), distance S (2). (E) Radiographie de la hanche d'un chien en position debout avec le faisceau longitudinal au plan du bassin : pourcentage de couverture dorsale par distance (A/B) (1), pente dorsale du toit acétabulaire (2), angle acétabulaire (3). (T) Tomodensitométrie transversale d'une hanche de chien : angle acétabulaire (1), indice acétabulaire (2), angle sectoriel acétabulaire dorsal (3), angle centre-bord latéral / angle de Wiberg (4), angle du rebord acétabulaire dorsal (5), angle d'antéversion acétabulaire (6).

de l'acétabulum et la droite perpendiculaire à l'axe sagittal du bassin (Mostafa et al., 2022).

- **indice de distraction (A4)** : Distance entre le centre de la tête fémorale et le centre de l'acétabulum divisé par le rayon de la tête fémorale (Smith et al., 1990).
- **pourcentage de couverture dorsale (A5 / D1 / E1)** : aussi connu sous le nom de *Dorso Lateral Subluxation score* en anglais, ce pourcentage représente la couverture dorsale de la tête fémorale par l'acétabulum. Il est calculé en rapportant la distance de la tête fémorale médiane recouverte par le bord acétabulaire crânial sur le diamètre de la tête fémorale (Schachner & Lopez, 2015). Il existe deux manières de l'obtenir, soit sur une projection radiographique du rebord dorsal acétabulaire soit en vue standard avec hanche étendue. Dans ce dernier cas il faut prendre comme distance le rebord dorsal de l'acétabulum vu par transparence au travers de la tête fémorale (Mostafa et al., 2022). Une autre mesure liée à ce rapport est le rapport de surface de la tête fémorale recouverte par le rebord dorsal de l'acétabulum (Mostafa et al., 2022).

5.2.1.1.2 Vue en compression (B)

L'animal est en décubitus dorsal avec les fémurs perpendiculaires à la table, l'opérateur tient les tibias (eux-mêmes perpendiculaires aux fémurs). Une pression externe, dirigée médialement, est exercée à la hauteur des articulations coxo-fémorales. Elle entraîne la mise en compression de la tête fémorale dans l'acétabulum. On visualise ainsi le maximum de congruence qu'il est possible d'obtenir entre la tête du fémur et sa cavité articulaire. Lorsque la congruence est parfaite, le centre de la tête fémorale est superposé au centre de l'acétabulum. Cette vue est utilisée afin d'observer la congruence de l'articulation et dans le cadre de la mesure de la laxité articulaire (i.e., mesure de l'indice de distraction).

5.2.1.1.3 Vue en contrainte (C)

Afin de pallier le manque de sensibilité de la radiographie standard pour la mise en évidence de la laxité articulaire, deux méthodes sont actuellement utilisées : une, sous brevet, nommée *PennHip* et une autre, libre de droits nommée méthode *Vezzoni*. Ces deux méthodes fonctionnent de la manière suivante. L'individu est placé de la même manière que pour une vue en compression, puis un dispositif (i.e., le « distracteur »), permettant l'écartement des membres postérieurs, est placé entre les cuisses de l'animal. Les deux postérieurs, manipulés à la hauteur des jarrets, sont rapprochés l'un vers l'autre. Le distracteur exerce alors un contre-appui en zone proximale des fémurs, ce qui déplace les têtes fémorales latéralement, mettant ainsi en évidence la laxité de l'articulation (Smith et al., 1990 ; Broeckx et al., 2018).

- **indice de distraction (C1)** : L'indice de distraction a été initialement décrit dans la méthode PennHip. Il permet d'obtenir une évaluation de la laxité articulaire dès 16 semaines de vie et permet de prédire la sévérité de l'arthrose à 2 ans (Smith et al., 1990 ; Runge et al., 2010). Ce procédé, breveté aux USA n'est que très peu utilisé en France. Un dispositif similaire a été développé par Vezzoni (Broeckx et al., 2018). L'indice de distraction est de plus en plus utilisé et correspond au rapport, mesuré sur une radiographie « en contrainte », entre la distance séparant le centre géométrique de la tête fémorale et le centre de la cavité acétabulaire divisée par le rayon de la tête fémorale correspondante. Plus l'indice est proche de zéro, plus la coaptation de la tête fémorale dans la cavité articulaire est bonne et moins la laxité articulaire est élevée. L'interprétation de cet indice peut être combiné avec les modifications ostéo-articulaires engendrées par la DCF suivant le tableau S2.

5.2.1.1.4 *Vue en position « naturelle » (D et E)*

L'animal est laissé debout en position naturelle tandis qu'un porte-cassette est placé en position caudale par rapport à la queue et aux membres postérieurs. La cassette et la grille sont maintenues perpendiculaires à l'angle correspondant à l'inclinaison du pelvis, ce qui permet de diriger le faisceau parallèlement à l'inclinaison du bassin (Slocum & Devine, 1990).

Des variantes consistent à placer, l'individu en décubitus ventral sur une table et à mettre des supports au niveau des membres postérieurs du chien afin de bien les maintenir à la verticale (Ogden et al., 2012).

La direction du faisceau de rayon X peut alors se faire perpendiculairement à l'axe du bassin (D) ou bien longitudinalement permettant ainsi d'observer le rebord acétabulaire dorsal (E).

- **pourcentage de couverture dorsale (A1 / D1 / E1)** : voir A1 ;
- **distance S (D2)** : Distance la plus courte entre l'aspect cranio-ventro-médial de l'acétabulum et la tête fémorale (Ogden et al., 2012) ;
- **pente dorsale du toit acétabulaire (E2)** : angle mesurant la pente obtenu entre la tangente au rebord acétabulaire et la droite passant par les intersections entre les tangentes et les droites passant par l'aspect dorsal du rebord acétabulaire. Les angles des deux hanches peuvent aussi être combinés en prenant l'angle formé par le croisement des deux tangentes (Charette et al., 2001) ;
- **angle acétabulaire (E3)** : similaire dans sa construction à l'angle de couverture crâniale, mais en plus du changement de vue, la droite de référence est parallèle au plan sagittal.

5.2.1.1.5 Tomodensitométrie (T)

La tomodensitométrie (TDM) est une technique d'imagerie en coupe qui permet d'obtenir des images en 2D ou 3D de l'articulation. Elle est peu utilisée chez le chien pour évaluer la dysplasie de la hanche mais les mesures qui peuvent être faites sont similaires à celles faites chez l'humain.

- **angle acétabulaire (T1)** : angle formé par une ligne passant par les extrémités médiale et latérale de l'acétabulum et une ligne perpendiculaire à l'axe sagittal du bassin ;
- **indice acétabulaire (T2)** : pourcentage obtenu en divisant la distance entre les points D et E (DE) par la largeur de l'acétabulum (L) et en multipliant le résultat par 100 (Lopez et al., 2008) ;
- **angle sectoriel acétabulaire dorsal (T3)** : angle formé par l'axe horizontal du pelvis et la droite entre le centre de la tête fémorale et le bord dorsolatéral de la surface lunaire (Lopez et al., 2008) ;
- **angle centre-bord latéral / angle de Wiberg (T4)** : angle formé par deux lignes passant par le centre de la tête fémorale, l'une est verticale et l'autre passe par le rebord latéral du labrum (Lopez et al., 2008) ;
- **angle du rebord acétabulaire dorsal (T5)** : angle formé par une ligne tracée du bord médial du sourcil acétabulaire jusqu'au bord latéral du rebord acétabulaire dorsal et par une ligne passant par le bord médial du sourcil, prolongée latéralement et parallèle à l'axe horizontal pelvien (Lopez et al., 2008) ;
- **angle d'antéversion acétabulaire (T6)** : angle formé par une ligne passant par les extrémités antérieure et postérieure de l'acétabulum et une ligne perpendiculaire à l'axe horizontal du pelvis (Schachner & Lopez, 2015).

5.2.1.2 Chez l'humain

Chez l'humain, la radiographie n'est plus utilisée comme méthode de dépistage. En effet sa fiabilité est moins bonne avant que le centre de la tête fémorale ne s'ossifie, aux alentours de 4 à 6 mois. Elle est donc utilisée comme outil de contrôle de l'évolution. Les différentes mesures faites chez l'humain sont représentées en figure 5.

5.2.1.2.1 Bassin de face ou antéro-postérieur

Le patient est en décubitus dorsal ou bien debout, les membres inférieurs sont soumis à une rotation interne de 15° à 25° par rapport à la hanche.

- **angle de couverture externe ou VCE (A1)** : aussi nommé angle de recouvrement du toit de l'acétabulum, angle de Wiberg et angle de couverture latérale, il est formé par deux droites passant par le centre de la tête fémorale lors d'une radio du bassin de face. L'une est verticale et l'autre passe par l'extrémité latérale du toit acétabulaire (ligne dense). Il vaut plus de 25° pour une hanche normale et semble être le même angle que celui de Norberg Olsson d'où un angle droit aurait été soustrait. Cet angle permet de situer la tête fémorale par rapport à l'acétabulum et, lorsque la tête fémorale n'est pas déplacée, donne une indication du degré de développement du toit de l'acétabulum (Wiberg, 1939).
- **Angle acétabulaire / Sharp Angle / HTE (A2)** : angle formé par la ligne passant par le rebord latéral de la cavité acétabulaire et la pointe inférieure du U radiologique (*Köhler teardrop*) et par une ligne horizontale reliant les deux cartilages tri-radiés (ou les 2 pointes inférieures du U radiologique) (Sharp, 1961) ;
- **angle d'obliquité du toit de l'acétabulum / angle de Tönnis / indice acétabulaire / angle HME (A3)** : angle formé par deux lignes passant par l'extrémité médiale du toit acétabulaire (point interne). L'une est horizontale et la deuxième passe par l'extrémité latérale du toit acétabulaire. Une hanche normale aura une valeur inférieure ou égale à 11° (Tönnis, 1976).
- **indice fémoro-épiphyseal du toit acétabulaire / fémoro-épiphyseal acetabular roof index (A4)** : angle formé entre la ligne passant par les extrémités médiale et latérale de l'acétabulum et la ligne passant par la partie rectiligne médiale de la cicatrice physéenne de la tête fémorale (Wyatt et al., 2017).
- **arche gothique / Gothic Arc Angle (A5)** : l'arche gothique est formée par trois courbes : la base est formée par le sourcil de l'acétabulum, l'arc médial est formé par la ligne partant du rebord latéral de l'acétabulum vers l'articulation sacro iliaque, l'arc latéral est constitué d'un arc d'os spongieux dense s'étendant de la plaque quadrilatérale vers les épines iliaques antéro-supérieure et antéro-inférieure. Cet angle est formée par une droite passant par le centre de la tête fémorale et la

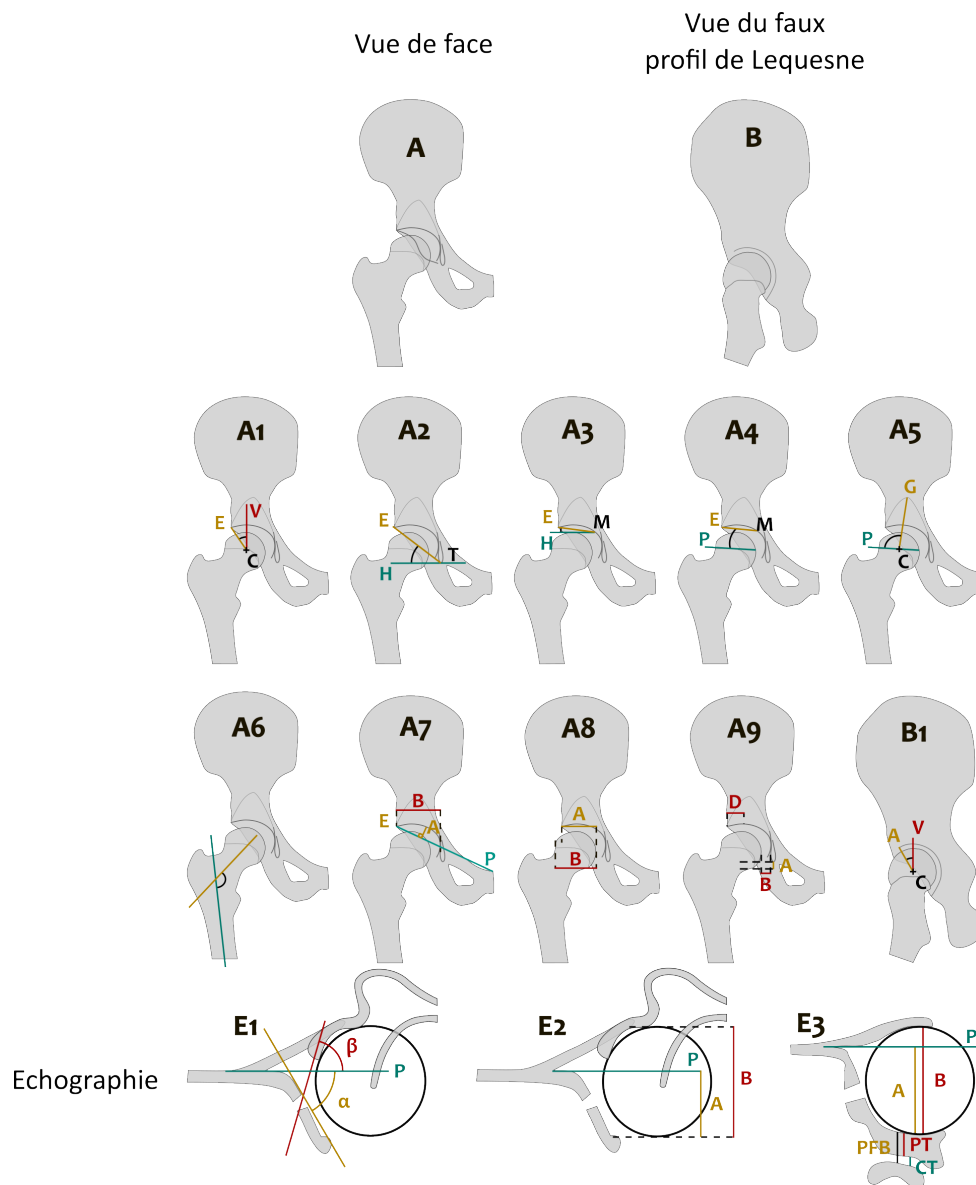


FIGURE 5 - Représentations graphiques des différentes mesures d'imagerie faites chez l'humain.

(A) Radiographie de la hanche chez l'humain de face : angle de couverture externe / VCE (A1), angle acétabulaire / Sharp angle / HTE (A2), angle d'obliquité du toit de l'acétabulum / angle de Tönnis / indice acétabulaire / angle HME (A3), indice fémoro-épiphyseal du toit acétabulaire / fémoro-épiphyseal acetabular roof index (A4), arche gothique (A5), angle cervico diaphysaire (A6), profondeur de l'acétabulum (A7), indice d'extrusion de la tête fémorale de Heyman et Herndon / Reimers migration index (A8), distance de subluxation latérale et supérieure (A9). (B) Radiographie de la hanche chez l'humain en vue de faux profil de Lequesne : angle de couverture antérieure / VCA (B1). (E) Échographie de la hanche chez l'humain : angle alpha et beta (E1), couverture acétabulaire (E2), distance pubo-fémorale (E3).

pointe de l'arche gothique et par la droite passant par la partie rectiligne médiale de la cicatrice physéenne de la tête fémorale (Zimmerer et al., 2021).

- **Angle cervico diaphysaire (A6)** : angle formé par l'axe du col fémoral et l'axe de la diaphyse. Il vaut entre 135°-140° pour une hanche normale. Cependant cet angle est très variable entre individus et varie en fonction de l'âge (Tönnis, 1976).
- **profondeur et rapport de profondeur de l'acétabulum (A7)** : distance entre le point le plus profond de la cavité acétabulaire et la ligne EP, où E est le bord latéral de la ligne dense sous chondrale du toit du cotyle et P l'angle supérieur du pubis (Murray, 1965). Pour une hanche normale, la profondeur de l'acétabulum est supérieure à 9 mm. Le rapport est quant à lui obtenu en divisant la profondeur de l'acétabulum par sa largeur.
- **indice d'extrusion de la tête fémorale de Heyman et Herndon / Reimers migration index (A8)** : pourcentage obtenu en divisant la distance horizontale de la tête fémorale latérale recouverte par l'acétabulum par la largeur totale de la tête fémorale > 25 % (Jacobsen & Sonne-Holm, 2005).
- **Distance de subluxation latérale et supérieure (A9)** : distance correspondant à celle séparant le bord latéral du U radiologique au rebord médial de la tête fémorale. La distance de subluxation supérieure est quant à elle mesurée entre la portion inférieure de la tête fémorale et la portion inférieure de l'acétabulum ischial (Murphy et al., 1995).

5.2.1.2.2 *Faux profil de Lequesne*

Le patient est debout, le côté à radiographier contre la table placée à la verticale. Le côté opposé est repoussé en arrière de telle façon que le plan du dos fasse 60° avec le plan de la table verticale.

- **Angle de couverture antérieur / central edge angle / VCA (B1)** : angle formé par deux droites passant par le centre de la tête fémorale. L'une est verticale et l'autre passe par l'extrémité antérieure du toit acétabulaire (ligne dense). Il vaut plus de 25° pour une hanche normale.

5.2.1.3 Facteurs influençant le résultat d'imagerie

5.2.1.3.1 Anesthésie

Dans les différents protocoles de dépistage de la dysplasie de la hanche chez le chien par radiographie, il est demandé d'anesthésier l'individu ou d'utiliser une sédation profonde afin d'avoir un relâchement maximal des muscles. Cela permet d'avoir des résultats répétables. Les mesures ainsi obtenues ne prennent pas en compte les éventuelles compensations, chez le chien vigile, de la laxité articulaire par le tonus musculaire, qui expliquent l'absence de boiterie chez des chiens notés comme atteints. La molécule anesthésiante utilisée doit être notée sur le compte rendu radiographique. En effet toutes les molécules n'induisent pas un relâchement aussi important des muscles (Maitre et al., 2010 ; Bozkan & Sarierler, 2019). Une seule étude n'a pas montré de différence significative (Aronson et al., 1991).

5.2.1.3.2 Lecteurs

Les différentes mesures sont influencé par l'expérience des lecteurs. La corrélation entre différents lecteurs n'est d'ailleurs pas toujours bonne et reste grandement à améliorer (Verhoeven et al., 2008).

5.2.1.3.3 Qualité des radios

La qualité des radios influe sur la fiabilité de la mesure. La résolution musculaire et un bon positionnement de l'animal sont primordiaux pour pouvoir évaluer la hanche (Genevois et al., 2007 ; Verhoeven et al., 2010).

5.2.1.3.4 Fiabilité des méthodes

Malgré la standardisation de la méthode FCI, la fiabilité de mise en évidence de la laxité est imparfaite, ce qui va de pair avec un pourcentage non négligeable de faux-négatifs.

5.2.1.3.5 Populations

Le seuil de l'angle de Norberg-Olsson est soumis à critique (Culp et al., 2006 ; Gaspar et al., 2016) et pourrait être différent en fonction des races (Tomlinson & Johnson, 2000).

5.2.1.4 Amélioration

De nombreuses études se penchent récemment sur la possibilité d'analyser les radiographies de dépistage par le biais de modèles en intelligence artificielle (Ghasseminia et al., 2022). Pour le moment ces recherches se font essentiellement chez l'humain mais commencent aussi à apparaître dans le domaine canin malgré les conflits politiques que cela implique. Les modèles se concentrent actuellement sur des méthodes de réseaux neuronaux en *Deep Learning*.

5.2.2 Échographie

5.2.2.1 Chez le chien

L'échographie chez le chiot entre 2 et 9 semaines, bien que techniquement faisable, ne semble pas prédictive du développement de la DCF à l'âge adulte (Adams et al., 2000 ; Fischer et al., 2010), bien qu'une première étude en 1997 ait mis en évidence une association entre la distance maximale de distraction entre 6 et 9 semaines et une classification anormale de l'articulation à 1 an (O'Brien et al., 1997).

5.2.2.2 Chez l'humain

L'échographie est chez l'humain la méthode de diagnostic de référence. Elle permet d'aisément observer le cartilage par contraste avec les tissus mous ainsi que de faire un examen dynamique à l'inverse d'une radiographie figée. De fait, de nombreuses coupes et positions peuvent être pratiquées (Harcke & Grissom, 1990). Elle est utilisée pour le dépistage dans de nombreux pays bien que son utilisation soit associée à des résultats variables à travers différentes études. Son application généralisée comme ciblée ne semble pas réduire la prévalence des cas de dysplasie détectés tardivement ou nécessitant une chirurgie par rapport à un examen physique (Shorter et al., 2011). Son utilisation est donc encore aujourd'hui sujet à débat (Dezateux & Rosendahl, 2007). Actuellement l'échographie est utilisée dès l'âge d'un mois sans anesthésie. De nombreuses mesures ont été développées au fur et à mesure des années, avec une utilisation variable en fonction des pays. L'objectif de cette méthode est de pouvoir observer à la fois la morphologie de l'articulation ainsi que sa stabilité (Rosendahl et al., 1992).

Elle permet d'obtenir les mesures suivantes :

- **angle alpha (E1)** : angle formé par une ligne tangente au toit acétabulaire et une ligne passant le long du rebord latéral de l'ilium (Graf, 1984). Un angle supérieur à 60° est considéré comme normal.
- **angle beta (E1)** : angle formé par une ligne passant le long du rebord latéral de l'ilium ligne et une ligne tangente au toit cartilagineux passant par le sommet du labrum (Graf, 1984). Un angle inférieur à 55° est considéré comme normal.
- **couverture acétabulaire (E2)** : pourcentage de la tête fémorale recouverte médialement par l'acétabulum.
- **fond du toit acétabulaire / distance pubo-fémorale (E3)** : distance séparant le bord médial de la tête fémorale et le pubis. Cette distance correspond à l'épaisseur du pulvinar et du cartilage du pubis (Tréguier et al., 2013).

5.3 Scores composites

5.3.1 Chez le chien

La façon la plus répandue d'évaluer la sévérité de la DCF est de suivre la procédure de la Fédération Cynologique Internationale (FCI). Le bassin du chien est radiographié dans une position standardisée avec les membres postérieurs en extension. Dans certains pays, une seconde radiographie est effectuée avec les membres en abduction. Un score est ensuite créé à partir de l'angle de Norberg-Olsson (NO) et de paramètres subjectifs de la conformation de la hanche, prenant en compte la qualité de l'acétabulum et de la tête du fémur ainsi que la congruence de ces deux parties (CAcTf) et la présence d'arthrose. Le score est une échelle allant de A à E, pour plus de précision voir tableau S3 (Flückiger, 2007b ; Commission & Committee, 2022). Particularité allemande, le score FCI est divisé en deux sous-catégories 1 et 2.

Aux États-Unis la structure *Orthopedic Foundation for Animals* (OFA) classe la DCF selon la grille présentée dans le tableau S4. Ce score est basé sur une évaluation subjective de la conformation de l'articulation. Il n'inclut pas l'angle de Norberg-Olsson mais prend en compte la congruence articulaire, la profondeur de l'acétabulum, la couverture de la tête fémorale et la présence d'arthrose. Le score est divisé en 7 catégories : *Excellent, Good, Fair, Borderline, Mild, Moderate et Severe*.

En Angleterre et en Australie, le *British Veterinary Association* (BVA) et le *Australian Veterinary Association* (AVA) ont mis en place un système de score numérique (Flückiger, 2007b). Ce score est basé sur 9 items radiographiques (voir tableau 3). Chaque mesure est notée entre 0 et 6, le score total est la somme des 9 mesures et additionné pour les deux hanches, il peut ainsi aller de 0 à 108. Un score faible est synonyme d'une bonne hanche.

Enfin, en Suisse, un système de score numérique similaire a été développé par (Flückiger, 2007a) (voir tableau S6). Ce score est basé sur 6 items radiographiques. Chaque mesure est notée entre 0 et 5, le score total est la somme des 6 mesures et la valeur la plus élevée des deux hanches est retenue, il peut ainsi aller de 0 à 30. Un score faible est synonyme d'une bonne hanche.

Bien que reposant sur des méthodologies différentes, les diverses classifications de la DCF à travers le monde restent globalement comparables entre elles (voir tableau 1).

TABLE 1 - Correspondance des différents scores composites utilisés dans le monde pour l'évaluation de la DCF chez le chien

Les scores présentés ici, sont ceux utilisés par la FCI : Fédération Cynologique Internationale (Europe), l'Allemagne, l'OFA : Orthopedic Foundation for Animals (États-Unis), la BVA : British Veterinary Association (Angleterre et Australie) et la Suisse. Source : Flückiger (2007b).

FCI	Allemagne	OFA	BVA	Suisse
A	A1	Excellent	0-4	0
	A2	Good	5-7	1-2
B	B1	Good	8-10	3-4
	B2	Fair	11-18	5-6
C	C1	Borderline	19-25	7-9
	C2	Mild	26-35	10-12
D	D1	Moderate	36-43	13-15
	D2	Moderate	44-50	16-18
E	E1	Severe	51-78	19-21
	E2	Severe	79-106	22-24

5.3.2 Chez l'humain

5.3.2.1 Graf

Cette classification a été et est encore dans plusieurs pays, le diagnostic de référence pour la DDH (Graph et al., 2006) du fait de sa forte sensibilité et spécificité (Terjesen, 1996 ; Gokharman et al., 2019). Cependant cette méthode est de plus en plus décriée et de nombreuses variations ont été élaborées afin de pallier les limitations de cette méthodologie (Harcke & Grissom, 1990 ; Terjesen, 1996 ; Dezateux & Rosendahl, 2007). Elle repose sur les valeurs des angles alpha et bêta et la morphologie de l'articulation (voir tableau S7).

5.3.2.2 Classification de Tönnis (Arthrose)

Classification représentant un gradient de 3 degrés des changements dégénératifs de la hanche (voir tableau S8) (Busse et al., 1972).

5.4 Biomarqueurs chez le chien

5.4.1 Liquide synovial

A l'heure actuelle les biomarqueurs identifiés dans le liquide synovial sont associés à la présence d'arthrose. Cependant, comme mentionné dans le chapitre sur l'arthrose, cette affection est, chez le chien, essentiellement dûe à la présence d'une dysplasie coxo-fémorale sous-jacente. Ces marqueurs pourraient donc aussi être utilisé pour la DCF. Les différents biomarqueurs identifiés peuvent être regroupé selon les catégories suivantes : médiateurs pro-inflammatoire (cytokines et protéine C-réactive), enzymes et inhibiteurs (métalloprotéinases de la matrice, inhibiteur tissulaire des métalloprotéinases, phosphatase acide résistante aux tartrates), produits extracellulaires de la dégénérescence du cartilage (produit de dégradation et de synthèse du collagène II, protéoglycanes, protéine de la matrice oligomérique du cartilage, tenascin-C) (Fujita et al., 2005 ; Bakker et al., 2017 ; Herr et al., 2021).

5.4.2 Sérum et Urine

Dans l'étude de (Ahner et al., 2019), 4 biomarqueurs présents dans l'urine ou le sérum permettraient de discriminer les individus atteints des non-atteint (CTX-I and II, sérum MMP-9, and sérum PIICP). Des marqueurs similaires ont été trouvés par l'étude de (Nganvongpanit et al., 2008) dans le sérum. Des marqueurs dans l'urine de chiots mâles de 5 mois pourraient aussi discriminer les individus atteints avec le score OFA mais l'étude doit être reproduite avec un plus grand panel (Bozynski et al., 2020).

5.5 Tests génétiques chez le chien

Deux tests génétiques de risque ont été mis au point pour le dépistage de la dysplasie de la hanche chez le chien. Le premier basé sur les résultats de (Foundation of the University of Veterinary Medicine Hannover, 2009) tout comme le deuxième basé sur ceux de (Bartolomé et al., 2015) ont été critiqués sur leurs fiabilités en dehors des populations canines utilisées pour la création des tests (Manz et al., 2017 ; Bruun et al., 2020). Ils sont désormais retirés du marché. D'autres tests sont actuellement sur le marché, certains se basent sur le gène prometteur *FBN2* mais ne sont basés sur aucune publication scientifique.

5.6 Comparaison des mesures et scores en médecine vétérinaire

DLS et angle de couverture antérieur sont légèrement négativement corrélée avec l'hyperlaxité de l'articulation mais pas avec la dysplasie de la hanche en elle-même (Ginja et al., 2009a). De nombreuses études ont été menées afin de comparer les différentes mesures disponibles chez le chien. Faire un résumé de l'ensemble de celles-ci est complexe, les scores utilisés ne sont pas toujours les mêmes et que doit-on prendre comme valeur de référence afin d'établir la spécificité et la sensibilité de chaque indice ou de chaque score ? Il est cependant possible de les comparer (voir tableau 2 et 3).

TABLE 2 - Comparaison des différentes mesures et scores pour la détection de la DCF

Abbreviations : acétabulum / acétabulaire (Ac), tête fémorale (TF), angle de Norberg-Olsson (NO).

Mesure	FCI	NO	CAD	DAR	ID
Principe	Score composite	Angle entre TF et Ac	% de couverture de la tête fémorale par l'Ac	Calcul de la pente Ac	Mesure de la congruence
Age	12 mois			6 mois	4 mois
Positionnement		Hyperextension		Décubitus sternal	Distraction
Seuil	Note > D	Angle < 105°	Couverture < 50 %	Angle > 15°	Note > 0,3 + ou - arthrose
Remarques	Extension des hanches diminue la laxité			Superpositions des structures (difficulté de mesures)	Nécessité d'un outil de distraction

TABLE 3 - Comparaison des différents examens cliniques pour la détection de la DCF

Manœuvre	Barlow	Ortolani	Bardens
Principe	Déterminer l'angle de réduction	Déterminer l'angle de subluxation	Déterminer une laxité articulaire
Espèce	Humain / Chien		Chien
Age	Naissance à quelques mois		
Positionnement	Décubitus dorsal		Décubitus latéral
Seuil	Ressaut ressenti		Elévation > 1,3 cm
Remarques	Douleur de l'animal Mesure difficile		

5.7 Différences et similarités entre humain et chien

Les dysplasies bilatérales sont plus difficiles à dépister chez l'humain, car la boiterie n'est pas franche. L'échelle de temps est bien différente entre l'humain et le chien. Le dépistage est complexe chez l'humain comme chez le chien, un examen clinique est nécessaire mais est opérateur dépendant. Le diagnostic comprend un faisceau d'indices.

5.7.1 Age de dépistage

Les échelles de temps sont très différentes entre l'humain et le chien. En moyenne il est considéré qu'un facteur de 7,5 est à appliquer entre les âges des deux espèces, cependant les stades de développement ne sont pas alignés (Meeson et al., 2019) (voir figure 6). L'âge de la marche est très différent chez l'humain (12-15 mois) et chez le chien (J+1-3) cela est associé à une ossification très rapide de l'articulation chez le chiot qui est stimulée par sa marche. Cependant le diagnostic est posé beaucoup plus tardivement chez le chien (i.e., minimum 16 semaines et officiel à 1 an) alors que chez l'humain, les premiers indices sont cliniques et sont obtenus dès les dix premiers jours de vies.

Chez le chien il n'existe actuellement pas d'examen néo-natal permettant d'évaluer précisément cette laxité, bien que des tests de dépistages précoces par échographie soient en cours au États-Unis. Pour des questions économiques de nombreux spécialistes seraient intéressés par la mise en place de ce dépistage, mais l'échographie n'est à ce jour pas prédictive d'un développement futur de DCF (Adams et al., 2000 ; Ohlerth et al., 2003).

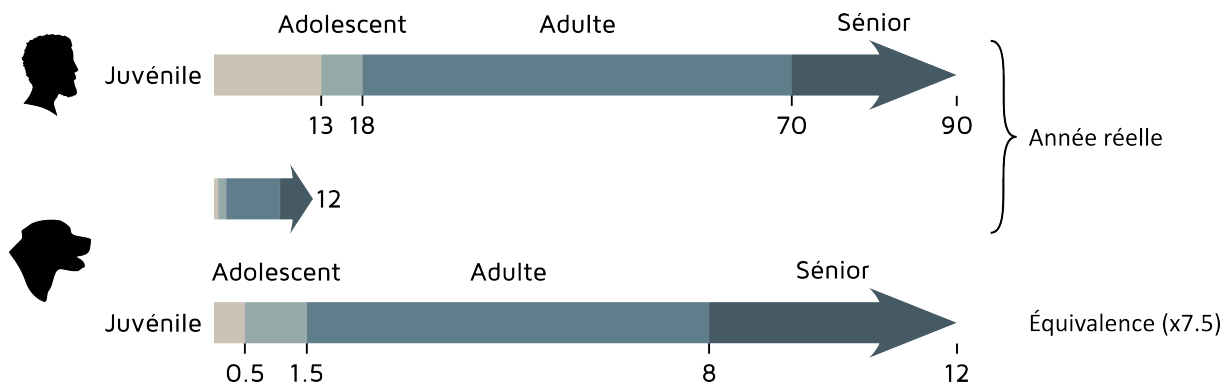


FIGURE 6 - Comparaison des stades de développement entre l'humain et le chien

Inspiré de (Meeson et al., 2019)

5.7.2 Pour l'examen clinique

Chez le chien, les manœuvres doivent être faites avant 6 mois (i.e., Ortolani, Bardens, Barlow) et sont peu prédictives (i.e., 35 % de faux positifs et 50 % de faux-négatifs), elles le sont pourtant chez l'humain. En effet chez l'humain si les signes de Bardens et d'Ortolani sont positifs alors l'individu sera forcément dysplasique.

Chez l'humain, la manœuvre de Barlow est faite de manière unilatérale à l'inverse de la manœuvre d'Ortolani. Cette dernière est différente chez l'humain et chez le chien. Elle se fait en décubitus latéral chez le chien et en décubitus dorsal chez l'humain. De plus, ces manœuvres ne se font pas au même stade d'ossification, ce qui est détecté n'est donc pas nécessairement identique.

5.7.3 Pour les imageries

L'ossification des structures osseuses de l'articulation joue un rôle important dans la technique d'imagerie utilisée aux différents âges chez les deux espèces. Lorsque le cartilage est encore majoritaire une radiographie ne permettra pas d'observer les structures anatomiques de manière fiable, l'échographie est alors employée.

Chez le chien l'ossification commence très tôt mais ne se termine que vers l'âge d'un an, âge du diagnostic définitif sauf pour les races géantes. Il est cependant possible d'obtenir des premiers indices avec une vue en contrainte dès 4 mois. Chez l'humain les radiographies ne deviennent interprétables qu'à partir d'un an et fiables à partir de 2,5 ans. C'est donc l'échographie qui est utilisée à partir de 4 semaines de vie. Enfin les radiologies sont faites sous anesthésie ou sédation profonde chez le chien et en éveil chez l'humain.

5.7.4 Laxité

Chez le chien, une laxité très importante peut être présente sans déformation osseuse, les lésions osseuses ne sont pas systématiques contrairement à l'humain où il n'y a pas de luxation sans déformation osseuse. De plus cette laxité chez l'humain n'est présente que dans les premiers jours de vie elle disparaît par la suite (i.e., 3 - 4 mois) du fait de la rigidification des structures molles (i.e., muscles et cartilages).

6 Etiologie

6.1 Facteurs environnementaux

6.1.1 Nutrition

Aucune carence alimentaire n'a pour l'instant été démontrée comme étant un facteur de risque de la DCF tant que les apports recommandés d'acides gras et d'acides aminés sont respectés (Richardson, 1992).

Un excès de calcium et de vitamine D semble favoriser l'apparition de la maladie chez les individus prédisposés. Un apport en vitamine C pendant la gestation et après celle-ci, permettait d'éliminer la prévalence de la maladie (étude non reproductible). Le chien produisant lui-même sa vitamine C, la carence n'est observée qu'en cas d'insuffisance hépatique. L'excès de vitamine peut causer une hypercalcémie et retarder le développement des os et des cartilages, cette pratique n'est donc pas recommandée.

L'osmolarité trop importante du liquide synovial a été mise en cause chez les chiens dysplasiques en raison d'un déséquilibre de concentration de sodium, potassium et chlore. Une réduction de l'apport anionique serait susceptible de réduire la subluxation à 30 semaines et améliorerait l'état des hanches à 2 ans (Kealy et al., 1993).

6.1.2 Poids

Le poids est un facteur de risque très important de la DCF. Il a été maintes fois démontré qu'un régime alimentaire trop riche favorise le développement de la maladie. Une prise de poids trop importante ou une croissance trop rapide résultent en une augmentation de la sévérité de la dysplasie (Lust et al., 1973 ; Richardson, 1992 ; Kealy et al., 1992 ; Fries & Remedios, 1995 ; Kasström, 2016). Le poids, par son augmentation, intensifierait les forces appliquées sur l'articulation et modifierait son mouvement, accroissant par ce fait l'arthrose de l'individu en surpoids (Brady et al., 2013). Cependant tous les articles ne font pas consensus. Les différences de résultats pourraient s'expliquer par des différences de méthodologie et de cohorte (Krontveit et al., 2010).

6.1.3 Exercice

L'activité physique a une double influence sur le développement de la maladie. Elle permet, lorsque modérée, de développer une musculature plus importante, stabilisant l'articulation. De plus elle agit indirectement sur le développement de la maladie par son action sur le poids de l'individu. En réduisant la masse et en musclant l'individu, elle réduit le stress mécanique sur le squelette immature du chiot. Il est donc recommandé de faire des marches courtes plusieurs fois par jours dans cette optique (Corral, 2018). Cependant une activité trop importante peut engendrer un effet contraire. La montée répétée d'escaliers est un facteur de risque et devrait donc être évitée avant l'âge de 3 mois (Krontveit et al., 2012). D'autres types d'activités intenses tel le lancer de

balle, la course ou le saut (e.g., agility, frisbee) doivent être pratiqués avec retenue, en particulier pour les tout jeunes animaux ou pour les vieux chiens atteints (Millis & Levine, 2013). Cependant, selon Greene et al. (2013), une durée plus importante d'exercice permettrait de réduire cette boiterie et cela indépendamment de l'intensité de cet exercice. D'autres études doivent donc être menées sur ce point.

6.1.4 Saison de naissance

Krontveit et al. (2012) ont montré que les chiens nés en automne étaient bien plus touchés que tous les autres chiens nés dans les autres saisons. Dans l'article de Loder & Todhunter (2017), ce sont les chiens nés en hiver et au printemps ainsi que ceux nés dans les régions plus au sud qui sont le plus touchés. Ces différences peuvent s'expliquer par les différences météorologiques entre les deux populations étudiées, la première étant faite en Norvège et la seconde aux Etats-Unis. L'effet saisonnier peut s'expliquer de différentes manières. Il a été suggéré par Loder & Todhunter (2017) que le déséquilibre entre l'apport alimentaire et le manque d'activité physique en hiver soit en partie responsable. Une autre possibilité suggérée par les auteurs serait des variations dans l'apport alimentaire et plus particulièrement une déficience en vitamine D.

6.1.5 Gestation

L'accouchement par le siège, des désordres neuromusculaires ainsi que des déficiences en fluide amniotique ont été incriminés dans la dysplasie de la hanche chez l'humain. La restriction des mouvements du fœtus in-utero augmente le risque de développer la maladie (Aronsson et al., 1994). Cela a aussi été montré expérimentalement sur les embryons de poulets (Nowlan et al., 2014). Cependant, chez le chien, aucune étude n'a pour le moment mis en évidence de tel résultat. En effet la dysplasie de la hanche chez le chiot n'est pas visible à la naissance, il peut néanmoins y avoir une influence de la vie prénatale sur la morphologie de la hanche et donc influencer le développement de la maladie.

L'évolution de la maladie peut aussi être altérée lors de la mise-bas des chiots. Ainsi des chiots nés par césarienne et élevés à la main auront une croissance très ralentie, comparativement à des chiots nés normalement et élevés par leur mère, en conséquence l'incidence de la DCF en sera réduite (Lust et al., 1973).

6.2 Facteurs hormonaux

Deux hormones ont principalement été étudiées : l'œstradiol-17 β et la relaxine. Ces deux hormones sont naturellement transmises aux chiots par la mère via le lait maternel et auraient toutes les deux un rôle dans la laxité de la hanche (Steinetz et al., 1987 ; Goldsmith et al., 1994).

6.2.1 Oestradiol

Lorsqu'administré à des chiots l'œstradiol-17 β voit son niveau de concentration augmenter dans leurs urines (Pierce & Bridges, 1967). Lorsque donné à des chiots ou à des chiennes en gestation, cette même hormone induit des changements défavorables de la tête fémorale ainsi qu'une instabilité de l'articulation (Gustafsson, 1975). Cet effet délétère a été confirmé plus récemment par (Hashem et al., 2006). L'étude de Beling et al. (1975) montre qu'une fois adulte les chiennes, atteintes ou non, métabolisent de la même manière l'œstradiol. Kasström et al. (1975) ont trouvé une concentration sanguine plus élevée d'œstradiol-17 β chez les chiots sains comparés aux chiots atteints. Les auteurs suggèrent que cette divergence de résultats pourrait s'expliquer par le fait que ce ne soit pas l'hyperœstrogénisme qui soit en cause mais la faible présence de protéines capable de se lier aux stéroïdes. D'autres études restent donc à mener.

6.2.2 Relaxine

Aussi bien l'œstradiol-17 β que la relaxine et leur combinaison induisent une perte de collagène et de glycosaminoglycanes à différents endroits. Steinetz et al. (2008) n'ont trouvé aucune différence de concentration pour ces deux hormones entre des mères atteintes et des mères contrôles. Seuls les chiots provenant de mères atteintes présentaient de l'œstradiol-17 β . Il a donc été suggéré que ces chiots atteints auraient une expression prématurée ou inappropriée des récepteurs ostrogéniques ou des récepteurs à la relaxine.

6.2.3 Sexe et stérilisation

Plusieurs études ont étudié l'effet du sexe de l'individu ainsi que de sa stérilisation, précoce ou non, sur le risque de développer la dysplasie de la hanche, mais les résultats divergent. Dans plusieurs études le sexe de l'individu ne semble pas interférer avec la maladie (Krontveit et al., 2010). Loder & Todhunter (2017), quant à eux, ont trouvé une prévalence plus importante de la maladie chez les femelles mais cela était dépendant de la race étudiée. Cela a été confirmé chez le berger allemand par (Hedhammar et al., 1979).

Chez les goldens retrievers, un effet négatif de la stérilisation précoce chez les mâles a pu être démontré (Torres de la Riva et al., 2013). De manière équivalente Spain et al. (2004) ont pu attester de cet effet délétère pour les deux sexes lorsque les individus

étaient stérilisés avant 5,5 mois. Par ailleurs, ces mêmes chiens semblent avoir été touchés plus sévèrement que leurs congénères atteints, stérilisés plus tardivement.

6.3 Facteurs génétiques

Depuis les années 1990, la dysplasie de la hanche a fait l'objet de nombreuses investigations génétiques, tant chez l'humain que chez le chien (voir tableaux 4 et 5). Ces travaux couvrent un large spectre technologique : des premières études basées sur les marqueurs microsatellites, au génotypage à haut débit via puces à SNP, jusqu'au séquençage complet du génome (WGS). Les populations humaines analysées sont très diverses, allant de cohortes européennes, asiatiques ou moyen-orientales à des études familiales ou cas-témoins. Chez le chien, un grand nombre de races ont été étudiées, soit indépendamment, soit au sein de cohortes mixtes.

Les méthodologies d'analyse ont, elles aussi, considérablement évolué au fil des décennies. Les premières études de liaison génétique ont été progressivement remplacées par des analyses d'association pangénomiques (GWAS), parfois enrichies de données transcriptomiques ou d'annotations fonctionnelles. Si ces évolutions méthodologiques ont permis une meilleure caractérisation des loci candidats, elles ont également introduit une hétérogénéité importante entre les études et entre les espèces.

Cette variabilité se reflète notamment dans les protocoles de phénotypage, souvent divergents d'une étude à l'autre, tant chez l'humain que chez le chien. L'évolution des génomes de référence (i.e., plusieurs mises à jour chez le chien, changement de version chez l'humain) et l'amélioration continue de leur annotation ajoutent un niveau de complexité à la comparaison des résultats. Sur le plan méthodologique, les études humaines s'appuient majoritairement sur des approches ciblées par gènes candidats, bien que plusieurs GWAS aient été publiées récemment. Chez le chien, malgré un nombre croissant d'études d'association, très peu de loci ont été validés de manière inter-race. Le contrôle de la structure de population constitue un défi méthodologique majeur, souvent abordé de manière hétérogène selon les études.

En résumé, bien que la littérature soit abondante, la diversité des populations, des approches méthodologiques et des critères phénotypiques freine l'émergence d'un consensus robuste. Dans ce contexte, la réinterprétation standardisée de ces résultats sur une base commune apparaît comme une étape indispensable pour dégager des éléments comparables entre espèces.

6.3.1 Comparaison génétique

À partir des données collectées dans la littérature scientifique, nous avons entrepris une comparaison systématique des loci chromosomiques associés à la dysplasie coxo-fémorale chez le chien et à la dysplasie développementale de la hanche chez l'humain. Pour permettre cette comparaison, il est nécessaire de projeter les loci identifiés dans chaque espèce sur un génome de référence commun. Compte tenu de la forte similarité entre les génomes humain et canin (Lindblad-Toh et al., 2005), notamment via des régions en synténie (voir figure 7), cette approche est biologiquement pertinente.

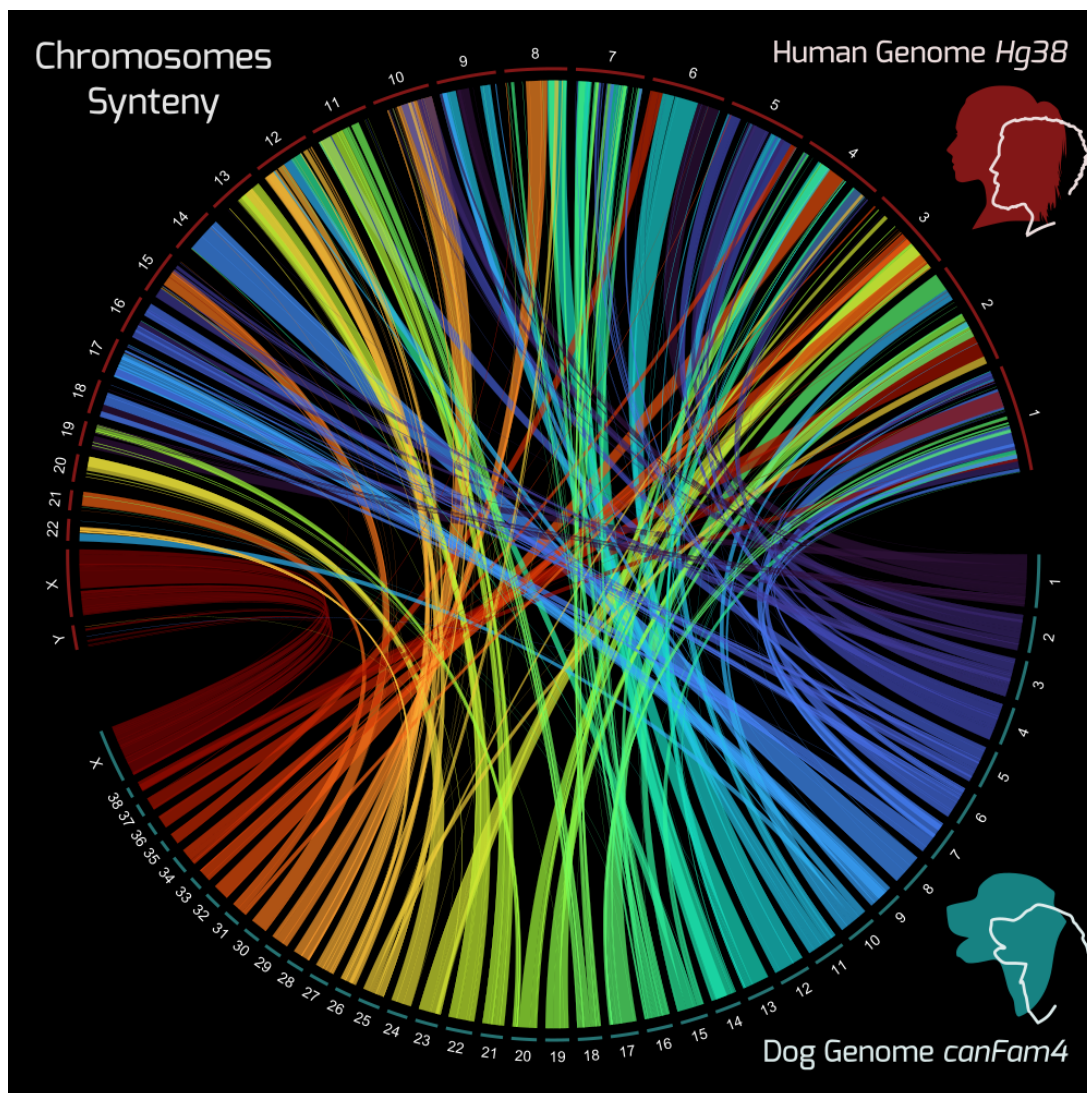


FIGURE 7 - Représentation graphique de la synténie (i.e., régions chromosomiques orthologues) entre le génome de référence humain *hg38* et le génome de référence canin *canFam4*.

Cette figure a été réalisée en utilisant un script R et le package *circize*, à partir du fichier *.chain liftOver* mis à disposition par le site Genome Browser de l'Université de Californie UCSC (<https://hgdownload.soe.ucsc.edu/goldenPath/canFam4/liftOver/>). Les barres colorées représentent les chromosomes canins (en bas) et humains (en haut). Les lignes reliant les deux ensembles indiquent les régions en synténie, c'est-à-dire les segments de chromosomes partageant une origine évolutive commune. La taille des chromosomes est proportionnelle à leur longueur en mégabases (Mb).

Nous avons ainsi développé une série de scripts en langage *R* reposant sur l'outil *liftOver* du package *Bioconductor* éponyme (Bioconductor Package, 2025), permettant de convertir des positions génomiques entre différentes versions de génomes de référence et entre espèces. Cette normalisation est indispensable étant donné l'hétérogénéité des technologies utilisées dans les études d'origine (e.g., microsatellites, puces SNP, WGS, exome) et les nombreuses mises à jour des génomes de référence (e.g., *hg18*, *hg19*, *hg38* chez l'humain ; *canFam2*, *canFam3*, *canFam4* chez le chien).

Les étapes méthodologiques ont été les suivantes. Dans un premier temps, pour les études canines anciennes basées sur des microsatellites, les séquences des amorces ont été extraites de la littérature puis alignées sur *canFam4* à l'aide de l'outil *in silico* PCR d'*UCSC*. Pour les études humaines et canines fournissant des séquences génomiques, celles-ci ont été positionnées respectivement sur *hg38* ou *canFam4* via *BLAST*. Pour les études humaines, le package *biomaRt* (Durinck et al., 2009) fut ensuite utilisée pour positionner les gènes référencés et la base de données *miRBase* pour les microARN (Griffiths-Jones et al., 2006).

Les positions sur le génome de référence *hg18* et de *hg19* furent ensuite converties vers *hg38*. Pour les études canines, la conversion fut un peu plus complexe. Les données sous *canFam2* furent converties en *canFam3*, tout comme les positions en centimorgan par le biais de la carte de recombinaison génétique de (Auton et al., 2013). Puis l'ensemble des positions sous *canFam3* et *hg38* furent converties vers *canFam4*. A l'issue de ce processus, 1 496 loci canins et 2 512 loci humains ont été identifiés. Parmi les loci humains, 335 n'ont pas pu être transférés vers *canFam4*, principalement en raison de l'absence de régions orthologues identifiables.

Les loci ainsi harmonisés ont été représentés sur le génome *canFam4* à l'aide du package *karyoploteR*, modifiée pour permettre une visualisation adaptée à la densité des signaux. Le génome a de même été découpé en fenêtres de 1 Mb et le nombre d'études identifiant un locus dans chaque région a été comptabilisé. Afin d'identifier et de prioriser les régions génomiques potentiellement partagées entre l'humain et le chien, nous avons développé une métrique simple mais robuste : le produit du nombre d'études chez l'humain par celui chez le chien pour chaque région donnée. Cette approche permet d'exclure les régions spécifiques à une seule espèce (le produit étant nul si l'une des deux espèces n'est pas représentée) tout en valorisant les régions investiguées chez les deux espèces, même si individuellement peu d'études les rapportent.

Les résultats sont présentés sous forme de tableau pour les régions les plus fortement représentées chez l'humain et le chien et en commun (voir tableau 4) ainsi que sous forme de caryotype enrichi et de heatmap en figure 8. Cette analyse met en évidence une concentration marquée des signaux dans certaines régions chromosomiques. Ainsi, la région la plus fréquemment rapportée dans la littérature humaine

est localisée sur le chromosome 24 canin (chr24 :24,778,688-25,769,834, *canFam4*), correspondant à la région chr20 :35,284,038-36,407,469 sur le génome humain (*hg38*) et contenant notamment les gènes *UQCC1* et *GDF5*, bien connus pour leur implication dans la croissance. Du côté canin, la région la plus représentée est située sur le chromosome 3 (chr3 :39,775,940-40,770,337, *canFam4*), correspondant à chr15 :100,620,429-101,661,887 chez l'humain.

Cette analyse systématique de la littérature offre non seulement une vue d'ensemble de la distribution des loci associés à la dysplasie de la hanche, mais permet également d'identifier des régions génomiques d'intérêt commun entre les deux espèces. Certaines d'entre elles, bien que modestement représentées dans chaque espèce prise isolément, apparaissent comme des zones d'intérêt lorsqu'on considère la convergence des résultats. Ces régions communes pourraient refléter des mécanismes pathogéniques partagés, ouvrant la voie à une meilleure compréhension de la physiopathologie de la dysplasie de la hanche à travers une approche comparative interespèces.

TABLE 4 - Régions chromosomiques d'1 Mb les plus étudiées chez l'humain, chez le chien et en commun

Le nombre d'études et de loci pour chaque région chromosomique est donné sous la forme littérature humaine / littérature canine. Seule les 5 régions les plus étudiées par espèce sont présentées ici.

Espèce	Région	Nb études	Nb loci	Gènes présents dans la région
Humain	chr24:24778688-25769834	14/3	113/5	FER1L4, PHF20, CNBD2, EPB41L1, DLGAP4, EIF6, FAM83C, UQCC1, GDF5, CEP250, C24H20orf173, ERGIC3, SPAG4, CPNE1, RBM12, NFS1, ROMO1, RBM39, SCAND1, AAR2
	chr27:39712756-40705574	10/0	19/0	VDR, ASB8, TMEM106C, COL2A1, SENP1, PFKM, CCDC184, H1-7, ZNF641, OR8BJ4, OR8S3, OR8S2, C27H12orf54, OR8S1, OR8S12, OR8S13, OR8S9, OR8S15, OR8S7, OR8S10
	chr9:25652737-26639379	9/0	22/0	SPOP, KAT7, DLX4, SAMD14, ACSF2, CACNA1G, SLC35B1, FAM117A, TAC4, DLX3, ITGA3, PDK2, PPP1R9B, SGCA, COL1A1, TMEM92, XYLT2, MRPL27, EME1, LRRRC59, CHAD, RSAD1, EPN3, SPATA20, ABCC3, ANKRD40, LUC7L3, ANKRD40CL
	chr1:98645891-99642313	8/0	17/0	CENPP, ASPN, PTPDC1, PHF2, WNK2, FGD3, IPPK, BARX1, FAM120A, NINJ1, CARD19, SUSD3, MIR8893, BICD2, ECM2
	chr23:8855833-9839814	8/1	15/3	TTC21A, CX3CR1, RPSA, MYRIP, SCN11A, WDR48, GORASP1, CSRNP1, XIRP1, CCR8, SLC25A38
Chien	chr3:39775940-40770337	1/8	2/17	TARS3, PCSK6, CHSY1, LRRK1, ALDH1A3, ASB7, TM2D3, SNRPA1, SELENOS
	chr4:22881213-23876047	1/5	1/24	CDH23, PSAP, ASCC1, MICU1, C4H10orf105, VSIR, CHST3, SPOCK2, ANAPC16, DDIT4, DNAJB12
	chr10:48753657-49748629	2/5	2/13	PRKCE, SRBD1, EPAS1
	chr20:43722196-44715882	2/5	4/12	ATP13A1, EXOSC7, KIF15, GATAD2A, SUGP1, TMEM161A, SUGP2, CLEC3B, ZDHHC3, TMEM42, KIAA1143, ZNF501, ZNF502, MIR8801, GMIP, LPAR2, PBX4, CILP2, YJEFN3, NDUFA13, TSSK6, MAU2, TM6SF2, HAPLN4, NCAN, NR2C2AP, RFXANK, BORCS8, MEF2B, SLC25A42, ARMC6, HOMER3, DDX49, COPE, CERS1, GDF1, UPF1
	chr33:3895663-4869578	1/5	1/12	EPHA6
Commun	chr24:24778688-25769834	14/3	113/5	FER1L4, PHF20, CNBD2, EPB41L1, DLGAP4, EIF6, FAM83C, UQCC1, GDF5, CEP250, C24H20orf173, ERGIC3, SPAG4, CPNE1, RBM12, NFS1, ROMO1, RBM39, SCAND1, AAR2
	chr7:22742399-23731198	6/3	15/3	PAPPA2, COP1, TNR, MIR1843
	chr1:112595815-113592237	8/2	9/2	CEACAM1, LIPE, ERF, POU2F2, BCKDHA, AXL, CXCL17, CNFN, MEGF8, TMEM145, PRR19, PAFAH1B3, CIC, GSK3A, ZNF526, DEDD2, ZNF574, GRIK5, ATP1A3, RABAC1, ERFL, ARHGEF1, CD79A, RPS19, DMRTC2, LYPD4, ERICH4, DMAC2, B3GNT8, EXOSC5, TMEM91, B9D2, TGFB1, CCDC97, HNRNPUL1, CYP2S1
	chr9:34532530-35519172	5/3	8/15	BCAS3, RPS6KB1, INTS2, BRIP1, MED13, TBX4, TBX2
	chr4:12932860-13927694	3/4	3/7	ANK3, CDK1, RHOTB1

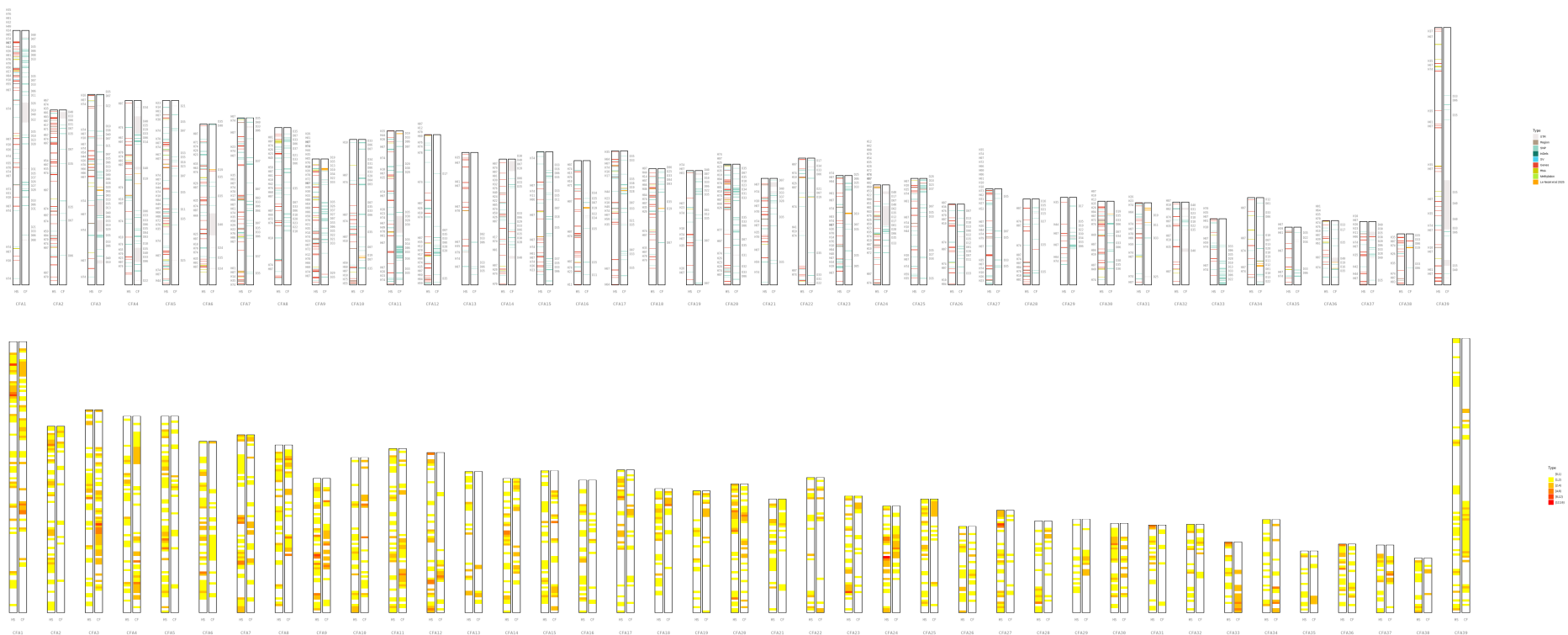


FIGURE 8 - Représentation sous forme de caryotype des loci identifiés dans la littérature humaine et dans la littérature canine.

La littérature humaine est représentée sur le chromosome de gauche (préfix « H ») et la littérature canine sur chromosome de droite (préfix « D »). Toutes les positions sont représentées sur le génome de référence canFam4. Le graphique du haut représente l'ensemble des loci par études (liste des codes de chaque étude en figure suivante) et par types (couleurs de chaque loci). Le graphique du bas représente la heatmap des différentes études pour chaque fenêtre de 1 Mb.

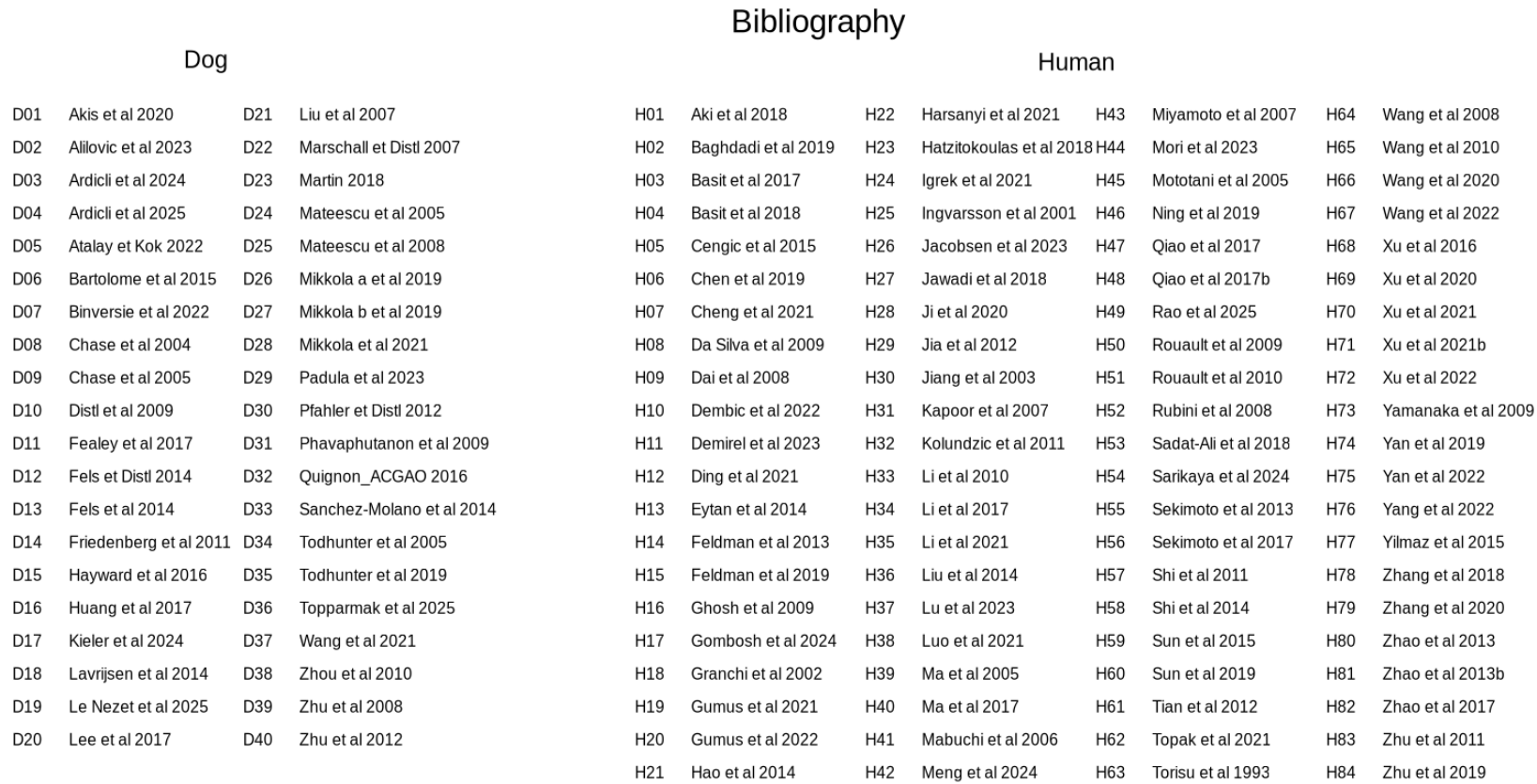


FIGURE 8 - Représentation sous forme de caryotype des loci identifiés dans la littérature humaine et dans la littérature canine. (suite)
La littérature humaine est représentée sur le chromosome de gauche (préfix « H ») et la littérature canine sur chromosome de droite (préfix « D »). Toutes les positions sont représentées sur le génome de référence canFam4. Le graphique du haut représente l'ensemble des loci par études (liste des codes de chaque étude en figure 8) et par types (couleurs de chaque loci). Le graphique du bas représente la heatmap des différentes études pour chaque fenêtre de 1 Mb.

7 Conclusion

Cette revue vise à nourrir une réflexion comparative entre médecine vétérinaire et médecine humaine, en s'appuyant sur l'étude conjointe de la dysplasie de la hanche chez le chien et chez l'humain. En mettant en lumière les convergences anatomiques, diagnostique et étiologiques entre les deux espèces, elle ouvre la voie à une compréhension plus fine des mécanismes physiopathologiques et génétiques sous-jacents à ces affections articulaires.

Le passage évolutif de la quadrupédie à la bipédie constitue un tournant majeur dans l'histoire de l'articulation de la hanche humaine. Il a entraîné une réorientation des contraintes mécaniques, une modification de l'anatomie pelvienne, ainsi qu'un remodelage de l'architecture musculaire. Malgré ces divergences liées à la locomotion, le chien reste, par la similarité structurelle et fonctionnelle de son articulation coxo-fémorale, un modèle de choix pour étudier les dysplasies de la hanche. Cette proximité, alliée à la richesse des phénotypes observés chez les races canines et à la disponibilité de données génomiques, confère à l'espèce canine une place centrale dans une approche de type *One Health*, au croisement de la médecine humaine, vétérinaire et de la biologie évolutive.

L'analyse des loci issus de la littérature, harmonisée via une approche comparative rigoureuse, a permis de cartographier les régions génomiques d'intérêt sur le génome canin (i.e., *canFam4*) et de révéler plusieurs régions communes aux deux espèces. Ces convergences, bien qu'émergentes de données fragmentées, renforcent la pertinence d'une approche comparative interespèces dans l'étude des maladies articulaires.

Ainsi, la comparaison interespèces ne se limite pas à une simple analogie : elle devient un véritable levier pour décrypter les bases biologiques communes des maladies, améliorer les diagnostics, affiner les stratégies thérapeutiques et *in fine*, faire progresser la santé de toutes les espèces concernées.

8 Supplementary material

TABLE S1 - Muscles de la hanche, leur groupe musculaire et leurs mouvements principaux

Les mouvements sont indiqués par un X dans la colonne correspondante. Abréviations : Rot. : rotation, Int. : interne, Ext. : externe, M. : musculus. + indique un mouvement faible et - indique un mouvement incertain. (A) fibres antérieures, (P) fibres postérieures, (F) en position fléchie, (E) en position étendue. Source : Sajko & Stuber (2009), Kenhub Hip Muscles (2025), and Physiopedia Hip Muscles (2025).

Localisation	Groupe musculaire	Muscle (nom latin)	Mouvements					
			Flexion	Extension	Rot. Int.	Rot. Ext.	Abduction	Adduction
Ceinture pel- viennne ventrale	Ilio-psoas	Iliaque (M. iliacus)	X			X		
		Psoas (M. psoas major)	X			X	+/-	+/-
Muscles dorsaux et latéraux de la ceinture pelvienne	Glutéaux (muscles fessiers superficiels)	Petit (M. gluteus minimus)			X		X(A)	
		Moyen (M. gluteus medius)	X(A)	X(P)	X(A)	X(P)	X	
		Grand (M. gluteus maximus)		X		X	X(A)	X(P)
		Tenseur du facia lata (M. tensor fasciae latae)	X		X		X	
	Pelvi-trochantériens (muscles fessiers profonds)	Piriforme (M. piriformis)				X	X(F)	
		Jumeau supérieur (M. gemellus superior)				X	X	
		Jumeau inférieur (M. gemellus inferior)				X	X	
		Obturateur interne (M. obturator internus)				X	X	
		Obturateur externe (M. obturator externus)				X		X
		Carré fémoral (M. quadratus femoris)				X		X
	Quadriceps fémoral	Droit fémoral (M. rectus femoris)	X					
		Sartorius (M. sartorius)	X			X(F)	X(F)	
	Adducteurs	Pectiné (M. pectineus)	X		+	+		X
		Gracile (M. gracilis)	+					X
		Court adducteur (M. adductor brevis)	+			-		X
		Long adducteur (M. adductor longus)	+(E)	+(F)	-			X
		Grand adducteur (M. adductor magnus)	X(A)	X(P)	+(P)	+(A)		X
Loge fémorale postérieure	Ischio-jambiers	Biceps fémoral (M. biceps femoris)		X		X		
		Semi-tendineux (M. semitendinosus)		X	+			
		Semi-membraneux (M. semimembranosus)		X	+			

TABLE S2 - Critères d'interprétation des résultats en fonction de l'ID et des modifications ostéo-articulaires (Guevar & Snaps, 2008)

Abbréviations : dysplasie coxo-fémorale (DCF).

Indice de distraction	Modifications ostéo-articulaires	Conclusions
$\leq 0,3$	Aucune	Animal indemne
$\leq 0,3$	Légère à modérée	Dysplasie de la hanche confirmée (rare)
$> 0,3$	Aucune	Au-delà de 0,3, risque de développement de DCF, faible si proche de 0,3 élevé si proche de 0,7
$> 0,7$	Aucune	Au-delà de 0,7, risque élevé de développement de DCF
$> 0,3$	Légère à modérée	DCF confirmée
Indice non calculé	Sévère	DCF confirmée. Les sévères modifications ostéo-articulaires rendent difficile l'obtention d'un indice de distraction

TABLE S3 - Score utilisé par la Fédération Cynologique Internationale (FCI) pour caractériser la dysplasie coxo-fémorale chez le chien

Abbréviations : angle de Norberg-Olsson (NO), acétabulum (Ac), tête fémorale (TF). Source : Flückiger (2007b) and Commission & Committee (2022).

Score FCI	Angle NO	Positionnement de la tête fémoral	Interligne articulaire	Os sous-chondral du rebord acétabulaire cranial	Rebord acétabulaire cranial	Arthrose	Conclusion
A	NO $\geq$ 105°	Parfaitement centrée dans l'Ac Centre médial au rebord Ac dorsal	Etroit et régulier	En ligne étroite d'épaisseur régulière	En arc de cercle bien délimité Parallèle au contour de la TF sauf au niveau de la fovea capitis	Absente	Hanche indemne
B	NO $\geq$ 100°	Centrée dans Ac Centre médial ou sur rebord Ac dorsal	Etroit	En ligne étroite d'épaisseur régulière	Horizontal dans sa partie latérale et forme une ligne droite dans un plan transversal Pas forcément parallèle au contour de la TF	Absente	Hanche sensiblement normale
C	NO $\approx$ 100°	Pas correctement centrée dans la Ac Centre situé sur rebord Ac dorsal ou latéralement Subluxation latérale ou caudale possible		Léger épaissement latéral ou léger amincissement médial possible	Légèrement aplati, (i.e., s'écarte cranio-latéralement de la TF) Divergent (non parallèles) au contour de la TF	Signes d'arthrose peuvent être présents	Dysplasie légère
D	NO > 90°	Pas correctement centrée dans Ac Centre situé latéralement au rebord acétabulaire dorsal Subluxation latérale ou caudale possible		Léger épaissement latéral ou léger amincissement médial	Aplati (i.e., s'écarte cranio-latéralement de la TF) Nettement divergent au contour de la TF	Signes d'arthrose peuvent être présents	Dysplasie moyenne
E	NO < 90°	Centre situé latéralement au rebord acétabulaire dorsal Luxé ou subluxé Remodelage ou déformation possible de la TF et de l'Ac		Epaissi latéralement, jusqu'à la hauteur du rebord cranio-latéral Cet épaissement peut être absent lors de luxation de l'articulation	Aplati et s'écarte cranio-latéralement de la TF, la partie latérale peut disparaître Nettement divergent au contour de la TF	Signes d'arthrose peuvent être présents	Dysplasie sévère

TABLE S4 - Score utilisé par la structure américaine *Orthopedic Foundation for Animals* (OFA)*Abbreviations : acétabulum (Ac), tête fémorale (TF).*

Note	Observations
Excellent	La conformation est de qualité supérieure. La TF s'intègre étroitement dans un Ac bien formé avec un interligne articulaire minimal et une très bonne couverture dorsale de la TF
Bon	Un peu moins bien que « excellent », mais l'articulation de la hanche est congruente et bien formée. La TF s'intègre bien dans l'Ac et une bonne couverture dorsale est présente.
Correcte	Irrégularités mineures présentes. L'interligne articulaire est plus large que sur une bonne hanche. La TF glisse légèrement hors de l'Ac. L'Ac peut également apparaître légèrement moins profond.
Limite	Absence de consensus pour classer la hanche comme normale ou non. Habituellement, l'incongruence est légèrement supérieure à celle d'une hanche classée « correcte », mais absence de signes d'arthrose.
Légère	Subluxation significative consécutive à la sortie partielle de la TF hors de l'Ac, provoquant un élargissement de l'interligne articulaire. L'Ac est généralement peu profond et ne couvre que partiellement la TF.
Moyenne	La TF est à peine assise dans un Ac peu profonde. Présence de signes d'arthrose secondaires à la hauteur du col fémoral et de la tête (i.e., remodelage). Présence de changement du rebord acétabulaire (e.g., ostéophytes ou éperons osseux) et de divers degrés de densification trabéculaires de la plaque osseuse sous-chondrale (i.e., sclérose).
Sévère	Des preuves marquées de dysplasie de la hanche sont présents. La TF est subluxée ou luxée, l'Ac étant peu profond. Des modifications arthrosiques significatives sont observées le long du col fémoral, de la TF et de la cavité acétabulaire.

TABLE S5 - Score utilisé par la *British Veterinary Association* (BVA) pour caractériser la dysplasie coxo-fémorale chez le chien

Abbréviations : acétabulum / acétabulaire (Ac), tête fémorale (TF), angle de Norberg-Olsson (NO), Ostéophytose (OP). Pour chaque item présenté en colonne, un score de 0 à 6 est attribué en fonction de la description. Le score final est la somme des scores obtenus pour les neufs items et pour les deux hanches (maximum 108). Source : Flückiger (2007b).

Score	Angle NO	Subluxation	Rebord Ac Crânial (RAC)	Rebord Ac Dorsal (RAD)	Rebord Ac Efficace Crânial (RAEC)	Fosse Ac (FA)	Rebord Ac Caudal (RACa)	Ostéophytose fémorale de la tête et du col	Remodelage de la TF
0	105° et plus	TF bien centrée dans l'Ac	Courbe régulière, parallèle à la TF tout le long	RAD légèrement courbe	Jonction tranchante et propre de RAD et RAC	fine ligne osseuse courbé médialement et caudalement à partir de l'extrémité caudale du RAC	Ligne propre	Profil arrondi lisse	Absent
1	100° à 104°	Centre de TF médial au RAD. Espace articulaire latéral ou médial légèrement augmenté.	Interlignes articulaires légèrement divergentes	Perte de la courbe S uniquement en présence d'autres changements dysplasiques	Jonction indistincte de RAD et RAC	Légère augmentation de la densité osseuse médiale à FA. Ligne « fine » brumeuse ou perdue.	Petite OP au RACa latéral	Légère OP sous forme « d'anneaux » ou ligne verticale dense adjacente à la fosse trochantérique (i.e., ligne de Morgan)	La TF ne forme pas un disque en raison de l'OP ou de la perte osseuse.
2	95° à 99°	Centre TF superposé au RAD. Augmentation évidente de l'interligne articulaire médial.	RAC plat sur la majeure partie de sa longueur	Très petite exostose sur RAD crânien	Très petite exostose ou très petite facette	Ligne fine perdue dans AF et ventral AE brumeuse en raison d'un nouvel os. Encoche du RACa clair	Petite OP au RACa latéral et médial	Légère OP visible sur la ligne d'horizon ou la densité sur la TF médiale	Une perte osseuse ou un anneau de tête / cou fémoral d'OP

TABLE S5 - Score utilisé par la British Veterinary Association (BVA) pour caractériser la dysplasie coxo-fémorale chez le chien

Abbréviations : acétabulum / acétabulaire (Ac), tête fémorale (TF), angle de Norberg-Olsson (NO), Ostéophytose (OP). Pour chaque item présenté en colonne, un score de 0 à 6 est attribué en fonction de la description. Le score final est la somme des scores obtenus pour les neuf items et pour les deux hanches (maximum 108). Source : Flückiger (2007b).

Score	Angle NO	Subluxation	Rebord Ac Crânial (RAC)	Rebord Ac Dorsal (RAD)	Rebord Ac Efficace Crânial (RAEC)	Fosse Ac (FA)	Rebord Ac Caudal (RACa)	Ostéophytose fémorale de la tête et du col	Remodelage de la TF
3	90° à 94°	Centre TF juste latéral au RAD. 1/2 TF dans l'Ac	Légère bilabiation	Exostose évidente sur RAD, en particulier crânial ou mineure « perte de bord »	Facette ou petite exostose ou une légère bilabiation	Remodelage incomplet de l'Ac avec une face médiale latérale à FA. Ventral AE perdu. FA brumeux. Encoche irrégulière	Importante OP et encoches étroites au RACa	OP distincte dans la « formation des anneaux »	Perte osseuse évidente et l'OP distincte donnant une légère apparence conique
4	89° à 85°	Centre de TF clairement latéral au RAD. 1/4 TF dans l'Ac	Bilabiation modérée	Exostose bien latérale à RAD ou « perte de bord » modérée	Facette évidente ou exostose évidente ou bilabiation modérée	Remodelage marqué. Face médiale de l'Ac clairement latérale à FA. Ventral AE perdu. Encoche partiellement fermée.	OP marquée et « accrochage » de l'extrémité latérale de RACa	Collier complet d'OP évident	Remodelage brut. La perte d'os et l'exostose évidente donne l'apparence de « champignons »
5	84° à 80°	Centre de TF bien latéral au RAD. La TF touche juste le RAD	Bilabiation prononcée	Exostose marquée tout au long du RAD ou « perte de bord » prononcée	Exostose brute ou facette ou bilabiation brute	Remodelage brut. Nouvel os dense dans tout l'Ac. Encoche perdue et FA obscurci	Distorsion brute due à la masse de nouveaux os dans l'Ac. Encoche complètement perdue	OP massive donnant l'apparence de « champignons »	Remodelage très brut avec une perte d'os marquée et beaucoup de nouveaux os

TABLE S5 - Score utilisé par la British Veterinary Association (BVA) pour caractériser la dysplasie coxo-fémorale chez le chien

Abbréviations : acétabulum / acétabulaire (Ac), tête fémorale (TF), angle de Norberg-Olsson (NO), Ostéophytose (OP). Pour chaque item présenté en colonne, un score de 0 à 6 est attribué en fonction de la description. Le score final est la somme des scores obtenus pour les neufs items et pour les deux hanches (maximum 108). Source : Flückiger (2007b).

Score	Angle NO	Subluxation	Rebord Ac Crânial (RAC)	Rebord Ac Dorsal (RAD)	Rebord Ac Efficace Crânial (RAEC)	Fosse Ac (FA)	Rebord Ac Caudal (RACa)	Ostéophytose fémorale de la tête et du col	Remodelage de la TF
6	79° et moins	Luxation pathologique complète	RAC entièrement en pente	Exostose massive du RAD crânial à caudal	Remodelage complet de la RAEC. Exostose massive ou facette brute	Remodelage complet et nouvelle surface articulaire, bien latérale à FA. Encoche perdue	Absent	OP massive et remplissage de la fosse trochantérique et en dessous de la TF	TF mal formée en raison du mauvais développement du centre fémoral

TABLE S6 - Score utilisé en Suisse pour caractériser la dysplasie coxo-fémorale chez le chien

Abbreviations : Interligne Articulaire (IA), angle de Norberg-Olsson (NO), acétabulum / acétabulaire (Ac), tête fémorale (TF), col fémoral (CF), ostéophytose (OP). Pour chaque item présenté en colonne, un score de 0 à 5 est attribué en fonction de la description. Le score final est la somme des scores obtenus pour les six items pour chaque hanche et la valeur la plus haute des deux est gardée (maximum 30). Source : Flückiger (2007a).

Score	Angle NO et IA	Relation Centre de la TF (CTF) et Rebord Ac Dorsal (RAD), caractérisation IA	Rebord Ac Crânial (RAC)	Os sous-chondral du rebord Ac crânial	TF et CF	Ligne de Morgan
0	NO $\geq$ 105° IA congruent	CTF médial au RAD (> 2 mm) IA étroit	Parallèle à la tête fémorale tout le long	Fin et régulier	TF ronde et lisse CF bien démarqué	Non visible
1	NO $\geq$ 105° mais IA légèrement élargi ou NO < 105° avec IA étroit	CTF médial au RAD (1-2 mm) IA minimalement divergent	RAC horizontal sur 1/4 de sa partie latérale	Régulier	TF ronde CF peu démarqué (cylindrique)	Epaulement arrondi en vue avec les membres inférieurs en abduction Bosse lisse non notée
2	100° $\leq$ NO < 105°	CTF superposé au RAD IA légèrement divergent	RAC légèrement aplati ou légère OP	Légèrement épaissi latéralement et légèrement réduit médialement	TF légèrement aplati CF avec légère OP	Fine saillie linéaire ($\leq$ 1 mm de large)
3	90° $\leq$ NO < 100°	CTF latéral au RAD (1-5 mm) IA modérément divergent	RAC modérément aplati, légère OP et bilabiation	Modérément épaissi latéralement et modérément réduit médialement	TF modérément aplati CF avec légère OP	Saillie bien définie ($\leq$ 3 mm de large)
4	80° $\leq$ NO < 90°	CTF latéral au RAD (6-10 mm) IA nettement divergent	RAC nettement aplati, OP modérée	Nettement épaissi latéralement et peut être absent médialement	TF modérément aplati CF avec OP modérée	Saillie irrégulière large (> 3 mm de large)
5	NO < 80°	CTF latéral au RAD (> 10 mm) ou luxation	RAD absent et Ac nettement déformé	Se fond avec le rebord latéral du pelvis ou est absent	TF sévèrement aplati CF avec OP importante	Saillie incorporé ou superposés à l'OP générale

TABLE S7 - Classification de la DDH selon la méthode de Graf

Abbréviations : acétabulum (Ac), tête fémorale (TF). Source : Graf et al. (2006).

Type	Sous- type	Centrage	Angle alpha	Angle béta	Toit cartilagineux	Périchondre	Echos	Autres
I	a	Centrée	> 60°	< 55°				
	b		> 55°					
II	a	Centrée	50° - 59° (< 3 mois)					
	b		50° - 59° (> 3 mois)					
	c		43° - 49°	< 77°				
D		Presque décentrée	43° - 49°	> 77°				
III	a	Décentrée	NA		Déplacement partiellement céphalée et caudal	Déplacement Vertical	Absent	
	b						Présent	
V		Décentrée	NA			Déplacement horizontal	Interposition du cartilage hyalin du bord latéral de l'Ac ainsi que le labrum fibrocartilagineux entre la TF et l'Ac	

TABLE S8 - Classification de l'arthrose selon Tönnis

Abbreviations : acétabulum (Ac), tête fémorale (TF). Source : Busse et al. (1972).

Score	Caractéristiques radiographiques
0	Absence d'arthrose
1	Réduction faible de l'interligne articulaire <i>Lipping</i> faible au bord de l'articulation Sclérose faible de la tête fémorale ou de l'Ac
2	Petits kystes sur la TF ou l'Ac Réduction de l'interligne articulaire Perte de la sphéricité de la TF
3	Kystes importants Réduction importante / disparition de l'interligne articulaire Déformation importante de la TF Nécrose avasculaire

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Sample : FH (Femoral Head) Phenotype : CDH (Congenital Dislocation of the Hip), DDH (Developmental Dysplasia of the Hip), OA (Osteoarthritis), CEA (Center Edge Angle), SA (Sharp Angle), ARO (Acetabular Roof Obliquity), MJS (Minimal Joint Space), AI (Acetabular Index), VCE (Vertical Center Edge angle), HTE (Sharp Angle), FN (Femoral Neck), HD (Hip Dysplasia), FNF (Femoral Neck Fracture), AD (Acetabular Dysplasia), EI (Extrusion Index), TA (Tönnis Angle), SL (Shenton's Line). Conclusion : LD (Linkage Disequilibrium), GWAS (Genome Wide Association Study), SNP (Single Nucleotide Polymorphism).

Article	Type	Population	Sample	Phenotype	Conclusion
Cilliers & Beighton (1990)	Model	Afrikaner		Beukes syndrom	Autosomal dominant inheritance
Torisu et al. (1993)	Genes	Japan		DDH (clinical and radiographic evidence)	HLA DR4 and DQw4 positively associated and HLA DR9 and DQW1 negatively associated
Beighton et al. (1994)	Genes	Afrikaner	Blood	Beukes	No segregation with any polymorphism
Roby et al. (1999)	STR	Afrikaner		Beukes	1 haplotype of 11 cM on chr 4
Sollazzo et al. (2000)		Italy		Non-syndromic dysplasia of the hip and Graph diagnosis	The analysis favoured a two-locus model
Ingvarsson et al. (2001)	Region	Iceland	Blood	OA from absent to total hip replacement score	3 loci on chr 16, 19 and 20
Granchi et al. (2002)	Region	Italy	Blood	OA idiopathic vs secondary to DDH	2 Alleles associated to DDH and one haplotype
Jiang et al. (2003)	STR	China			12 alleles present and one in desequilibrium with CDiH
Cvjeticanin & Marinkovic (2005)		Serbia			Significant increase of genetic homozygosity in the CDiH patient, no difference in ABO and Rh blood, type 6 HRC more present in CDH patient and it increase with severity of disease
Jiang et al. (2005)	STR	China			No association with SNP from PCOL2, no polymorphism of Sp1 of COL2A1
Ma et al. (2005)	STR	China		DDH (radiographic evidence)	3 markers with one allele in desequilibrium with CDiH
Mototani et al. (2005)	SNP	Japan		Hip OA with 40% AD	2219 SNP < 0,01, but none genome wide significant, CALM1 as susceptible gene, 1 haplotype associated and higher levels of CALM1 expression in hip and knee OA cartilage
Loughlin et al. (2006)	SNP	UK		KL > 2	No association between CALM1 and OA in UK population
Mabuchi et al. (2006)	STR	Japan	Blood	Hip OA and AD (Nakamura CEA < 20, SA > 45, ARO > 15, MJS < 2,5mm)	1 Loci chr 13
Kapoor et al. (2007)	Genes	UK	Blood	CEA, SA	Wild allele Xba I and mutant Fok I > in DDH but not significant

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Miyamoto et al. (2007)	SNP	Japan, China	Blood	Hip OA with some AD	3 SNP validated and 2 Group of SNP in LD associated to Hip OA
Dai et al. (2008)	SNP	China	Blood or Buccal	CDH unilateral and bilateral + severity (0-3)	Allele T > in severe cases
Rubini et al. (2008)	STR	Italy	Blood or Buccal	Graph and Wilson	No association
Wang et al. (2008)	RNA and Genes	China	Hip capsule	DDH	mRNA of COL1a1, COL3a1 and protein of COL1a1 decreased in DDH patient
Silva et al. (2009)	Methylation	UK	Cartilage		Loss of DNA methylation at CpG site of MMP13, MMP9 and MMP3
Egli et al. (2009)	SNP	UK	Cartilage, synovium, ligamentum teres femoris	Primary OA with THR	Allelic imbalance in GDF5 is joint wide
Evangelou et al. (2009)	SNP	UK, Japan, Netherland, Spain, USA, Iceland, Belgium		Hip OA Kellgren/Lawrence (0-4)	Japanese population add variability, SNP associated but depend of population studied
Ghosh et al. (2009)	SNP	UK	Blood	DDH (clinical and radiographic evidence)	No mutation find in patient and control
Rouault et al. (2009)	SNP and STR	France	Blood	VCE < 20, SA > 12	No association and no distortion in haplotype and allelic transmission
Yamanaka et al. (2009)	Genes	Japan	Blood	Osteotomy secondary to AD, CEA < 20 and OA Tonnis grading (0-3)	No difference
Yamanaka et al. (2009)	Genes	Japan	Blood	3 groups of MJS recovery after surgery	ER Pvu II pp associated with narrower joint space width
Feldman et al. (2009)	Region	USA	Blood or Buccal	TA > 10, CEA < 20, E, Acetabular coverage (ant, post), FN Angle	Two regions with low LOD score
Li et al. (2010)	STR	China		DDH	1 Loci of 11,7 cM on chr 17
Rouault et al. (2010)	SNP	France	Blood	VCE < 20, SA > 12	2 SNP in strong LD associated

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Wang et al. (2010)	SNP	China	Blood or Buccal	DDH unilateral and bilateral + severity (0-3)	rs3744448 associated with DDH in males
Kolundžić et al. (2011)	SNP	Croatia	Blood	OA secondary to DDH	2 SNP and their combination associated to DDH-related OA
Shi et al. (2011)	STR	China	Blood or Buccal	DDH unilateral and bilateral + severity (0-3)	8 alleles exist and 1 is more frequent in DDH and another one less frequent
Zhu et al. (2011)	SNP	China	Blood or Buccal	DDH unilateral and bilateral + severity (0-3)	3 SNP in LD and no association to DDH
Feldman et al. (2012)	Pedigree	USA		SL, CEA < 20, TA > 10, EI > 10 mm, FN angle	X-linked inheritance ruled out, mode of inheritance appears to be autosomal dominant with variable expression and penetrance
Jia et al. (2012)	SNP	China	Blood	DDH unilateral and bilateral + severity (0-3)	SNP associated to DDH
Tian et al. (2012)	SNP	China	Blood	DDH unilateral and bilateral	1 SNP associated to DDH
Feldman et al. (2013)	Region and SNP	USA	Blood or Buccal	3 DDH sign: SL, CEA < 20, TA > 10, EI > 10 mm, FN angle	9 SNP in LOD significant region in chr 3, 96 shared variants in candidate region, 6 variants involved in bone formation
Liu et al. (2013)	SNP	China	Blood	KL > 2	rs143383 significantly associated to OA
Sekimoto et al. (2013)	SV	Japan	Blood	AD (SA > 42)	2 Contigs of CNV identified on chr9 and 4, CNV loss associated with AD on chr 9
Zhao et al. (2013b)	SNP	China	Blood	DDH (ultrasound and radiographic evidence)	3 variants identified, more variations in DDH patients
Zhao et al. (2013a)	SNP	China	Blood	AA > 25, CEA < 20, SA > 45, SL disrupted, AHI diff 84-85%	Both SNP associated to DDH
Eytan et al. (2014)	Genes and RNA	Israel	Blood, Buccal or Skin	PPKP1 and CDH	1 mutation segregate with CDH and PPKP1, c.451 + 1G>A results in an in-frame deletion and a loss of 30 amino acids in p34
Feldman et al. (2014)	SNP	USA		3 or more sign of DDH vs 1000G	Allele A frequency in affected suggestive
Hao et al. (2014)	SNP	China	Blood or Buccal	DDH unilateral and bilateral + severity (0-3) (clinical and radiographic evidence)	1 SNP associated in dominant model
Liu et al. (2014)	SNP	China	Blood	Acetabulum angle > 25, SA > 45, CEA < 20, SL, AHI diff 84-85%	
Shi et al. (2014)	SNP	China	Blood	Unilateral or bilateral DDH	Genotypes different between the two studies, no significant association

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Aref-Eshghi et al. (2015)	Genes	Canada	Cartilage FH	Primary OA	4 Genes differently expressed in OA
Čengić et al. (2015)	SNP	Croatia	Blood	CEA < 20 + severe OA vs only OA	Both SNP associated to DDH
Sun et al. (2015)	SNP	China	Blood or Buccal	DDH (radiographic evidence)	51 SNP significant < 10e-4 and 21 SNP in UQCC, but SNP poorly associated in validation
Watson et al. (2015)	Region and SNP	Afrikaner	Blood	Beukes HD	1,33 Mb region with 16 genes, 6 variants and only 1 shared in UFSP2, all 17 BHD het and penetrance of 80% equivalent to rs796052130
Yilmaz et al. (2015)	STR	Turkey	Blood	DDH	No association
Xu et al. (2016)	SNP	China		DDH (clinical and radiographic evidence, complete dislocation)	No association
Basit et al. (2017)	Genes and Region	Saudi	Blood	Isolated form of DDH	None of the genes associated, shared haplotype in chr 1 by all DDH, 59 new variants and only 2 linked to DDH
Li et al. (2017)	SNP	China	Blood	DDH (clinical and radiographic evidence)	2 SNP associated to DDH
Ma et al. (2017)	SNP and Region	China	Blood	Unilateral or bilateral DDH + severity (0-3)	1 SNP in TGFB1 nominally significant, 2 SNP in IL6 and TGFB1 nomillay significant and 1 haplotype linked to DDH
Qiao et al. (2017b)	SNP	China	Blood	DDH (clinical and radiographic evidence)	No association of 2 SNP with DDH
Qiao et al. (2017a)	SNP	China	Blood or Buccal	Unilateral or bilateral complete dislocation	rs10250905 in TXNDC3 in DDH, protective T allele is lower in DDH
Sekimoto et al. (2017)	SV	Japanese	Blood	AD (SA >= 42, CEA <= 25, Acetabular roof obliquity >=15, MJSW)	More CNV in AD patient with worse radiographic measure
Wang et al. (2017)	Genes	China	Bone marrow	Trauma fracture vs DDH	Scin mRNA higher in DDH
Zhao et al. (2017)	InDels	China		Perkin quadrant, Acetabular index > 25, SL, CEA < 20	2 in frame mutation associated to DDH, BMP2K variants present in DDH sporadic and in unaffected relatives
Aki et al. (2018)	Genes	Japan	FH cartilage	Secondary OA to AD, CEA < 20 vs Neck of Femur Fracture	23 upregulated and 12 downregulated genes (FC > 10 or < 0.1), Collagen fibril organization and Extracellular matrix with q-values < 0.1, 51 genes upregulated mostly ECM-receptor interaction, focal adhesion, protein digestion and absorption pathways

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Basit et al. (2018)	Genes	Saudi	Blood	DDH (clinical and radiographic evidence)	1 common SNP in CX3CR1 but not penetrant (modifier possibly), 1 CNV (+15Kb) in chr6 associated with DDH, 36 (17 new) rares variants in autosomal dominant inheritance model with incomplete penetrance
Hatzikotoulas et al. (2018)	SNP	UK	Buccal	Self reported DDH, 25% radiographic validated	11 SNP on chr20 in 1 loci
Hatzikotoulas et al. (2018)	SNP and Genes	UK	Buccal	HD (indication for hip replacement + pelvic radiographs for 25%)	3 SNP in chr20, GDF5: 2 SNP with high epeigenomic scores, 1 SNP as probable causal variant, 5 Genes significantly associated, OA and DDH positively genetically correlated
Jawadi et al. (2018)	SNP	Saudi	Blood		Negative result of all 4 SNP with VDR
Sadat-Ali et al. (2018)	SNP	Saudi	Blood	DDH (clinical, ultrasound and radiographic evidence)	Allele T > in cases vs mother, negatif otherwise (Father, Sibling, Control)
Zhang et al. (2018)	SNP	China	Blood or Buccal	DDH (clinical and radiographic evidence)	5 SNP associated and 1 haplotype overly represented in DDH
Baghdadi et al. (2019)	Genes	Iran	Femoral Neck cartilage	DDH diagnosis at the orthopedics clinic	Excess of methylation of GDF5 in DDH
Feldman et al. (2019)	SNP	USA		3 signs (SL, CEA < 20, TA > 10, EI > 10 mm, FN angle) vs 1000Genome	20 SNP shared by 3 DDH and relevant to chondrogenesis or osteogenesis and 1 variant with autosomal dominant mode of inheritance with incomplete penetrance
Sun et al. (2019b)	SNP and RNA	China	Blood, Buccal, Hip joint capsule and ligament	DDH (radiographic evidence)	SNP associated with DDH, Higher expression of WIF1 in DDH dosage dependent with A allele, low association between 2 SNP, GDF5 expression downregulated with allelic differences in hip cartilage of DDH and upregulated in the repaired cartilage in a rabbit cartilage defect model
Yan et al. (2019)	SNP and RNA	China	Hip joint capsule or hip cartilage	Complete dislocation of the hip	4 SNP (GWS) in 3 chr, No difference in expression
Yan et al. (2019)	Genes	China		DDH	3 genes set enriched (positive regulation of biological process, single-organism process, biological regulation) and 3 enriched pathway (type II diabetes mellitus, JAK-STAT signaling, endocytosis)
Zhu et al. (2019)	SNP and Genes	China	Blood	Unilateral or bilateral DDH vs unaffected father	From 1344 genes 5 potential mutation in DDH genes, molecular function, cellular component and biological process enrichment investigated

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Gumus et al. (2021)	SNP	Turkey	Blood	DDH (clinical, ultrasound and radiographic evidence)	CX3CR1 polymorphism significantly associated with DDH
Luo et al. (2021)	Genes	China		DDH	25(OH)D decreased as well as miR-140-5p and miR-140-3p in DDH
Wang et al. (2020)	Genes	China	Articular cartilage	Periacetabular osteotomy vs FNF	DDH patient have lower expression level of lncRNA H19, COL2A1, Sox9 and Aggrecan but higher MMP3 and Adamts5
Xu et al. (2020)	SNP and Genes	China	Blood, Buccal or Hip joint cartilage and ligament	DDH (radiographic evidence)	SNP associated to DDH in genotypes models and lower expression of COL11A2 in DDH
Zhang et al. (2020)	Region	China	Blood	Familial DDH bilateral hip dislocation (Tonnis grade 4)	2 loci on chr 8 of 2.4 Mb and chr 12 of 0.6 Mb
Cheng et al. (2021)	Genes and RNA	China	Proliferating primary chondrocytes from Hip cartilage	DDH with Crowe type II vs FNF (Garden III-IV)	1833 significant mRNA DEGs for DDH and 982 significant GO terms, 175 significant pathways, 186 significant miRNA DEGs for DDH and 11 significant GO terms, one significant pathway, 175 differential expressed genes from the miRNA-mRNA expressed regulatory networks, 111 GO terms and 14 pathways for DDH
Cheng et al. (2021)	Genes	China		OA vs FNF compared to DDH vs FNF	1% showed regulation in the same direction and EFNA1 is up regulated in DDH and down in OA
Harsanyi et al. (2021b)	SNP	Slovakia	Blood	Isolated DDH (Graph > stade I)	GDF5 SNP barely associated to DDH
Harsanyi et al. (2021a)	SNP	Slovakia	Blood	DDH (ultrasound evidence)	SNP in GDF5 significantly associated
İgrek et al. (2021)	SNP	Turkey	blood	Hartofilakidis and IHDI classifications	No statistically significant relationship of TGFB1 and IL-6 SNPs for DD
Li et al. (2021)	Genes	China	Hip joint capsule	DDH (MRI and radiographic evidence) vs FNF	419 upregulated and 675 downregulated genes in DDH patients, GO and KEGG terms investigated, Slower remodelling and restoration of the ECM, type I and III collagens weaker in hip joint capsule of DDH, 14 genes up and 3 down regulated, DDH fibroblasts in the hip joint capsule inclined to differentiate into myofibroblasts, reduced cell viability but no difference in apoptotic rate, reduced Type I collagen staining, increased percentage in the G2/M phase, but decreased percentage in the S phase and EdU-positive staining reduced

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Topak et al. (2021)	Genes	Turkey	Blood	Treated DDH (clinical, ultrasound and radiographic evidence)	No significant difference for Ca, P, ALP, vitamin D but mean VDR levels lower in DDH
Xu et al. (2021a)	SNP and Genes	China	Blood, Buccal, Hip joint tissue	DDH (clinical and radiographical evidence)	DDH have higher expression levels of FRZB
Xu et al. (2021b)	SNP	China	Blood	Unilateral or bilateral DDH (clinical and radiographic evidence)	None of the 4 variations detected
Dembic et al. (2022)	SNP, Genes and SV	Danish	Blood	DDH (clinical and radiographic evidence) vs 2000 Danish diabetic patients	2640 rare and shared variants in a total of 2276 genes, carried by related individuals accross 28 families, only two variants are predicted to be damaging in more than 1 family, large protein network with 8 clusters, significantly interaction than random, novel variants in 17 known genes, one CNV variant detected in 2 families, 36 genes carrying 37 very damaging variants, with more than one in many families and 1 gene involved in 9 families and 4 in 5 families
Gumus et al. (2022)	RNA	Turkey	Plasma	DDH (medical anamnesis, clinical, ultrasound and radiographic evidence)	miR-96 levels decreased in DDH group, 145 biological processes, 96 cellular components, 63 molecular functions (MF) were enriched by the 872 genes targeted by mirR-96 and 26 KEGG pathways
Wang et al. (2022)	RNA and Genes	China	Cartilage	DDH and Kellgren-Lawrence grade (OA) vs FNF	21 miRNAs dysregulated. miR-34a-5p negatively correlated to SESN2 by blocking its expression, it supress mTORC1 signalling and inhibit chondrocyte autophagy, SESN2 expression reduced in DDH and negatively correlated with OA development
Xu et al. (2022b)	Genes	China	Blood	Hip dislocation	KANSL1 missense segregate with dominant mode of inheritance, 1 new variant
Yan et al. (2022)	SNP	China	Blood	DDH (complete or incomplete dislocation)	2 rare variants identified in LRP1 in two families, 5 missense variants and 3 splice-site variants and LRP1 is LoF-intolerant genes
Yang et al. (2022)	Genes				No collinear genes, proteins structures are diverse, co-expression is the most common, KEGG and GO terms investigated, some mutations in the genes are involved in cancers
Demirel et al. (2023)	Genes	Turkey	Blood from umbilical cords	Graf type 2 or 3	IGF-1 and IGF-BP3 levels are similar but IGF-BP5 lower in DDH
Jacobsen et al. (2023)	SNP	Norway, UK		DDH or OA secondary to DDH	2 SNP reached GWS on chr 1 in LD with COL11A1, only COL11A1 gene significantly associated, 10 GWS SNP in GDF5 and UQC1

TABLE S9 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez l'humain

Article	Type	Population	Sample	Phenotype	Conclusion
Lu et al. (2023)	Genes	China			Novel variant in FNLA causing sulfhydryl side chain disruption
Mori et al. (2023)	SNP	Japanese, UK	Blood	DDH (hip pain, radiographic findings, CEA < 20)	4 GWS loci in chr 1 15 17, no replication of the japanese SNP in UK population but another SNP reached GWS in chr 12
Mori et al. (2023)	Genes	Japanese	FH cartilage	Secondary OA with DDH vs FNF	1079 and 822 genes were significantly expressed, GP6 signaling pathway top hit and ferroptosis signaling pathway, 42 genes assigned, no significant pathway enrichment, 1 ingenuity pathway enriched Japanese (ferroptosis signaling pathway), ToppGene disease analysis no enrichment, GCH1 with lower expression level in DDH
Mori et al. (2023)	Genes	UK		M161, M162, and M163 in ICD10	96 genes assigned to UK GWAS with no overlap, no significant pathway enrichment. No ingenuity pathway enriched. ToppGene two disease enriched in UK (Osteoarthritis of the hip, Bipolar disorder) with four DDH GWAS-related genes
Gombosh et al. (2024)	SNP and SV	Israel	Blood		1 loci on chr 1, 1 variant in TRIM33 as likely causative variant, all affected homozygous, no other homozygotes in controls
Gombosh et al. (2024)	Genes	Israel	Fibroblast	Cell from WT and mutant patient	DLX5 downregulated in mutant, BGLAP and ALPL upregulated, RUNX2 and TRIM33 not different
Meng et al. (2024)	SNP	China	Blood	DDH (clinical and radiographic evidence)	Rare genotype of HIBCH rs291412 has a protective effect (dominant model), common genotype of FTCDNL1 rs769956 is a risk factor (recessive model). T allele of HIBCH and A allele in FTCDNL1 reduce expression of corresponding gene
Sarıkaya et al. (2024)	Genes	Turkey	Ligament teres	Tonnis stage	Correlation between MMP-2, MMP-9, and ADAMTS-7 and all histological parameters.
Miao et al. (2025)	Genes	China	Blood	DDH (radiographic evidence)	COMP c.1133A>C prioritised as likely pathogenic in trio and confirmed by pedigree
Rao et al. (2025)	RNA	China	Joint capsule tissues	DDH radiographically confirmed Crowe II-IV	155 DE-CMRGs and 5 biomarkers downregulated in DDH, only CCNB2 significantly downregulated

TABLE S10 - Liste des études génétiques sur la Dysplasie Coxo-Fémorale réalisées chez le chien

Races : PWD (Portuguese Water Dog), LR (Labrador Retriever), Gh (Greyhound), GS (German Sheperd), BMD (Bernese Mountain Dog), Rot (Rottweiler), Nfl (Newfoundland), BC (Border Collie), GR (Golden Retriever), DDBx (Dogue de Bordeaux), EngS (English Spaniel). Phenotype : DI (Distraction Index), NO (Norberg Angle), OA (Osteoarthritis), DLS (Dorso Lateral Score), OFA (Orthopedic Foundation for Animals score), PCA (Principal Component Analysis), OSS (femoral capital ossification), FCI (Fédération Cynologique Internationale score), EHR (Extended Hip Radiography score), Sub (Subluxation), CrAE (Cranial Acetabular Edge), FHCD AE (femoral head center position in relation to dorsal acetabular edge), BVA (British Veterinary Association score), HD (Hip Dysplasia). Conclusion : QTL (Quantitative Trait Loci), CFA (Canis Familiaris Autosome), SNP (Single Nucleotide Polymorphism), GWS (Genome Wide Significance), EBV (Estimated Breeding Value), DE (Differentially Expressed).

Article	Type	Breed	Sample	Phenotype	Conclusion
Wang et al. (1999)	Region	Boyken Spaniel	Blood	DI $\geq$ 0.6	1 primer mildly associated but need more investigation
Chase et al. (2002)	STR	PWD	Blood	NO	2 QTL, one for each hip side on chr1
Mäki et al. (2004)		GS, GR, LR, Rott		Mean grade (1-6) left and right hip	Major gene variance significant in all breeds, with allele effects indicating dominance, but unfavourable allele frequencies remained high, especially in LR
Chase et al. (2005)	STR	PWD	Blood	OA scale (0-3)	1 QTL close to IGF1 on chr3
Mateescu et al. (2005)	STR	LR, Gh, LRxGh	Blood	DI, DLS, NO	2 QTL on chr6
Todhunter et al. (2005)	STR	LR, Gh, LRxGh	Blood	DI, DLS, NO, OFA, PCA	14 QTL
Tsai & Murphy (2006)	STR	Boykin spaniel	Blood	DI $>$ 0,6	Not enough power
Liu et al. (2007)	STR	LR, Gh, LRxGh	Blood	OSS	4 general QTL and 4 specific to one side
Marschall & Distl (2007)	STR	GS	Blood	FCI A vs B/C/D/E	15 QTL on 13 chr
Mateescu et al. (2008)	STR	LR, Gh, LRxGh	Blood	Necropsy score $>$ 0	4 QTL
Zhu et al. (2008)	STR	LR, Gh, LRxGh, GS	Blood	DI $>$ 0,7	2 QTL on chr11 and chr29
Foundation of the University of Veterinary Medicine Hannover (2009)	SNP	GS, Rottweiler, BMD, Hovawart			Risk score of 17 markers
Phavaphutanon et al. (2009)	STR	LR	Blood	NO $<$ 105, DLS $<$ 45%, DI $>$ 0.7, OFA, PCA	8 QTL on 6 chr
Zhou et al. (2010)	SNP	LR, Gh, LRxGh, GS, BMD, BC, Rot, Nfl	Blood	NO or OA (OFA, Necropsy)	4 SNP near EVC, EVC2, PTRD, COL15A1 and MAGP1 for NO and 2 SNP near REG3A and PARD3B for OA

TABLE S10 - Liste des études génétiques sur la Dysplasie Coxo-Fémorale réalisées chez le chien

Article	Type	Breed	Sample	Phenotype	Conclusion
Friedenberg et al. (2011)	Genes	LR, Gh, LRxGh, BC, GS, GR, Nfl, Rot, GD	Blood	NO, DI, DLS, EHR	Haplotype in intron 30 segregating with CHD in LR, intron 30 of FBN2 7bp homozygous deletion AGC haplotype worsen hip score
Friedenberg et al. (2011)	RNA	LRxGH	Capsule, Round ligament, Cartilage	OA, NO, DLS, EHR	Dogs with OA have 50% more FBN2 mRNA, dogs with wildtype haplotype have more FBN2 expression in FH, KIN11 mRNA expression greater in OA joint capsule
Guo et al. (2011)	SNP	LR, Gh, LRxGh, GR, GS, Rot, Nfl		NO	EBV and GBV are correlated, CHD can be predicted from genomic data with 100 - 500 SNP
Pfahler & Distl (2012)	SNP	BB	Blood	FCIc	3 QTL and 1 Haplotype on chr 37
Zhu et al. (2012)	STR	LR, Gh, LRxGh, GR, GS, Nfl, Rot, BM		NO	36 QTL on 15 chr and 17 QTL on 11 chr
Fels et al. (2014)	SNP	GS	Blood	FCI (A vs CDE)	10 SNP associated and 3 to 9 SNP validated on 5 chr
Fels & Distl (2014)	SNP	GS	Blood	FCI (A vs CDE)	5 suggestive SNP on 4 chr and 3 SNP validated 3 chr
Lavrijsen et al. (2014)	SNP	LR	Blood or Buccal	FCI (A vs CDE)	1 SNP on chr8, 18 SNP on 7 chr and 31 SNP on 6 chr
Sánchez-Molano et al. (2014)	SNP	LR, GR, GS	Blood	FCI (AB or A vs DE)	26 Loci (165 SNP) in 12 chr, CHST3 gene, 9 SNPs in 7 Loci on chr 4 and 20 SNP on 10 chr
Sánchez-Molano et al. (2014)	SNP	LR	Saliva	NO, CrAE	4 SNP for CrAE and NO on chr 1 and 21
Hayward et al. (2016)	SNP	69 breeds	Blood	NO	1 Loci on chr 28 in gene CTBP2
Fealey et al. (2017)	SNP	GS, GR, LR, Nfl, LRxGH, 46 others	Blood	PCA (13 measures), NO, Angle of inclination	2 Loci for PCs but not for NO chr 15 and 16
Huang et al. (2017)	SNP	LR, GR, GS, English Setter, Nfl, 347 other	Blood	NO	3 SNP associated to HD
Korea Genetic Information Research Institute Co., Ltd. (2017)	SNP				21 SNP use but only 5 significant
Manz et al. (2017)	SNP	GS	Blood	FCI	Genetic test unsuitable for individual risk assessment

TABLE S10 - Liste des études génétiques sur la Dysplasie Coxo-Fémorale réalisées chez le chien

Article	Type	Breed	Sample	Phenotype	Conclusion
Mars Inc (2018)	Genes		Blood	FCI (A vs CDE)	10 SNP on 6 chr
Mikkola et al. (2019b)	SNP	GS	Blood	FCI (A vs DE or CDE)	11 SNP on chr9, 6 SNP on chr9 and 9 SNP chr 1 and 9
Mikkola et al. (2019a)	SNP	GS	Blood	FHCDAE and NO	2 Loci associated to FHCDAE on chr 9 and 28 (NOG, CACUL1)
Mikkola et al. (2019a)	SNP	GS	Blood	OA and FCI (A vs CDE)	1 Loci on chr 1 (NOX3, ARID1B)
Todhunter et al. (2019)	Genes	LR, LRxGH, BexCS	Ligament, Whole Hip, Capsule	NO < 100, DLS < 45%, DI	COL1A1, COL1A2, and COL3A1 highly expressed, 2 pathways upregulated, 2 downregulated, 149 Genes DE and 24 already associated in litterature for the capsule and 29 Genes DE and 4 already associated in litterature for the ligament
Akis et al. (2020)	SNP	16 breeds	Blood	FCI (A vs BCDE)	1 SNP on chr 19
Bruun et al. (2020)	SNP	LR (Danish)		FCI (A/B vs D/E)	No significant correlation between the Dysgen test result and the radiographic hip status
Kang et al. (2020)	SNP	GS and BM (Korean)	Blood	NO < 79	1 SNP on chr 4 for left NO
Jiang et al. (2021)	SNP	GS, EngS, LR, GR, Rot, Nfl (USA)		NO	Both model give similar medium level prediction
Mikkola et al. (2021)	SNP	10 Breeds		FCI (A vs CDE)	4 SNP on 4 chr, Neddylation pathway enriched and multiple breed specific loci
Wang et al. (2021)	SNP	GS (UK, Finland, Sweden)	Buccal or Blood	FCIc, FCI (A vs BCDE) or BVA > 10	1 Loci and 2 SNP 3 chr and country specific loci
Atalay & Kok (2022)	SNP	BM, LR, GS	Blood	NO < 105	2 SNP inf CX3CR1 and 1 interaction between 2 SNP (co-dominant)
Binversie et al. (2022)	SNP	13 to 38 Breeds		OFA or Hospital HD (low vs high risk breed)	66 loci identified in 10 different scenarios
Alilovic et al. (2023)	Genes	Tornjak	blood	FCI (A/B vs C/D/E)	CHD dog has lower HA and higher PIICP and MMP9
Padula et al. (2023)	Genes	GS	blood	DJD secondary to primary disease	2 DGE upregulated and 1 downregulated in 3 pathway
Ardicli et al. (2024)	SNP	GS, LR	blood	3 class from absence to severe for CHD and ED	None of the SNP associated to CHD but 3 to ED
Kieler et al. (2024)	SNP	LR	Blood	FCI (A to E) or HD-EBV	5 SNP on chr24, 0,7 Mb Locus and individuals homozygous for risk allele have lower EBV

TABLE S10 - Liste des études génétiques sur la Dysplasie Coxo-Fémorale réalisées chez le chien

Article	Type	Breed	Sample	Phenotype	Conclusion
Ardicli et al. (2025)	SNP	GS, LR, German Wirehaired Pointer	blood	FCI A/B vs C/D/E	1 SNP weakly associated in multivariate model
Topparmak et al. (2025)	SNP and RNA			CHD (physical and radiological examination)	4 SNP detected, 2 associated to CHD and 11 miRNAs targeting 3' UTR CHST14

TABLE S11 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez d'autres organismes

Sample : FH (Femoral Head), Ch. (Chondrocytes). Conclusion : DDH (Developmental Dysplasia of the Hip), OA (Osteoarthritis), LoF (Loss of Function).

Article	Type	Organism	Sample	Phenotype	Conclusion
Bo et al. (2012)	Pheno and Genes	Wistar Rat	Hip joint, Articular cartilage	Swaddling DDH	Size of Acetabulum and FH smaller, with abnormal development, loss of proteoglycans, cartilage degenerates and fewer Ch. appears in DDH model, Expressions of Collagen X and MMP-13 increased in Articular Cartilage of the DDH Model as well as mRNA levels Collagen X, MMP-13, and ADAMTS-5
Feldman et al. (2017)	Genes	Mice 4167		C57BL/6NTac-[KO]CX3CR1 mouse	Acetabulum diameter and inter space joint larger in KO mice with impaired gait
Wang et al. (2017)	Genes	Rat, Mice		Swaddling DDH	Scin mRNA higher in DDH, and increase osteoblast in rats, Scin mRNA upregulated with Ccs and downregulated chondrogenesis in mice
Kiapour et al. (2018)	Genes	BALB/cJ bp3J Mice strain cross with FVB-NJ strains	Hip	Gdf5 loss-of-function	Gdf5 loss results in substantial dysmorphologies to the proximal femur and acetabulum. Downstream regulatory sequences of Gdf5/GDF5 drive stronger, more ubiquitous prenatal hip expression and uniquely control postnatal hip expression
Chen et al. (2019a)	Genes	L929 fibroblast	Cell	Cas9 KO PAPP2 or Cas9/PAPP-A2 sgRNA	PAPP-A2 is downregulated, increasing the concentration of IGFBP-5, downregulating proteins involved in the IGF signaling pathway
Chen et al. (2019a)	Genes	ICR mice	Acetabular and FH cartilage	Cas9/PAPP-A2 sgRNA	COL2 and ACAN protein and mRNA downregulated, no significant results in the soft tissue
Chen et al. (2019b)	Genes	Wistar	Hip joint	Swaddling DDH	IGF1 downregulated while IGF1R up regulated at 14 days
Feldman et al. (2019)	SNP	Tenm3 mice			Mutation slow down chondrogenesis
Ning et al. (2019)	Genes	Wistar rats Ch.	Ch. from acetabulum cartilage	Two shRNAs targeting H19	lncRNAH19 KD reduce Ch. proliferation capacity through direct binding to miRNA let-7 but not apoptosis
Ning et al. (2019)	Genes	Wistar rats	Ch. from acetabulum cartilage	Swaddling DDH	LncRNA H19, Col2A1 are downregulated in DDH
Sun et al. (2019a)	Genes	BMSC	Ch.like cells	GDF5 siRNA or Ad-GDF5	GDF5 regulated chondrogenesis, enhance ectopic cartilage generation and enhanced migration

TABLE S11 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez d'autres organismes

Article	Type	Organism	Sample	Phenotype	Conclusion
Sun et al. (2019a)		Rabbit			3D-bioprinted GDF5-conjugated BMSC-laden scaffold in rabbit knees conferred long-term chondroprotection
Ji et al. (2020)	SNP	Ch.		Gain and loss WISP2 or PPAR	WISP-2 overexpression induces Ch. apoptosis. Silencing WISP-2 (shRNAs) decrease WISP2 level, enhance cell viability, reduce apoptosis. WISP-2 regulates PPAR? expression during Ch. apoptosis and PPAR? attenuates the effects of WISP-2 on Ch. viability and apoptosis
Ji et al. (2020)	SNP	Wistar rat	Hip acetabular cartilage complex, acetabulum and FH	Swaddling DDH	104 mRNAs were upregulated and 142 mRNAs were downregulated, 29 KEGG pathway enriched, decline in the size of the acetabulum and FH, same level of proteoglycans but unevenly distributed, increase of Ch. apoptosis
Liu et al. (2020)	Genes and RNA	New Zealand rabbits	Acetabular roof cartilage	Swaddling DDH	Acetabulum presented with an increased inclination, the FH with lateral displacement, and SL was not continuous, collagen type X and RUNX2 expression downregulated in DDH samples, 13miRNA upregulated and 15 downregulated, miR-129-5p expression is reduced in DDH and its downregulation inhibits Ch. proliferation and hypertrophy and reduces mineralization via GDF11/Smad3
Luo et al. (2021)	Genes	Wistar rats		Swaddling DDH	miR-140-5p downregulated in the DDH group (day 2 and 8) miR-140-3p only at day 2
Wang et al. (2020)	Genes	ICMS		Intermittent cyclic mechanical stress	ICMS promoted cartilage degeneration and caused reorientation of the cytoskeleton and lncRNA H19 helped inhibit cartilage degeneration. lncRNA H19 and Dusp5 are both direct targets of miR-483e5p
Wang et al. (2020)	Genes	Mice	Hip and knee joints	Swaddling DDH	Smaller and dislocated FH with worse cartilage conditions, lower expression level of lncRNA H19, COL2A1 and Aggrecan but higher MMP3 and Adamts5, pUbi-H19 effectively alleviated cartilage degeneration triggered by DDH, whereas si-H19 worsened the cartilage degeneration
Xu et al. (2020)	Genes	Mice	Cell	rs9277935 polymorphism	Allele T increase luciferase expression of COL11A2
Ding et al. (2021)	RNA	Rabbits Ch. and human primary Ch.	Cell	Swaddling DDH	Downregulation of miR-1-3p expression inhibits Ch. hypertrophy and reduces mineralization in vitro via SOX9 and reduce RUNX2 and collagen X reduced
Ding et al. (2021)	RNA	New Zealand rabbits	Hip joint	Swaddling DDH	Rabbits with DDH exhibited thick ARC and low miR-1-3p expression
Xu et al. (2021a)	Genes	ATDC5	Cell		KO FRZB up and down regulate multiple genes from different pathway, FRZB modulated cell adhesion pathway and cell spreading by regulating integrins expressions

TABLE S11 - Liste des études génétiques sur la Dysplasie Développementale de la Hanche réalisées chez d'autres organismes

Article	Type	Organism	Sample	Phenotype	Conclusion
Xu et al. (2021a)	Genes	BMSC	Cell		FRZB induced chondrogenesis and Wnt signaling pathway in vitro in BMSCs
Xu et al. (2021a)	Genes	BMSC and ATDC5 cell	Synovial fluid	DDH (clinical and radiographic evidence)	Downregulated miRNA-454 expression causes FRZB upregulation in the synovial fluid of DDH patients
Xu et al. (2022a)	Genes	Ch. C28/I2 and Osteoblast hFOB1.19	Cell	pcDNA3.1-WNT1 or si-WNT1	WNT1 promotes proliferation of osteoblast and influences NOG, RBPMS2, BMP2, BMP4, WNT1, WISP2 expression
Xu et al. (2022a)	Genes	mESC	Cell	CrisprCas9 KANSL1 missense mutant	Mutation do not impact pluripotent stem cell state, KANSL1 expression not affected but decrease cell proliferation and cartilage generation, MMP13 upregulated in mutant. KANSL1 missense variant can cause degeneration of cartilage
Xu et al. (2022a)	Genes	Mice BALB/c	Hip joint	pcDNA3.1-WNT1 or si-WNT1	pcDNA3.1 WNT1 improve development and increase BMP2, BMP4, WNT1 and WISP2 levels. Silencing WNT1 promotes DDH and decrease BMP2, BMP4, WNT1 and WISP2 levels
Xu et al. (2022a)	Genes	Mice	Cartilage	CrisprCas9 KANSL1 missense mutant	Missense variant leads to defect in acetabular cartilage
Yan et al. (2022)	Genes	BMSC			Decreased chondrogenic ability of Lrp1 +/2 BMSCs.
Yan et al. (2022)	Genes	ATDC5		shRNA-Lrp	LRP1 deficiency suppresses autophagic degradation of β -catenin. Inhibition of β -catenin rescues the decreased chondrogenic ability caused by LRP1 deficiency in ATDC5. The level of autophagy increased gradually during chondrogenesis
Yan et al. (2022)	Genes	Mice KI	Hip joint	CRISPR-Cas9 Lrp1R1783W	c.5347C > T can cause LRP1 LoF with a gene dosage effect and mice with mutation mimic DDH
Liu et al. (2024)	Genes and RNA	Wistar rats	Acetabular cartilage and Proliferative zone	Swaddling DDH	Ch. in the proliferative zone of the acetabular cartilage in DDH rats show increased proliferation and decreased hypertrophic differentiation, 986 upregulated and 468 downregulated at 5 days, 655 upregulated and 216 downregulated at 10 days. Downregulated DEGs in the DDH group are more critical for DDH development. Cartilage development enriched in both GO and KEGG analyses, 2 gene modules with functions related to bone and cartilage development, 6 genes classified as hub genes, hub genes are downregulated in DDH group while novel miRNA are upregulated.

9 Bibliography

- ADAMS, W. M. et al., 2000. Comparison of Two Palpation, Four Radiographic and Three Ultrasound Methods for Early Detection of Mild to Moderate Canine Hip Dysplasia. *Veterinary Radiology & Ultrasound* [en ligne]. Vol. 41, no. 6, p. 484–490 [visit  le 2022-11-04]. ISSN 1740-8261. Disp.   l'adr. DOI: 10.1111/j.1740-8261.2000.tb01875.x.
- AERSSSENS, J. et al., 1998. Interspecies Differences in Bone Composition, Density, and Quality: Potential Implications for in Vivo Bone Research. *Endocrinology*. Vol. 139, no. 2, p. 663–670. ISSN 0013-7227. Disp.   l'adr. DOI: 10.1210/endo.139.2.5751.
- AHNER, C. E. et al., 2019. Protein Biomarkers in Serum and Urine for Determining Presence or Absence of Hip Dysplasia in a Canine Model. *Journal of Orthopaedic Research®*. Vol. 37, no. 4, p. 916–920.
- AKI, T. et al., 2018. A Whole-Genome Transcriptome Analysis of Articular Chondrocytes in Secondary Osteoarthritis of the Hip. *PLOS ONE* [en ligne]. Vol. 13, no. 6, e0199734 [visit  le 2024-02-02]. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0199734.
- AKIS, I. et al., 2020. The Association of Genetic Polymorphisms of Bone Formation Genes with Canine Hip Dysplasia. *Iranian Journal of Veterinary Research*. Vol. 21, no. 1, p. 40–45. ISSN 1728-1997. Disp.   l'adr. PMID: 32368224.
- ALEXANDER, J. W., 1992. The Pathogenesis of Canine Hip Dysplasia. *Veterinary Clinics of North America: Small Animal Practice* [en ligne]. Vol. 22, no. 3, p. 503–511 [visit  le 2020-08-05]. ISSN 0195-5616. Disp.   l'adr. DOI: 10.1016/S0195-5616(92)50051-1.
- ALILOVIC, I. et al., 2023. Breed-Specific Evaluation of Serum Biochemical Markers in Canine Hip Dysplasia Observed in a Tornjak Dog Population. *American Journal of Veterinary Research* [en ligne]. Vol. 84, no. 11 [visit  le 2023-11-09]. ISSN 0002-9645, ISSN 1943-5681. Disp.   l'adr. DOI: 10.2460/ajvr.23.07.0170.
- ARCHER, C. W. et al., 2003. Development of Synovial Joints. *Birth Defects Research Part C: Embryo Today: Reviews* [en ligne]. Vol. 69, no. 2, p. 144–155 [visit  le 2020-08-07]. ISSN 1542-9768. Disp.   l'adr. DOI: 10.1002/bdrc.10015.
- ARDICLI, S. et al., 2024. The Markers of the Predictive DNA Test for Canine Hip Dysplasia May Have a Stronger Relationship with Elbow Dysplasia. *Heliyon* [en ligne], e37716 [visit  le 2024-09-12]. ISSN 2405-8440. Disp.   l'adr. DOI: 10.1016/j.heliyon.2024.e37716.
- ARDICLI, S. et al., 2025. The Effectiveness of Genetic Markers and the Role of Environmental Factors in Hip Dysplasia and Osteochondritis Dissecans of the Shoulder in German Shepherd, Labrador Retriever, and German Wirehaired Pointer (Deutsch Drahthaar) Dogs. *Gene Reports* [en ligne]. Vol. 38, p. 102153 [visit  le 2025-02-03]. ISSN 2452-0144. Disp.   l'adr. DOI: 10.1016/j.genrep.2025.102153.
- AREF-ESHGHI, E. et al., 2015. Overexpression of MMP13 in Human Osteoarthritic Cartilage Is Associated with the SMAD-independent TGF- β Signalling Pathway. *Arthritis Research & Therapy* [en ligne]. Vol. 17, no. 1, p. 264 [visit  le 2022-11-10]. ISSN 1478-6354. Disp.   l'adr. DOI: 10.1186/s13075-015-0788-x.
- ARONSON, E. et al., 1991. The Effect of Anesthesia on the Radiographic Appearance of the Coxofemoral Joints. *Veterinary Radiology* [en ligne]. Vol. 32, no. 1, p. 2–5 [visit  le 2022-11-04]. ISSN 1740-8261. Disp.   l'adr. DOI: 10.1111/j.1740-8261.1991.tb00074.x.

- ARONSSON, D. D. et al., 1994. Developmental Dysplasia of the Hip. *Pediatrics-English Edition*. Vol. 94, no. 2, p. 201–208.
- ATALAY, s. & KOK, S., 2022. The Effects of Polymorphisms in the CX3CR1 Gene on the Development of Canine Hip Dysplasia. *Turkish Journal of Veterinary and Animal Sciences*. No. 46, p. 74–87.
- AUTON, A. et al., 2013. Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *PLOS Genetics* [en ligne]. Vol. 9, no. 12, e1003984 [visit  le 2025-06-30]. ISSN 1553-7404. Disp.   l'adr. DOI: 10.1371/journal.pgen.1003984.
- BAGHDADI, T. et al., 2019. DNA Hypermethylation of GDF5 in Developmental Dysplasia of the Hip (DDH). *Molecular Genetics & Genomic Medicine* [en ligne]. Vol. 7, no. 9, e887 [visit  le 2024-02-01]. ISSN 2324-9269. Disp.   l'adr. DOI: 10.1002/mgg3.887.
- BAKKER, E. de et al., 2017. Canine Synovial Fluid Biomarkers for Early Detection and Monitoring of Osteoarthritis. *Veterinary Record*. Vol. 180, no. 13, p. 328–329.
- BALDINGER, A. et al., 2020. Prevalence of Canine Hip Dysplasia in 10 Breeds in France, a Retrospective Study of the 1997-2017 Radiographic Screening Period. *Plos one*. Vol. 15, no. 7, e0235847.
- BARLOW, T. G., 1962. Early Diagnosis and Treatment of Congenital Dislocation of the Hip. *The Journal of Bone and Joint Surgery. British volume* [en ligne]. Vol. 44-B, no. 2, p. 292–301 [visit  le 2022-11-04]. ISSN 0301-620X. Disp.   l'adr. DOI: 10.1302/0301-620X.44B2.292.
- BARR, A. R. S. et al., 1987. Clinical Hip Dysplasia in Growing Dogs: The Long-term Results of Conservative Management. *Journal of Small Animal Practice* [en ligne]. Vol. 28, no. 4, p. 243–252 [visit  le 2023-12-11]. ISSN 0022-4510, ISSN 1748-5827. Disp.   l'adr. DOI: 10.1111/j.1748-5827.1987.tb03879.x.
- BARTOLOM , N. et al., 2015. A Genetic Predictive Model for Canine Hip Dysplasia: Integration of Genome Wide Association Study (GWAS) and Candidate Gene Approaches. *PLoS One*. Vol. 10, no. 4, e0122558.
- BASIT, S. et al., 2017. Exome Sequencing Identified Rare Variants in Genes HSPG2 and ATP2B4 in a Family Segregating Developmental Dysplasia of the Hip. *BMC medical genetics*. Vol. 18, no. 1, p. 34. ISSN 1471-2350. Disp.   l'adr. DOI: 10.1186/s12881-017-0393-8.
- BASIT, S. et al., 2018. Whole Genome SNP Genotyping in a Family Segregating Developmental Dysplasia of the Hip Detected Runs of Homozygosity on Chromosomes 15q13.3 and 19p13.2. *Congenital Anomalies* [en ligne]. Vol. 58, no. 2, p. 56–61 [visit  le 2024-01-26]. ISSN 1741-4520. Disp.   l'adr. DOI: 10.1111/cga.12235.
- BEIGHTON, P. et al., 1994. Autosomal Dominant (Beukes) Premature Degenerative Osteoarthropathy of the Hip Joint Unlinked to COL2A1. *American Journal of Medical Genetics* [en ligne]. Vol. 53, no. 4, p. 348–351 [visit  le 2024-02-02]. ISSN 1096-8628. Disp.   l'adr. DOI: 10.1002/ajmg.1320530408.
- BELING, C. G. et al., 1975. Metabolism of Estradiol in Greyhounds and German Shepherd Dogs: An Investigation with Special Reference to Hip Dysplasia. *Acta Radiologica. Diagnosis* [en ligne]. Vol. 16, no. 344_suppl, p. 109–120 [visit  le 2020-08-06]. ISSN 0567-8056. Disp.   l'adr. DOI: 10.1177/0284185175016S34410.
- BINVERSIE, E. E. et al., 2022. Across-Breed Genetic Investigation of Canine Hip Dysplasia, Elbow Dysplasia, and Anterior Cruciate Ligament Rupture Using Whole-Genome Sequencing. *Frontiers in Genetics* [en ligne]. Vol. 13 [visit  le 2022-12-06]. ISSN 1664-8021.

- BIOCONDUCTOR PACKAGE, 2025. liftOver [en ligne]. Bioconductor. [visité le 2025-07-28]. Disp. à l'adr. DOI: 10.18129/B9.BIOC.LIFTOVER.
- BLOEBAUM, R. D. et al., 1993. Comparison of Human and Canine External Femoral Morphologies in the Context of Total Hip Replacement. *Journal of Biomedical Materials Research* [en ligne]. Vol. 27, no. 9, p. 1149–1159 [visité le 2022-10-05]. ISSN 0021-9304, ISSN 1097-4636. Disp. à l'adr. DOI: 10.1002/jbm.820270905.
- BO, N. et al., 2012. Early Cartilage Degeneration in a Rat Experimental Model of Developmental Dysplasia of the Hip. *Connective Tissue Research* [en ligne]. Vol. 53, no. 6, p. 513–520 [visité le 2024-02-02]. ISSN 0300-8207. Disp. à l'adr. DOI: 10.3109/03008207.2012.700346.
- BOEHME, K. A. & ROLAUFFS, B., 2018. Onset and Progression of Human Osteoarthritis—Can Growth Factors, Inflammatory Cytokines, or Differential miRNA Expression Concomitantly Induce Proliferation, ECM Degradation, and Inflammation in Articular Cartilage? *International Journal of Molecular Sciences* [en ligne]. Vol. 19, no. 8, p. 2282 [visité le 2023-12-10]. ISSN 1422-0067. Disp. à l'adr. DOI: 10.3390/ijms19082282.
- BOZKAN, Z. & SARIERLER, M., 2019. Influence of Anesthesia on Grading of Canine Hip Dysplasia. *Topics in Companion Animal Medicine*. Vol. 37, p. 100368. ISSN 1946-9837. Disp. à l'adr. DOI: 10.1016/j.tcam.2019.100368.
- BOZYNSKI, C. C. et al., 2020. Protein Biomarkers for Early Diagnosis of Canine Hip Dysplasia. In: *Veterinary Clinical Pathology* [en ligne]. Vol. 50, p. 99–118 [visité le 2021-12-20]. Disp. à l'adr. DOI: 10.1111/vcp.13011.
- BRADY, R. B. et al., 2013. Evaluation of Gait-Related Variables in Lean and Obese Dogs at a Trot. *American journal of veterinary research*. Vol. 74, no. 5, p. 757–762.
- BRAND, R. A., 2008. Etiology of Congenital Dislocation of the Hip: Carl E. Badgley MD (1893-1973) The 11th President of the AAOS 1942. *Clinical Orthopaedics & Related Research* [en ligne]. Vol. 466, no. 1, p. 90–103 [visité le 2022-11-04]. ISSN 0009-921X. Disp. à l'adr. DOI: 10.1007/s11999-007-0020-7.
- BRASS, W., 1989. Hip Dysplasia in Dogs. *Journal of Small Animal Practice* [en ligne]. Vol. 30, no. 3, p. 166–170 [visité le 2020-08-04]. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.1989.tb01525.x.
- BROECKX, B. et al., 2018. Comparison of Three Methods to Quantify Laxity in the Canine Hip Joint. *Veterinary and Comparative Orthopaedics and Traumatology*. Vol. 31, no. 1, p. 23–29.
- BRUUN, C. S. et al., 2020. Validation of DNA Test for Hip Dysplasia Failed in Danish Labrador Retrievers. *Animal Genetics*.
- BUSSE, J. et al., 1972. Eine neue Methode zur röntgenologischen Beurteilung eines Hüftgelenkes — Der Hüftwert. *Archiv für orthopädische und Unfall-Chirurgie, mit besonderer Berücksichtigung der Frakturenlehre und der orthopädisch-chirurgischen Technik* [en ligne]. Vol. 72, no. 1, p. 1–9 [visité le 2025-08-22]. ISSN 1434-3916. Disp. à l'adr. DOI: 10.1007/BF00415854.
- ČENGIĆ, T. et al., 2015. Association of TGFB1 29C/T and IL6 -572G/C Polymorphisms with Developmental Hip Dysplasia: A Case–Control Study in Adults with Severe Osteoarthritis. *International Orthopaedics* [en ligne]. Vol. 39, no. 4, p. 793–798 [visité le 2024-01-31]. ISSN 1432-5195. Disp. à l'adr. DOI: 10.1007/s00264-015-2675-0.

- CHARETTE, B. et al., 2001. Palpation and Dorsal Acetabular Rim Radiographic View for Early Detection of Canine Hip Dysplasia. *Veterinary and Comparative Orthopaedics and Traumatology* [en ligne]. Vol. 14, no. 03, p. 125–132 [visité le 2023-12-11]. ISSN 0932-0814, ISSN 2567-6911. Disp. à l'adr. DOI: 10.1055/s-0038-1632686.
- CHASE, K. et al., 2002. Genetic Basis for Systems of Skeletal Quantitative Traits: Principal Component Analysis of the Canid Skeleton. *Proceedings of the National Academy of Sciences of the United States of America* [en ligne]. Vol. 99, no. 15, p. 9930–9935 [visité le 2022-11-03]. ISSN 0027-8424. Disp. à l'adr. DOI: 10.1073/pnas.152333099.
- CHASE, K. et al., 2005. Genetic Regulation of Osteoarthritis: A QTL Regulating Cranial and Caudal Acetabular Osteophyte Formation in the Hip Joint of the Dog (*Canis Familiaris*). *American journal of medical genetics. Part A*. Vol. 135, no. 3, p. 334.
- CHEN, Y. et al., 2019a. Abnormal Expression of Pappa2 Gene May Indirectly Affect Mouse Hip Development through the IGF Signaling Pathway. *Endocrine* [en ligne]. Vol. 65, no. 2, p. 440–450 [visité le 2024-02-01]. ISSN 1559-0100. Disp. à l'adr. DOI: 10.1007/s12020-019-01975-0.
- CHEN, Y. et al., 2019b. Expression of PAPP-A2 and IGF Pathway-Related Proteins in the Hip Joint of Normal Rat and Those with Developmental Dysplasia of the Hip. *International Journal of Endocrinology*. Vol. 2019, p. 7691531. ISSN 1687-8337. Disp. à l'adr. DOI: 10.1155/2019/7691531.
- CHENG, B. et al., 2021. Integrative Analysis of MicroRNA and mRNA Sequencing Data Identifies Novel Candidate Genes and Pathways for Developmental Dysplasia of Hip. *CARTILAGE* [en ligne], p. 1947603521990859 [visité le 2021-09-02]. ISSN 1947-6035. Disp. à l'adr. DOI: 10.1177/1947603521990859.
- CILLIERS, H. J. & BEIGHTON, P., 1990. Beukes Familial Hip Dysplasia: An Autosomal Dominant Entity. *American journal of medical genetics* [en ligne]. Vol. 36, no. 4, p. 386–390 [visité le 2023-06-12]. ISSN 1096-8628. Disp. à l'adr. DOI: 10.1002/ajmg.1320360403.
- COLEMAN, S. S., 1968. Congenital Dysplasia of the Hip in the Navajo Infant. *Clinical Orthopaedics and Related Research®* [en ligne]. Vol. 56, p. 179–194 [visité le 2022-11-04]. ISSN 0009-921X.
- COMHAIRE, F. H. & SNAPS, F., 2008. Comparison of Two Canine Registry Databases on the Prevalence of Hip Dysplasia by Breed and the Relationship of Dysplasia with Body Weight and Height. *American Journal of Veterinary Research* [en ligne]. Vol. 69, no. 3, p. 330–333 [visité le 2021-11-24]. Disp. à l'adr. DOI: 10.2460/ajvr.69.3.330.
- COMHAIRE, F. H. et al., 2009. Precision, Reproducibility, and Clinical Usefulness of Measuring the Norberg Angle by Means of Computerized Image Analysis. *American Journal of Veterinary Research* [en ligne]. Vol. 70, no. 2, p. 228–235 [visité le 2023-12-11]. Disp. à l'adr. DOI: 10.2460/ajvr.70.2.228.
- COMMISSION, F. S. & COMMITTEE, F. G., 2022. FCI Hip Dysplasia Classification [en ligne]. FCI/DKK Hip Dysplasia Conference in Copenhagen [visité le 2025-08-20].
- CORLEY, E. A. et al., 1997. Reliability of Early Radiographic Evaluations for Canine Hip Dysplasia Obtained from the Standard Ventrodorsal Radiographic Projection. *Journal of the American Veterinary Medical Association*. Vol. 211, no. 9, p. 1142–1146. ISSN 0003-1488. Disp. à l'adr. PMID: 9364228.
- CORRAL, C., 2018. Canine Hip Dysplasia: Aetiology and Treatment. *The Veterinary Nurse* [en ligne]. Vol. 9, no. 5, p. 246–250 [visité le 2020-07-23]. ISSN 2044-0065. Disp. à l'adr. DOI: 10.12968/vetrn.2018.9.5.246.

- CULP, W. T. N. et al., 2006. Evaluation of the Norberg Angle Threshold: A Comparison of Norberg Angle and Distraction Index as Measures of Coxofemoral Degenerative Joint Disease Susceptibility in Seven Breeds of Dogs. *Veterinary Surgery* [en ligne]. Vol. 35, no. 5, p. 453–459 [visité le 2020-08-05]. ISSN 1532-950X. Disp. à l'adr. DOI: 10.1111/j.1532-950X.2006.00174.x.
- CVJETICANIN, S. & MARINKOVIC, D., 2005. Genetic Variability in the Group of Patients with Congenital Hip Dislocation. *Genetika*. Vol. 41, no. 8, p. 1142–1146. ISSN 0016-6758. Disp. à l'adr. PMID: 16161637.
- CZEIZEL, A. et al., 1974. Incidence of Congenital Dislocation of the Hip in Hungary. *British Journal of Preventive & Social Medicine* [en ligne]. Vol. 28, no. 4, p. 265 [visité le 2023-12-07]. Disp. à l'adr. DOI: 10.1136/jech.28.4.265.
- DAI, J. et al., 2008. Association of a Single Nucleotide Polymorphism in Growth Differentiate Factor 5 with Congenital Dysplasia of the Hip: A Case-Control Study. *Arthritis Research & Therapy* [en ligne]. Vol. 10, no. 5, R126 [visité le 2022-11-04]. ISSN 1478-6354. Disp. à l'adr. DOI: 10.1186/ar2540.
- DEMBIC, M. et al., 2022. Whole Exome Sequencing of 28 Families of Danish Descent Reveals Novel Candidate Genes and Pathways in Developmental Dysplasia of the Hip. *Molecular Genetics and Genomics* [en ligne] [visité le 2022-12-06]. ISSN 1617-4623. Disp. à l'adr. DOI: 10.1007/s00438-022-01980-5.
- DEMIREL, E. et al., 2023. Investigation of Igf-1, Igf-Bp3 and Igf-Bp5 Levels in Umbilical Cord Blood of Infants with Developmental Dysplasia of the Hip. *Turkish Journal of Medical Sciences*. Vol. 53, no. 3, p. 659–665. ISSN 1303-6165. Disp. à l'adr. DOI: 10.55730/1300-0144.5628.
- DEZATEUX, C. & ROSENDAHL, K., 2007. Developmental Dysplasia of the Hip. *The Lancet* [en ligne]. Vol. 369, no. 9572, p. 1541–1552 [visité le 2020-08-06]. ISSN 0140-6736. Disp. à l'adr. DOI: 10.1016/S0140-6736(07)60710-7.
- DING, R. et al., 2021. Downregulation of miR-1-3p Expression Inhibits the Hypertrophy and Mineralization of Chondrocytes in DDH. *Journal of Orthopaedic Surgery and Research* [en ligne]. Vol. 16, no. 1, p. 512 [visité le 2021-09-02]. ISSN 1749-799X. Disp. à l'adr. DOI: 10.1186/s13018-021-02666-1.
- DURINCK, S. et al., 2009. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor Package biomaRt. *Nature Protocols*. Vol. 4, no. 8, p. 1184–1191. ISSN 1750-2799. Disp. à l'adr. DOI: 10.1038/nprot.2009.97.
- EGLI, R. J. et al., 2009. Functional Analysis of the Osteoarthritis Susceptibility–Associated GDF5 Regulatory Polymorphism. *Arthritis & Rheumatism* [en ligne]. Vol. 60, no. 7, p. 2055–2064 [visité le 2024-02-01]. ISSN 1529-0131. Disp. à l'adr. DOI: 10.1002/art.24616.
- EITEL, F. et al., 1981. Bone Regeneration in Animals and in Man. *Archives of orthopaedic and traumatic surgery* [en ligne]. Vol. 99, no. 1, p. 59–64 [visité le 2023-12-06]. ISSN 1434-3916. Disp. à l'adr. DOI: 10.1007/BF00400911.
- EVANGELOU, E. et al., 2009. Large-Scale Analysis of Association between GDF5 and FRZB Variants and Osteoarthritis of the Hip, Knee, and Hand. *Arthritis & Rheumatism* [en ligne]. Vol. 60, no. 6, p. 1710–1721 [visité le 2024-01-26]. ISSN 1529-0131. Disp. à l'adr. DOI: 10.1002/art.24524.
- EYTAN, O. et al., 2014. A Novel Splice-site Mutation in the AAGAB Gene Segregates with Hereditary Punctate Palmoplantar Keratoderma and Congenital Dysplasia of the Hip in a Large Family. *Clinical and Experimental Dermatology* [en ligne]. Vol. 39, no. 2, p. 182–186 [visité le 2024-02-02]. ISSN 0307-6938. Disp. à l'adr. DOI: 10.1111/ced.12213.

- FEALEY, M. J. et al., 2017. Genetic Mapping of Principal Components of Canine Pelvic Morphology. *Canine genetics and epidemiology*. Vol. 4, no. 1, p. 4.
- FELDMAN, G. et al., 2009. The Otto Aufranc Award: Identification of a 4 Mb Region on Chromosome 17q21 Linked to Developmental Dysplasia of the Hip in One 18-Member, Multigeneration Family. *Clinical Orthopaedics and Related Research®* [en ligne]. Vol. 468, no. 2, p. 337–344 [visité le 2022-11-04]. ISSN 0009-921X. Disp. à l'adr. DOI: 10.1007/s11999-009-1073-6.
- FELDMAN, G. et al., 2017. A Murine Model for Developmental Dysplasia of the Hip: Ablation of CX3CR1 Affects Acetabular Morphology and Gait. *Journal of Translational Medicine* [en ligne]. Vol. 15, no. 1, p. 233 [visité le 2021-01-07]. ISSN 1479-5876. Disp. à l'adr. DOI: 10.1186/s12967-017-1335-0.
- FELDMAN, G. et al., 2019. Novel Mutation in Teneurin 3 Found to Co-Segregate in All Affecteds in a Multi-Generation Family with Developmental Dysplasia of the Hip. *Journal of Orthopaedic Research* [en ligne]. Vol. 37, no. 1, p. 171–180 [visité le 2024-01-25]. ISSN 1554-527X. Disp. à l'adr. DOI: 10.1002/jor.24148.
- FELDMAN, G. J. et al., 2012. Variable Expression and Incomplete Penetrance of Developmental Dysplasia of the Hip: Clinical Challenge in a 71-Member Multigeneration Family. *The Journal of Arthroplasty* [en ligne]. Vol. 27, no. 4, p. 527–532 [visité le 2022-11-04]. ISSN 0883-5403. Disp. à l'adr. DOI: 10.1016/j.arth.2011.10.016.
- FELDMAN, G. J. et al., 2013. Developmental Dysplasia of the Hip: Linkage Mapping and Whole Exome Sequencing Identify a Shared Variant in CX3CR1 in All Affected Members of a Large Multigeneration Family. *Journal of Bone and Mineral Research* [en ligne]. Vol. 28, no. 12, p. 2540–2549 [visité le 2021-01-07]. ISSN 1523-4681. Disp. à l'adr. DOI: 10.1002/jbmr.1999.
- FELDMAN, G. J. et al., 2014. Linkage Mapping and Whole Exome Sequencing Identify a Shared Variant in CX3CR1 in a Large Multi-Generation Family. *The Journal of Arthroplasty* [en ligne]. Vol. 29, no. 9, Supplement, p. 238–241 [visité le 2022-11-04]. ISSN 0883-5403. Disp. à l'adr. DOI: 10.1016/j.arth.2014.05.014.
- FELS, L. & DISTL, O., 2014. Identification and Validation of Quantitative Trait Loci (QTL) for Canine Hip Dysplasia (CHD) in German Shepherd Dogs. *PLOS ONE* [en ligne]. Vol. 9, no. 5, e96618 [visité le 2020-05-13]. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0096618.
- FELS, L. et al., 2014. Multiple Loci Associated with Canine Hip Dysplasia (CHD) in German Shepherd Dogs. *Mammalian Genome*. Vol. 25, no. 5–6, p. 262–269.
- FISCHER, A. et al., 2010. Static and Dynamic Ultrasonography for the Early Diagnosis of Canine Hip Dysplasia. *The Journal of Small Animal Practice*. Vol. 51, no. 11, p. 582–588. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.2010.00995.x.
- FLÜCKIGER, M. A., 2007a. How to Take and Read Hip Joint Radiographs in a Structured Way. *Eur J Companion Anim Pract*. Vol. 17, no. 02, p. 133–134.
- FLÜCKIGER, M. A., 2007b. Scoring Radiographs for Canine Hip Dysplasia-The Big Three Organisations in the World. *European Journal of Companion Animal Practice*. Vol. 17, no. 2, p. 135–140.
- FOUNDATION OF THE UNIVERSITY OF VETERINARY MEDECINE HANNOVER, 2009. Analysis for the Genetic Disposition for Hip Dysplasia in Canidae [en ligne]. [visité le 2020-09-28]. 2123777A1.

- FRIEDENBERG, S. G. et al., 2011. Evaluation of a Fibrillin 2 Gene Haplotype Associated with Hip Dysplasia and Incipient Osteoarthritis in Dogs. *American journal of veterinary research*. Vol. 72, no. 4, p. 530–540.
- FRIES, C. L. & REMEDIOS, A. M., 1995. The Pathogenesis and Diagnosis of Canine Hip Dysplasia: A Review. *The Canadian Veterinary Journal*. Vol. 36, no. 8, p. 494.
- FROST, H. M., 1989. Pathogenesis of Congenital Hip Dysplasia (CDH). A Proposal. *Veterinary and Comparative Orthopaedics and Traumatology* [en ligne]. Vol. 02, no. 01, p. 1–10 [visité le 2023-12-08]. ISSN 0932-0814, ISSN 2567-6911. Disp. à l'adr. DOI: 10.1055/s-0038-1633181.
- FUJITA, Y. et al., 2005. Direct and Indirect Markers of Cartilage Metabolism in Synovial Fluid Obtained from Dogs with Hip Dysplasia and Correlation with Clinical and Radiographic Variables. *American journal of veterinary research*. Vol. 66, no. 12, p. 2028–2033.
- GANZ, R. et al., 2008. The Etiology of Osteoarthritis of the Hip. *Clinical Orthopaedics and Related Research* [en ligne]. Vol. 466, no. 2, p. 264–272 [visité le 2023-12-10]. ISSN 0009-921X. Disp. à l'adr. DOI: 10.1007/s11999-007-0060-z.
- GASPAR, A. R. et al., 2016. The Norberg Angle Is Not an Accurate Predictor of Canine Hip Conformation Based on the Distraction Index and the Dorsolateral Subluxation Score. *Preventive Veterinary Medicine* [en ligne]. Vol. 135, p. 47–52 [visité le 2020-05-25]. ISSN 0167-5877. Disp. à l'adr. DOI: 10.1016/j.prevetmed.2016.10.020.
- GENEVOIS, J.-P. et al., 2007. Canine Hip Dysplasia Radiographic Screening. Prevalence of Rotation of the Pelvis along Its Length Axis in 7,012 Conventional Hip Extended Radiographs. *Veterinary and Comparative Orthopaedics and Traumatology*. Vol. 20, no. 04, p. 296–298.
- GENEVOIS, J.-P. et al., 2008. Prevalence of Hip Dysplasia According to Official Radiographic Screening, among 31 Breeds of Dogs in France. *Veterinary and Comparative Orthopaedics and Traumatology: V.C.O.T.* Vol. 21, no. 1, p. 21–24. ISSN 0932-0814. Disp. à l'adr. DOI: 10.3415/VCOT-07-02-0014.
- GHASSEMINIA, S. et al., 2022. Automated Diagnosis of Hip Dysplasia from 3D Ultrasound Using Artificial Intelligence: A Two-Center Multi-Year Study. *Informatics in Medicine Unlocked* [en ligne], p. 101082 [visité le 2022-09-13]. ISSN 2352-9148. Disp. à l'adr. DOI: 10.1016/j.imu.2022.101082.
- GHOSH, S. et al., 2009. Fibrillin 1 Gene with R2726W Mutation Is Absent in Patients with Primary Protrusio Acetabuli and Developmental Dysplasia of the Hip. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*. Vol. 15, no. 5, CR199–202. ISSN 1643-3750. Disp. à l'adr. PMID: 19396033.
- GINJA, M. et al., 2008a. Early Hip Laxity Examination in Predicting Moderate and Severe Hip Dysplasia in Estrela Mountain Dog. *Journal of Small Animal Practice* [en ligne]. Vol. 49, no. 12, p. 641–646 [visité le 2022-11-04]. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.2008.00629.x.
- GINJA, M. et al., 2008b. Passive Hip Laxity in Estrela Mountain Dog — Distraction Index, Heritability and Breeding Values. *Acta Veterinaria Hungarica* [en ligne]. Vol. 56, no. 3, p. 303–312 [visité le 2022-11-04]. ISSN 1588-2705, ISSN 0236-6290. Disp. à l'adr. DOI: 10.1556/avet.56.2008.3.4.
- GINJA, M. et al., 2009a. Comparison of Clinical, Radiographic, Computed Tomographic, and Magnetic Resonance Imaging Methods for Early Prediction of Canine Hip Laxity and Dysplasia. *Veterinary Radiology & Ultrasound* [en ligne]. Vol. 50, no. 2, p. 135–143 [visité le 2022-11-04]. ISSN 1740-8261. Disp. à l'adr. DOI: 10.1111/j.1740-8261.2009.01506.x.

- GINJA, M. et al., 2009b. Hip Dysplasia in Estrela Mountain Dogs: Prevalence and Genetic Trends 1991–2005. *The Veterinary Journal* [en ligne]. Vol. 182, no. 2, p. 275–282 [visité le 2022-11-04]. ISSN 1090-0233. Disp. à l'adr. DOI: 10.1016/j.tvjl.2008.06.014.
- GINJA, M. et al., 2010. Diagnosis, Genetic Control and Preventive Management of Canine Hip Dysplasia: A Review. *Veterinary Journal* (London, England: 1997). Vol. 184, no. 3, p. 269–276. ISSN 1532-2971. Disp. à l'adr. DOI: 10.1016/j.tvjl.2009.04.009.
- GOKHARMAN, F. D. et al., 2019. Optimizing the Time for Developmental Dysplasia of the Hip Screening: Earlier or Later? *Ultrasound Quarterly*. Vol. 35, no. 2, p. 130–135. ISSN 1536-0253. Disp. à l'adr. DOI: 10.1097/RUQ.0000000000000348.
- GOLDSMITH, L. T. et al., 1994. Transmission of Relaxin from Lactating Bitches to Their Offspring via Suckling. *Biology of Reproduction*. Vol. 50, no. 2, p. 258–265.
- GOMBOSH, M. et al., 2024. Developmental Dysplasia of the Hip Caused by Homozygous TRIM33 Pathogenic Variant Affecting Downstream BMP Pathway. *Journal of Medical Genetics* [en ligne] [visité le 2024-08-19]. ISSN 0022-2593, ISSN 1468-6244. Disp. à l'adr. DOI: 10.1136/jmg-2024-109928.
- GONG, J. K. et al., 1964. Composition of Trabecular and Cortical Bone. *The Anatomical Record*. Vol. 149, p. 325–331. ISSN 0003-276X. Disp. à l'adr. DOI: 10.1002/ar.1091490303.
- GRAF, R., 1984. Fundamentals of Sonographic Diagnosis of Infant Hip Dysplasia. *Journal of Pediatric Orthopedics*. Vol. 4, no. 6, p. 735–740. ISSN 0271-6798. Disp. à l'adr. DOI: 10.1097/01241398-198411000-00015.
- GRANCHI, D. et al., 2002. Association of Two Gene Polymorphisms With Osteoarthritis Secondary to Hip Dysplasia. *Clinical Orthopaedics and Related Research®* [en ligne]. Vol. 403, p. 108–117 [visité le 2022-11-04]. ISSN 0009-921X.
- GRAPH, R. et al., 2006. Hip Sonography: Diagnosis and Management of Infant Hip Dysplasia [en ligne]. Springer Berlin Heidelberg [visité le 2024-01-04]. ISBN 978-3-540-30957-4. Disp. à l'adr. DOI: 10.1007/3-540-30958-6.
- GREENE, L. M. et al., 2013. Associations among Exercise Duration, Lameness Severity, and Hip Joint Range of Motion in Labrador Retrievers with Hip Dysplasia. *Journal of the American Veterinary Medical Association*. Vol. 242, no. 11, p. 1528–1533.
- GRIFFITHS-JONES, S. et al., 2006. miRBase: microRNA Sequences, Targets and Gene Nomenclature. *Nucleic Acids Research*. Vol. 34, p. D140–144. ISSN 1362-4962. Disp. à l'adr. DOI: 10.1093/nar/gkj112.
- GUEVAR, M. & SNAPS, F., 2008. La méthode PennHIP – un moyen de dépistage précoce de la dysplasie de la hanche dans l'espèce canine. *Pratique Médicale et Chirurgicale de l'Animal de Compagnie* [en ligne]. Vol. 43, no. 2, p. 55–62 [visité le 2020-08-04]. ISSN 0758-1882. Disp. à l'adr. DOI: 10.1016/j.anicom.2008.04.001.
- GUMUS, E. et al., 2021. The Association Between BMP-2, UQCC1 and CX3CR1 Polymorphisms and the Risk of Developmental Dysplasia of the Hip. *Indian Journal of Orthopaedics* [en ligne]. Vol. 55, no. 1, p. 169 [visité le 2021-10-25]. ISSN 0019-5413. Disp. à l'adr. DOI: 10.1007/s43465-020-00235-y.
- GUMUS, E. et al., 2022. The Link between miR-96 Levels and the Developmental Dysplasia of the Hip. *GMJ*. Vol. 33, p. 249–252.

- GUO, G. et al., 2011. Canine Hip Dysplasia Is Predictable by Genotyping. *Osteoarthritis and Cartilage*. Vol. 19, no. 4, p. 420–429. ISSN 1522-9653. Disp. à l'adr. DOI: 10.1016/j.joca.2010.12.011.
- GUSTAFSSON, P.-O., 1975. Estradiol Induced Skeletal Changes: The Long Term Effect of Prenatal and Postnatal Administration in Beagles. *Acta Radiologica. Diagnosis* [en ligne]. Vol. 16, no. 344_suppl, p. 53–67 [visité le 2020-08-06]. ISSN 0567-8056. Disp. à l'adr. DOI: 10.1177/0284185175016S34406.
- HALL, B. K., 2005. Long Bone Growth: A Case of Crying Wolff. In: *Bones and Cartilage: Developmental and Evolutionary Skeletal Biology*. Elsevier Academic Press San Diego, p. 407–415.
- HAO, Z. et al., 2014. Association of a Single Nucleotide Polymorphism in HOXB9 with Developmental Dysplasia of the Hip: A Case-Control Study. *Journal of Orthopaedic Research* [en ligne]. Vol. 32, no. 2, p. 179–182 [visité le 2022-11-10]. ISSN 1554-527X. Disp. à l'adr. DOI: 10.1002/jor.22507.
- HARCKE, H. T. & GRISSOM, L. E., 1990. Performing Dynamic Sonography of the Infant Hip. *AJR. American journal of roentgenology*. Vol. 155, no. 4, p. 837–844. ISSN 0361-803X. Disp. à l'adr. DOI: 10.2214/ajr.155.4.2119119.
- HARSANYI, S. et al., 2021a. Association Analysis of GDF5 and Contributing Factors in Developmental Dysplasia of the Hip in Infants. *Ortopedia, Traumatologia, Rehabilitacja*. Vol. 23, no. 5, p. 335–339. ISSN 2084-4336. Disp. à l'adr. DOI: 10.5604/01.3001.0015.4348.
- HARSANYI, S. et al., 2021b. Genetic Study of IL6, GDF5 and PAPP2 in Association with Developmental Dysplasia of the Hip. *Genes* [en ligne]. Vol. 12, no. 7, p. 986 [visité le 2021-09-02]. Disp. à l'adr. DOI: 10.3390/genes12070986.
- HASHEM, G. et al., 2006. Relaxin and β -Estradiol Modulate Targeted Matrix Degradation in Specific Synovial Joint Fibrocartilages: Progesterone Prevents Matrix Loss. *Arthritis research & therapy*. Vol. 8, no. 4, R98.
- HATZIKOTULAS, K. et al., 2018. Genome-Wide Association Study of Developmental Dysplasia of the Hip Identifies an Association with GDF5. *Communications Biology* [en ligne]. Vol. 1, no. 1, p. 1–11 [visité le 2021-01-07]. ISSN 2399-3642. Disp. à l'adr. DOI: 10.1038/s42003-018-0052-4.
- HAYWARD, J. J. et al., 2016. Complex Disease and Phenotype Mapping in the Domestic Dog. *Nature Communications* [en ligne]. Vol. 7, no. 1, p. 1–11 [visité le 2020-05-13]. ISSN 2041-1723. Disp. à l'adr. DOI: 10.1038/ncomms10460.
- HEDHAMMAR, A. et al., 1979. Canine Hip Dysplasia: Study of Heritability in 401 Litters of German Shepherd Dogs. *Journal of the American Veterinary Medical Association*. Vol. 174, no. 9, p. 1012–1016. ISSN 0003-1488. Disp. à l'adr. PMID: 570968.
- HENRIGSON, B. et al., 1966. On the Etiology and Pathogenesis of Hip Dysplasia: A Comparative Review. *Journal of Small Animal Practice* [en ligne]. Vol. 7, no. 11, p. 673–688 [visité le 2020-08-05]. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.1966.tb04393.x.
- HERR, S. A. et al., 2021. Evidence of Acrolein in Synovial Fluid of Dogs with Osteoarthritis as a Potential Inflammatory Biomarker. *BMC Musculoskeletal Disorders* [en ligne]. Vol. 22, no. 1, p. 894 [visité le 2021-10-25]. ISSN 1471-2474. Disp. à l'adr. DOI: 10.1186/s12891-021-04762-z.
- HINTON, R. et al., 2002. Osteoarthritis: Diagnosis and Therapeutic Considerations. *American Family Physician*. Vol. 65, no. 5, p. 841–848. ISSN 0002-838X. Disp. à l'adr. PMID: 11898956.

- HOAGLUND, F. T. et al., 1981. Congenital Hip Dislocation and Dysplasia in Southern Chinese. *International Orthopaedics*. Vol. 4, no. 4, p. 243–246. ISSN 0341-2695. Disp. à l'adr. DOI: 10.1007/BFO0266064.
- HUANG, M. et al., 2017. A Novel Iterative Mixed Model to Remap Three Complex Orthopedic Traits in Dogs. *PloS one*. Vol. 12, no. 6, e0176932.
- İĞREK, S. et al., 2021. The Association of Interleukin-6 (IL-6) -572G/C and Transforming Growth Factor Beta 1 (TGFB1) 29C/T Single Nucleotide Polymorphisms (SNPs) with Developmental Dysplasia of the Hip: A Case Control Study. *ACHOT* [en ligne]. Vol. 88, no. 5, p. 339–343 [visité le 2025-01-24]. ISSN 00015415, ISSN 2570981X. Disp. à l'adr. DOI: 10.55095/achot2021/050.
- INGVARSSON, T. et al., 2001. A Large Icelandic Family with Early Osteoarthritis of the Hip Associated with a Susceptibility Locus on Chromosome 16p. *Arthritis and Rheumatism*. Vol. 44, no. 11, p. 2548–2555. ISSN 0004-3591. Disp. à l'adr. DOI: 10.1002/1529-0131(200111)44:11<2548::aid-art435>3.0.co;2-s.
- JACOBSEN, K. K. et al., 2023. COL11A1 Is Associated with Developmental Dysplasia of the Hip and Secondary Osteoarthritis in the HUNT Study. *Osteoarthritis and Cartilage Open* [en ligne], p. 100424 [visité le 2024-01-08]. ISSN 2665-9131. Disp. à l'adr. DOI: 10.1016/j.ocarto.2023.100424.
- JACOBSEN, S. & SONNE-HOLM, S., 2005. Hip Dysplasia: A Significant Risk Factor for the Development of Hip Osteoarthritis. A Cross-Sectional Survey. *Rheumatology* [en ligne]. Vol. 44, no. 2, p. 211–218 [visité le 2022-11-04]. ISSN 1462-0324. Disp. à l'adr. DOI: 10.1093/rheumatology/keh436.
- JACOBSEN, S. et al., 2009. Hip Dysplasia and Osteoarthritis. *Acta Orthopaedica* [en ligne] [visité le 2022-11-04]. Disp. à l'adr. DOI: 10.1080/00016470510030517.
- JAWADI, A. H. et al., 2018. Association Analysis between Four Vitamin D Receptor Gene Polymorphisms and Developmental Dysplasia of the Hip. *Journal of Genetics* [en ligne]. Vol. 97, no. 4, p. 925–930 [visité le 2024-02-01]. ISSN 0973-7731. Disp. à l'adr. DOI: 10.1007/s12041-018-0984-y.
- JL, X. et al., 2020. WISP-2, an Upregulated Gene in Hip Cartilage from the DDH Model Rats, Induces Chondrocyte Apoptosis through PPAR γ in Vitro. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*. Vol. 34, no. 4, p. 4904–4917. ISSN 1530-6860. Disp. à l'adr. DOI: 10.1096/fj.201901915R.
- JIA, J. et al., 2012. Association of a Single Nucleotide Polymorphism in Pregnancy-Associated Plasma Protein-A2 with Developmental Dysplasia of the Hip: A Case-Control Study. *Osteoarthritis and Cartilage* [en ligne]. Vol. 20, no. 1, p. 60–63 [visité le 2022-11-04]. ISSN 1063-4584. Disp. à l'adr. DOI: 10.1016/j.joca.2011.10.004.
- JIANG, J. et al., 2003. [Transmission Disequilibrium Test for Congenital Dislocation of the Hip and HOXB9 Gene or COL1A1 Gene]. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics*. Vol. 20, no. 3, p. 193–195. ISSN 1003-9406. Disp. à l'adr. PMID: 12778441.
- JIANG, J. et al., 2005. [Association Analysis on the Polymorphisms of PCOL2 and Sp1 Binding Sites of COL1A1 Gene and the Congenital Dislocation of the Hip in Chinese Population]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi = Zhonghua Yixue Yichuanxue Zazhi = Chinese Journal of Medical Genetics*. Vol. 22, no. 3, p. 327–329. ISSN 1003-9406. Disp. à l'adr. PMID: 15952128.

- JIANG, L. et al., 2021. Genomic Prediction of Two Complex Orthopedic Traits Across Multiple Pure and Mixed Breed Dogs. *Frontiers in Genetics* [en ligne]. Vol. 12, p. 1830 [visité le 2021-12-09]. ISSN 1664-8021. Disp. à l'adr. DOI: 10.3389/fgene.2021.666740.
- KANG, J. M. et al., 2020. Genome-Wide Association Study to Identify Canine Hip Dysplasia Loci in Dogs. *Journal of Animal Science and Technology* [en ligne]. Vol. 62, no. 3, p. 306–312 [visité le 2024-01-25]. ISSN 2055-0391, ISSN 2672-0191. Disp. à l'adr. DOI: 10.5187/jast.2020.62.3.306.
- KANG, M. S. et al., 2019. Clinical Significance of Asymmetric Skin Folds in the Medial Thigh for the Infantile Screening of Developmental Dysplasia of the Hip. *Pediatrics & Neonatology* [en ligne]. Vol. 60, no. 5, p. 570–576 [visité le 2023-12-11]. ISSN 1875-9572. Disp. à l'adr. DOI: 10.1016/j.pedneo.2019.02.004.
- KAPOOR, B. et al., 2007. Vitamin D and Oestrogen Receptor Polymorphisms in Developmental Dysplasia of the Hip and Primary Protrusio Acetabuli – A Preliminary Study. *Journal of Negative Results in BioMedicine* [en ligne]. Vol. 6, no. 1, p. 7 [visité le 2022-11-04]. ISSN 1477-5751. Disp. à l'adr. DOI: 10.1186/1477-5751-6-7.
- KASSTRÖM, H., 2016. Nutrition, Weight Gain and Development of Hip Dysplasia: An Experimental Investigation in Growing Dogs with Special Reference to the Effect of Feeding Intensity. *Acta Radiologica. Diagnosis* [en ligne] [visité le 2020-07-23]. Disp. à l'adr. DOI: 10.1177/0284185175016S34412.
- KASSTRÖM, H. et al., 1975. Plasma Levels of Estradiol and Plasma Protein Binding of Sex Steroids in Dogs: An Investigation with Special Reference to Development of Hip Dysplasia in Growing Individuals. *Acta Radiologica. Diagnosis* [en ligne]. Vol. 16, no. 344_suppl, p. 121–133 [visité le 2020-08-06]. ISSN 0567-8056. Disp. à l'adr. DOI: 10.1177/0284185175016S34411.
- KAYES, K. & DIDELOT, W., 2012. Chapter 129 - Major Congenital Orthopedic Deformities. In: CORAN, A. G. (ed.). *Pediatric Surgery (Seventh Edition)* [en ligne]. Philadelphia: Mosby, p. 1699–1710 [visité le 2023-12-08]. ISBN 978-0-323-07255-7. Disp. à l'adr. DOI: 10.1016/B978-0-323-07255-7.00129-X.
- KEALY, R. D. et al., 1992. Effects of Limited Food Consumption on the Incidence of Hip Dysplasia in Growing Dogs. *JOURNAL-AMERICAN VETERINARY MEDICAL ASSOCIATION*. Vol. 201, p. 857–857.
- KEALY, R. D. et al., 1993. Effects of Dietary Electrolyte Balance on Subluxation of the Femoral Head in Growing Dogs. *American journal of veterinary research*. Vol. 54, no. 4, p. 555–562.
- Kenhub Hip Muscles, 2025 [en ligne]. Kenhub. [visité le 2025-09-18].
- KHAN, I. M. et al., 2007. The Development of Synovial Joints. In: *Current Topics in Developmental Biology* [en ligne]. Academic Press. Vol. 79, p. 1–36 [visité le 2020-08-07]. Disp. à l'adr. DOI: 10.1016/S0070-2153(06)79001-9.
- KIAPOUR, A. M. et al., 2018. The Role of Gdf5 Regulatory Regions in Development of Hip Morphology. *PLOS ONE* [en ligne]. Vol. 13, no. 11, e0202785 [visité le 2022-01-07]. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0202785.
- KIELER, I. N. et al., 2024. Genome Wide Association Study in Swedish Labrador Retrievers Identifies Genetic Loci Associated with Hip Dysplasia and Body Weight. *Scientific Reports* [en ligne]. Vol. 14, p. 6090 [visité le 2024-04-19]. ISSN 2045-2322. Disp. à l'adr. DOI: 10.1038/s41598-024-56060-y.

- KIMMEL, D. B. & JEE, W. S. S., 1982. A Quantitative Histologic Study of Bone Turnover in Young Adult Beagles. *The Anatomical Record* [en ligne]. Vol. 203, no. 1, p. 31–45 [visit  le 2023-12-06]. ISSN 1097-0185. Disp.   l'adr. DOI: 10.1002/ar.1092030104.
- KING, M. D., 2017. Etiopathogenesis of Canine Hip Dysplasia, Prevalence, and Genetics. *Veterinary Clinics of North America: Small Animal Practice* [en ligne]. Vol. 47, no. 4, p. 753–767 [visit  le 2022-11-04]. ISSN 0195-5616. Disp.   l'adr. DOI: 10.1016/j.cvsm.2017.03.001.
- KOLUNDŽIĆ, R. et al., 2011. Association of Interleukin-6 and Transforming Growth Factor- β 1 Gene Polymorphisms with Developmental Hip Dysplasia and Severe Adult Hip Osteoarthritis: A Preliminary Study. *Cytokine* [en ligne]. Vol. 54, no. 2, p. 125–128 [visit  le 2022-11-04]. ISSN 1043-4666. Disp.   l'adr. DOI: 10.1016/j.cyto.2011.02.004.
- KOREA GENETIC INFORMATION RESEARCH INSTITUTE CO., LTD., 2017. Composition of Multiplex Single Nucleotide Polymorphism Markers for Predicting or Diagnosing Canine Hip Dysplasia and Method for Prediction or Diagnosis Using the Same [en ligne]. [visit  le 2020-09-28]. KR101777161B1.
- KOTLARSKY, P. et al., 2015. Developmental Dysplasia of the Hip: What Has Changed in the Last 20 Years? *World Journal of Orthopedics* [en ligne]. Vol. 6, no. 11, p. 886–901 [visit  le 2023-12-08]. Disp.   l'adr. DOI: 10.5312/wjo.v6.i11.886.
- KRONTVEIT, R. I. et al., 2010. A Prospective Study on Canine Hip Dysplasia and Growth in a Cohort of Four Large Breeds in Norway (1998–2001). *Preventive veterinary medicine*. Vol. 97, no. 3–4, p. 252–263.
- KRONTVEIT, R. I. et al., 2012. Housing-and Exercise-Related Risk Factors Associated with the Development of Hip Dysplasia as Determined by Radiographic Evaluation in a Prospective Cohort of Newfoundlands, Labrador Retrievers, Leonbergers, and Irish Wolfhounds in Norway. *American journal of veterinary research*. Vol. 73, no. 6, p. 838–846.
- LAVRIJSEN, I. C. M. et al., 2014. Genome Wide Analysis Indicates Genes for Basement Membrane and Cartilage Matrix Proteins as Candidates for Hip Dysplasia in Labrador Retrievers. *PLoS ONE* [en ligne]. Vol. 9, no. 1, e87735 [visit  le 2020-05-13]. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0087735.
- LI, C. et al., 2021. Comprehensive Analysis of Pathological Changes in Hip Joint Capsule of Patients with Developmental Dysplasia of the Hip. *Bone & Joint Research*. Vol. 10, no. 9, p. 558–570.
- LI, L.-y. et al., 2010. [Gene Mapping of Developmental Dysplasia of the Hip in Chromosome 17q21 Region]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi = Zhonghua Yixue Yichuanxue Zazhi = Chinese Journal of Medical Genetics*. Vol. 27, no. 6, p. 620–625. ISSN 1003-9406. Disp.   l'adr. DOI: 10.3760/cmaj.issn.1003-9406.2010.06.004.
- LI, L. et al., 2017. CX3CR1 Polymorphisms Associated with an Increased Risk of Developmental Dysplasia of the Hip in Human. *Journal of Orthopaedic Research* [en ligne]. Vol. 35, no. 2, p. 377–380 [visit  le 2022-11-04]. ISSN 1554-527X. Disp.   l'adr. DOI: 10.1002/jor.23294.
- LINDBLAD-TOH, K. et al., 2005. Genome Sequence, Comparative Analysis and Haplotype Structure of the Domestic Dog. *Nature* [en ligne]. Vol. 438, no. 7069, p. 803–819 [visit  le 2020-09-01]. ISSN 1476-4687. Disp.   l'adr. DOI: 10.1038/nature04338.
- LIU, J. et al., 2024. Microstructure Changes and miRNA–mRNA Network in a Developmental Dysplasia of the Hip Rat Model. *iScience* [en ligne], p. 109449 [visit  le 2024-03-12]. ISSN 2589-0042. Disp.   l'adr. DOI: 10.1016/j.isci.2024.109449.

- LIU, J. et al., 2013. Rs143383 in the Growth Differentiation Factor 5 (GDF5) Gene Significantly Associated with Osteoarthritis (OA)-A Comprehensive Meta-analysis. *International Journal of Medical Sciences* [en ligne]. Vol. 10, no. 3, p. 312–319 [visité le 2024-02-01]. ISSN 1449-1907. Disp. à l'adr. DOI: 10.7150/ijms.5455.
- LIU, S. et al., 2014. Two Single-Nucleotide Polymorphisms in the DKK1 Gene Are Associated with Developmental Dysplasia of the Hip in the Chinese Han Female Population. *Genetic Testing and Molecular Biomarkers*. Vol. 18, no. 8, p. 557–561. ISSN 1945-0257. Disp. à l'adr. DOI: 10.1089/gtmb.2014.0044.
- LIU, T. et al., 2007. A Random Model for Mapping Imprinted Quantitative Trait Loci in a Structured Pedigree: An Implication for Mapping Canine Hip Dysplasia. *Genomics*. Vol. 90, no. 2, p. 276–284.
- LIU, X. et al., 2020. Chondrocyte Suppression Is Mediated by miR-129-5p via GDF11/SMAD3 Signaling in Developmental Dysplasia of the Hip. *Journal of Orthopaedic Research* [en ligne]. Vol. 38, no. 12, p. 2559–2572 [visité le 2024-02-06]. ISSN 1554-527X. Disp. à l'adr. DOI: 10.1002/jor.24713.
- LODER, R. T. & SHAFER, C., 2015. The Demographics of Developmental Hip Dysplasia in the Midwestern United States (Indiana). *Journal of Children's Orthopaedics* [en ligne]. Vol. 9, no. 1, p. 93–98 [visité le 2022-11-10]. ISSN 1863-2521. Disp. à l'adr. DOI: 10.1007/s11832-015-0636-1.
- LODER, R. T. & TODHUNTER, R. J., 2017. The Demographics of Canine Hip Dysplasia in the United States and Canada. *Journal of veterinary medicine*. Vol. 2017.
- LOPEZ, M. J. et al., 2008. Relationships among Measurements Obtained by Use of Computed Tomography and Radiography and Scores of Cartilage Microdamage in Hip Joints with Moderate to Severe Joint Laxity of Adult Dogs. *American Journal of Veterinary Research* [en ligne]. Vol. 69, no. 3, p. 362–370 [visité le 2025-09-23]. Disp. à l'adr. DOI: 10.2460/ajvr.69.3.362.
- LORENZ, A., 1920. Die Sogenannte Angeborene Hüftverrenkung : Ihre Pathologie Und Therapie [en ligne] [visité le 2023-12-07].
- LOUGHLIN, J. et al., 2006. The CALM1 Core Promoter Polymorphism Is Not Associated with Hip Osteoarthritis in a United Kingdom Caucasian Population. *Osteoarthritis and Cartilage* [en ligne]. Vol. 14, no. 3, p. 295–298 [visité le 2024-01-29]. ISSN 1063-4584. Disp. à l'adr. DOI: 10.1016/j.joca.2005.11.001.
- LU, Y.-L. et al., 2023. Identification of A Novel Variant of Filamin A Destroying the Attraction Between Benzene Rings and Sulfhydryl in Developmental Dysplasia of the Hip. *DNA and Cell Biology* [en ligne]. Vol. 42, no. 10, p. 638–644 [visité le 2023-10-26]. ISSN 1044-5498. Disp. à l'adr. DOI: 10.1089/dna.2023.0159.
- LUO, W. et al., 2021. 25(OH)D Status and Expression of miR-140 in the Serum of Patients with Developmental Dysplasia of the Hip. *Nutrition (Burbank, Los Angeles County, Calif.)* Vol. 81, p. 110896. ISSN 1873-1244. Disp. à l'adr. DOI: 10.1016/j.nut.2020.110896.
- LUST, G. et al., 1993. Joint Laxity and Its Association with Hip Dysplasia in Labrador Retrievers. *American journal of veterinary research*. Vol. 54, no. 12, p. 1990–1999. ISSN 1943-5681. Disp. à l'adr. PMID: 8116927.
- LUST, G. et al., 1973. Development of Hip Dysplasia in Dogs. *American Journal of Veterinary Research*. Vol. 34, no. 1, p. 87–91. ISSN 0002-9645. Disp. à l'adr. PMID: 4346059.

- MA, H. et al., 2005. Transmission Disequilibrium between Congenital Dislocation of the Hip and Homeobox-Containing Genes. Vol. 9, p. 190–191.
- MA, W. et al., 2017. Genetic Association Study of Common Variants in TGFB1 and IL-6 with Developmental Dysplasia of the Hip in Han Chinese Population. Scientific Reports [en ligne]. Vol. 7, no. 1, p. 10287 [visité le 2024-02-02]. ISSN 2045-2322. Disp. à l'adr. DOI: 10.1038/s41598-017-11185-1.
- MABUCHI, A. et al., 2006. Familial Osteoarthritis of the Hip Joint Associated with Acetabular Dysplasia Maps to Chromosome 13q. The American Journal of Human Genetics [en ligne]. Vol. 79, no. 1, p. 163–168 [visité le 2022-11-04]. ISSN 0002-9297. Disp. à l'adr. DOI: 10.1086/505088.
- MAITRE, P. et al., 2010. Description of the Type of Chemical Restraint Used by French Veterinarians to Perform Hip Dysplasia Screening Radiographs. Veterinary and Comparative Orthopaedics and Traumatology [en ligne]. Vol. 23, no. 4, p. 245–249 [visité le 2021-04-14]. ISSN 0932-0814, ISSN 2567-6911. Disp. à l'adr. DOI: 10.3415/VCOT-10-01-0004.
- MÄKI, K. et al., 2004. An Indication of Major Genes Affecting Hip and Elbow Dysplasia in Four Finnish Dog Populations. Heredity [en ligne]. Vol. 92, no. 5, p. 402–408 [visité le 2021-11-24]. ISSN 1365-2540. Disp. à l'adr. DOI: 10.1038/sj.hdy.6800434.
- MANZ, E. et al., 2017. Prospective Evaluation of a Patented DNA Test for Canine Hip Dysplasia (CHD). PLOS ONE [en ligne]. Vol. 12, no. 8, e0182093 [visité le 2020-05-13]. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0182093.
- MARS INC, 2018. Method for Detecting the Susceptibility of a Dog to Hip Dysplasia [en ligne]. [visité le 2020-09-28]. 3008206B1.
- MARSCHALL, Y. & DISTL, O., 2007. Mapping Quantitative Trait Loci for Canine Hip Dysplasia in German Shepherd Dogs. Mammalian Genome [en ligne]. Vol. 18, no. 12, p. 861–870 [visité le 2020-08-31]. ISSN 1432-1777. Disp. à l'adr. DOI: 10.1007/s00335-007-9071-z.
- MATEESCU, R. G. et al., 2005. Analysis of Allele Fidelity, Polymorphic Information Content, and Density of Microsatellites in a Genome-Wide Screening for Hip Dysplasia in a Crossbreed Pedigree. The Journal of heredity. Vol. 96, no. 7, p. 847–853. ISSN 0022-1503. Disp. à l'adr. DOI: 10.1093/jhered/esl09.
- MATEESCU, R. G. et al., 2008. Identification of Quantitative Trait Loci for Osteoarthritis of Hip Joints in Dogs. American Journal of Veterinary Research [en ligne]. Vol. 69, no. 10, p. 1294–1300 [visité le 2022-11-04]. Disp. à l'adr. DOI: 10.2460/ajvr.69.10.1294.
- MEESON, R. L. et al., 2019. Spontaneous Dog Osteoarthritis—a One Medicine Vision. Nature Reviews Rheumatology. Vol. 15, no. 5, p. 273–287.
- MENG, X.-h. et al., 2024. Two Genetic Variants in the HIBCH and FTCDNL1 Genes Are Associated with Susceptibility to Developmental Dysplasia of the Hips among the Han Chinese Population of Southwest China. Journal of Orthopaedic Surgery and Research [en ligne]. Vol. 19, no. 1, p. 464 [visité le 2024-08-19]. ISSN 1749-799X. Disp. à l'adr. DOI: 10.1186/s13018-024-04958-8.
- MIAO, Y. et al., 2025. Genetic Analysis of Familial Developmental Dysplasia of the Hip Associated With a Heterozygous Variant in the COMP Gene: A Case Report. Molecular Genetics & Genomic Medicine [en ligne]. Vol. 13, no. 9, e70135 [visité le 2025-10-05]. ISSN 2324-9269. Disp. à l'adr. DOI: 10.1002/mgg3.70135.

- MIKKOLA, L. et al., 2019a. Genetic Dissection of Canine Hip Dysplasia Phenotypes and Osteoarthritis Reveals Three Novel Loci. *BMC Genomics* [en ligne]. Vol. 20, no. 1, p. 1027 [visité le 2020-05-13]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/s12864-019-6422-6.
- MIKKOLA, L. et al., 2021. An Across-Breed Validation Study of 46 Genetic Markers in Canine Hip Dysplasia. *BMC Genomics* [en ligne]. Vol. 22, no. 1, p. 68 [visité le 2021-02-01]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/s12864-021-07375-x.
- MIKKOLA, L. I. et al., 2019b. Novel Protective and Risk Loci in Hip Dysplasia in German Shepherds. *PLOS Genetics* [en ligne]. Vol. 15, no. 7, e1008197 [visité le 2020-05-13]. ISSN 1553-7404. Disp. à l'adr. DOI: 10.1371/journal.pgen.1008197.
- MILLIS, D. & LEVINE, D., 2013. *Canine Rehabilitation and Physical Therapy*. Elsevier Health Sciences.
- MITCHELL, P. D. & REDFERN, R. C., 2007. The Prevalence of Dislocation in Developmental Dysplasia of the Hip in Britain over the Past Thousand Years. *Journal of Pediatric Orthopedics*. Vol. 27, no. 8, p. 890–892. ISSN 0271-6798. Disp. à l'adr. DOI: 10.1097/bpo.0b013e31815a6091.
- MIYAMOTO, Y. et al., 2007. A Functional Polymorphism in the 5' UTR of GDF5 Is Associated with Susceptibility to Osteoarthritis. *Nature Genetics* [en ligne]. Vol. 39, no. 4, p. 529–533 [visité le 2022-11-04]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/2005.
- MORGAN, S. J., 1992. The Pathology of Canine Hip Dysplasia. *Veterinary Clinics of North America: Small Animal Practice* [en ligne]. Vol. 22, no. 3, p. 541–550 [visité le 2023-12-08]. ISSN 0195-5616. Disp. à l'adr. DOI: 10.1016/S0195-5616(92)50054-7.
- MORI, Y. et al., 2023. Genome-Wide Association Study and Transcriptome of Japanese Patients with Developmental Dysplasia of the Hip Demonstrates an Association with the Ferroptosis Signaling Pathway. *International Journal of Molecular Sciences* [en ligne]. Vol. 24, no. 5, p. 5019 [visité le 2023-03-09]. ISSN 1422-0067. Disp. à l'adr. DOI: 10.3390/ijms24055019.
- MOSTAFA, A. A. et al., 2022. Modified FCI (Fédération Cynologique Internationale) Scoring of the Coxofemoral Joint in Labrador Retrievers Without and With Hip Dysplasia. *Frontiers in Veterinary Science*. Vol. 9.
- MOTOTANI, H. et al., 2005. A Functional Single Nucleotide Polymorphism in the Core Promoter Region of CALM1 Is Associated with Hip Osteoarthritis in Japanese. *Human molecular genetics* [en ligne]. Vol. 14, no. 8, p. 1009–1017 [visité le 2023-06-12]. ISSN 1460-2083. Disp. à l'adr. DOI: 10.1093/hmg/ddi093.
- MURPHY, S. B. et al., 1995. The Prognosis in Untreated Dysplasia of the Hip. A Study of Radiographic Factors That Predict the Outcome. *JBJS* [en ligne]. Vol. 77, no. 7, p. 985 [visité le 2023-12-14]. ISSN 0021-9355.
- MURRAY, R. O., 1965. The Aetiology of Primary Osteoarthritis of the Hip. *The British Journal of Radiology* [en ligne]. Vol. 38, no. 455, p. 810–824 [visité le 2023-12-14]. ISSN 0007-1285. Disp. à l'adr. DOI: 10.1259/0007-1285-38-455-810.
- MUSUMECI, G. et al., 2015. Osteoarthritis in the XXIst Century: Risk Factors and Behaviours That Influence Disease Onset and Progression. *International Journal of Molecular Sciences* [en ligne]. Vol. 16, no. 3, p. 6093–6112 [visité le 2023-12-10]. ISSN 1422-0067. Disp. à l'adr. DOI: 10.3390/ijm s16036093.

- NGANVONGPANIT, K. et al., 2008. Evaluation of Serum Chondroitin Sulfate and Hyaluronan: Biomarkers for Osteoarthritis in Canine Hip Dysplasia. *Journal of Veterinary Science*. Vol. 9, no. 3, p. 317–325.
- NING, B. et al., 2019. The H19/Let-7 Feedback Loop Contributes to Developmental Dysplasia and Dislocation of the Hip. *Physiological Research*. Vol. 68, no. 2, p. 275–284. ISSN 1802-9973. Disp. à l'adr. DOI: 10.33549/physiolres.933920.
- NOWLAN, N. C. et al., 2014. Immobilized Chicks as a Model System for Early-Onset Developmental Dysplasia of the Hip. *Journal of orthopaedic research*. Vol. 32, no. 6, p. 777–785.
- O'BRIEN, R. T. et al., 1997. Dynamic Ultrasonographic Measurement of Passive Coxofemoral Joint Laxity in Puppies. *Journal of the American Animal Hospital Association* [en ligne]. Vol. 33, no. 3, p. 275–281 [visit  le 2022-11-04]. ISSN 0587-2871. Disp. à l'adr. DOI: 10.5326/15473317-33-3-275.
- OFA, 2023. Breed Health Statistics [en ligne]. OFA, 2023-12-07. [visit  le 2023-12-07].
- OGDEN, D. M. et al., 2012. The S-Measurement in the Diagnosis of Canine Hip Dysplasia. *Veterinary Surgery* [en ligne]. Vol. 41, no. 1, p. 78–85 [visit  le 2022-11-10]. ISSN 1532-950X. Disp. à l'adr. DOI: 10.1111/j.1532-950X.2011.00874.x.
- OHLERTH, S. et al., 2003. Comparison of Three Distraction Methods and Conventional Radiography for Early Diagnosis of Canine Hip Dysplasia. *Journal of Small Animal Practice* [en ligne]. Vol. 44, no. 12, p. 524–529 [visit  le 2020-08-05]. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.2003.tb00114.x.
- OHLERTH, S. et al., 2001. Estimation of Genetic Population Variables for Six Radiographic Criteria of Hip Dysplasia in a Colony of Labrador Retrievers. *American Journal of Veterinary Research* [en ligne]. Vol. 62, no. 6, p. 846–852 [visit  le 2022-11-04]. Disp. à l'adr. DOI: 10.2460/ajvr.2001.62.846.
- ORTOLANI, M., 1948. La Lussazione Congenita Dell'anca: Nuovi Criteri Diagnostici e Profilattico-Correttivi. Cappelli.
- PACIFICI, M. et al., 2005. Mechanisms of Synovial Joint and Articular Cartilage Formation: Recent Advances, but Many Lingering Mysteries. *Birth Defects Research Part C: Embryo Today: Reviews* [en ligne]. Vol. 75, no. 3, p. 237–248 [visit  le 2020-08-07]. ISSN 1542-9768. Disp. à l'adr. DOI: 10.1002/bdrc.20050.
- PADULA, G. et al., 2023. Preliminary Transcriptomic Analysis of Peripheral Blood from German Shepherd Dogs with Degenerative Joint Disease for the Identification of Diagnostic Biomarkers. *Gene* [en ligne]. Vol. 872, p. 147455 [visit  le 2024-01-17]. ISSN 0378-1119. Disp. à l'adr. DOI: 10.1016/j.gene.2023.147455.
- PASCUAL-GARRIDO, C. et al., 2018. Canine Hip Dysplasia: A Natural Animal Model for Human Developmental Dysplasia of the Hip. *Journal of Orthopaedic Research* [en ligne]. Vol. 36, no. 7, p. 1807–1817 [visit  le 2022-10-05]. ISSN 1554-527X. Disp. à l'adr. DOI: 10.1002/jor.23828.
- PASTER, E. R. et al., 2005. Estimates of Prevalence of Hip Dysplasia in Golden Retrievers and Rottweilers and the Influence of Bias on Published Prevalence Figures. *Journal of the American Veterinary Medical Association* [en ligne]. Vol. 226, no. 3, p. 387–392 [visit  le 2022-11-04]. Disp. à l'adr. DOI: 10.2460/javma.2005.226.387.
- PEARCE, A. I. et al., 2007. Animal Models for Implant Biomaterial Research in Bone: A Review. *European Cells & Materials*. Vol. 13, p. 1–10. ISSN 1473-2262. Disp. à l'adr. DOI: 10.22203/ecm.v013a01.

- PFAHLER, S. & DISTL, O., 2012. Identification of Quantitative Trait Loci (QTL) for Canine Hip Dysplasia and Canine Elbow Dysplasia in Bernese Mountain Dogs. *PLOS ONE* [en ligne]. Vol. 7, no. 11, e49782 [visit  le 2020-05-13]. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0049782.
- PHAVAPHUTANON, J. et al., 2009. Evaluation of Quantitative Trait Loci for Hip Dysplasia in Labrador Retrievers. *American journal of veterinary research*. Vol. 70, no. 9, p. 1094–1101.
- Physiopedia Hip Muscles, 2025 [en ligne]. Physiopedia. [visit  le 2025-09-18].
- PIERCE, K. R. & BRIDGES, G. H., 1967. The Role of Estrogens in the Pathogenesis of Canine Hip Dysplasia. *Metabolism of Exogenous Estrogens. Journal of Small Animal Practice* [en ligne]. Vol. 8, no. 7, p. 383–389 [visit  le 2020-08-06]. ISSN 1748-5827. Disp.   l'adr. DOI: 10.1111/j.1748-5827.1967.tb04564.x.
- QIAO, L. et al., 2017a. A Novel Missense Variant in TXNDC3 Is Associated with Developmental Dysplasia of the Hip in Han Chinese Population. *International Journal of Clinical and Experimental Pathology* [en ligne]. Vol. 10, no. 10, p. 10483–10488 [visit  le 2024-02-01]. ISSN 1936-2625. Disp.   l'adr. PMID: 31966386.
- QIAO, L. et al., 2017b. Variants Rs3732378 and Rs3732379 of Gene CX3CR1 Are Not Associated with Developmental Dysplasia of the Hip in Han Chinese Population. *International Journal of Clinical and Experimental Pathology*. Vol. 10, p. 2095–2099.
- RAO, Y. et al., 2025. Dysregulated Mechanotransduction via CCNB2 Underlies Fibroblast Hypoplasia in Developmental Dysplasia of the Hip [en ligne]. 2025-05-07. [visit  le 2025-06-02]. ISSN 2693-5015. Disp.   l'adr. DOI: 10.21203/rs.3.rs-6367897/v1.
- REIJMAN, M. et al., 2005. Acetabular Dysplasia Predicts Incident Osteoarthritis of the Hip: The Rotterdam Study. *Arthritis & Rheumatism* [en ligne]. Vol. 52, no. 3, p. 787–793 [visit  le 2022-11-04]. ISSN 1529-0131. Disp.   l'adr. DOI: 10.1002/art.20886.
- REMEDIOS, A. M. & FRIES, C. L., 1995. Treatment of Canine Hip Dysplasia: A Review. *The Canadian Veterinary Journal* [en ligne]. Vol. 36, no. 8, p. 503–509 [visit  le 2020-08-04]. ISSN 0008-5286. Disp.   l'adr. PMID: 7585437.
- RICHARDSON, D. C., 1992. The Role of Nutrition in Canine Hip Dysplasia. *Veterinary Clinics of North America: Small Animal Practice*. Vol. 22, no. 3, p. 529–540.
- RISER, W. H. & SHIRER, J. F., 1966. Hip Dysplasia: Coxafemoral Abnormalities in Neonatal German Shepherd Dogs. *Journal of Small Animal Practice* [en ligne]. Vol. 7, no. 1, p. 7–12 [visit  le 2023-12-07]. ISSN 0022-4510, ISSN 1748-5827. Disp.   l'adr. DOI: 10.1111/j.1748-5827.1966.tb04372.x.
- ROBERTS, T. & MCGREEVY, P. D., 2010. Selection for Breed-Specific Long-Bodied Phenotypes Is Associated with Increased Expression of Canine Hip Dysplasia. *The Veterinary Journal*. Vol. 183, no. 3, p. 266–272.
- ROBLING, A. G. et al., 2006. Biomechanical and Molecular Regulation of Bone Remodeling. *Annu. Rev. Biomed. Eng.* Vol. 8, p. 455–498.
- ROBY, P. et al., 1999. Autosomal Dominant (Beukes) Premature Degenerative Osteoarthropathy of the Hip Joint Maps to an 11-cM Region on Chromosome 4q35. *The American Journal of Human Genetics* [en ligne]. Vol. 64, no. 3, p. 904–908 [visit  le 2024-02-02]. ISSN 0002-9297. Disp.   l'adr. DOI: 10.1086/302291.

- ROSENDAHL, K. et al., 1992. Ultrasound in the Early Diagnosis of Congenital Dislocation of the Hip: The Significance of Hip Stability versus Acetabular Morphology. *Pediatric Radiology*. Vol. 22, no. 6, p. 430–433. ISSN 0301-0449. Disp. à l'adr. DOI: 10.1007/BF02013504.
- ROUAULT, K. et al., 2010. Evidence of Association between GDF5 Polymorphisms and Congenital Dislocation of the Hip in a Caucasian Population. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. Vol. 18, p. 1144–9. Disp. à l'adr. DOI: 10.1016/j.joca.2010.05.018.
- ROUAULT, K. et al., 2009. Do HOXB9 and COL1A1 Genes Play a Role in Congenital Dislocation of the Hip? Study in a Caucasian Population. *Osteoarthritis and Cartilage* [en ligne]. Vol. 17, no. 8, p. 1099–1105 [visité le 2024-01-29]. ISSN 1063-4584. Disp. à l'adr. DOI: 10.1016/j.joca.2008.12.012.
- RUBINI, M. et al., 2008. Exclusion of COL2A1 and VDR as Developmental Dysplasia of the Hip Genes. *Clinical Orthopaedics and Related Research®* [en ligne]. Vol. 466, no. 4, p. 878–883 [visité le 2022-11-04]. ISSN 0009-921X. Disp. à l'adr. DOI: 10.1007/s11999-008-0120-z.
- RUNGE, J. J. et al., 2010. Distraction Index as a Risk Factor for Osteoarthritis Associated with Hip Dysplasia in Four Large Dog Breeds*. *Journal of Small Animal Practice* [en ligne]. Vol. 51, no. 5, p. 264–269 [visité le 2022-11-04]. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.2010.00937.x.
- SADAT-ALI, M. et al., 2018. Genetic Influence in Developmental Dysplasia of the Hip in Saudi Arabian Children Due to GDF5 Polymorphism. *Biochemical Genetics* [en ligne]. Vol. 56, no. 6, p. 618–626 [visité le 2024-02-01]. ISSN 1573-4927. Disp. à l'adr. DOI: 10.1007/s10528-018-9864-7.
- SAJKO, S. & STUBER, K., 2009. Psoas Major: A Case Report and Review of Its Anatomy, Biomechanics, and Clinical Implications. *The Journal of the Canadian Chiropractic Association* [en ligne]. Vol. 53, no. 4, p. 311–318 [visité le 2025-09-18]. ISSN 0008-3194. Disp. à l'adr. PMID: 20037696.
- SÁNCHEZ-MOLANO, E. et al., 2014. Quantitative Trait Loci Mapping for Canine Hip Dysplasia and Its Related Traits in UK Labrador Retrievers. *BMC Genomics* [en ligne]. Vol. 15, no. 1 [visité le 2020-05-13]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/1471-2164-15-833.
- SARIKAYA, B. et al., 2024. Histological Structure and Immunohistochemical Properties of the Ligamentum Teres in Patients With Developmental Dysplasia of the Hip. *Cureus* [en ligne] [visité le 2024-06-04]. ISSN 2168-8184. Disp. à l'adr. DOI: 10.7759/cureus.59748.
- SCHACHNER, E. R. & LOPEZ, M. J., 2015. Diagnosis, Prevention, and Management of Canine Hip Dysplasia: A Review. *Veterinary Medicine : Research and Reports* [en ligne]. Vol. 6, p. 181–192 [visité le 2020-08-04]. ISSN 2230-2034. Disp. à l'adr. DOI: 10.2147/VMRR.S53266.
- SCHWEND, R. M. et al., 2014. Evaluation and Treatment of Developmental Hip Dysplasia in the Newborn and Infant. *Pediatric Clinics of North America*. Vol. 61, no. 6, p. 1095–1107. ISSN 1557-8240. Disp. à l'adr. DOI: 10.1016/j.pcl.2014.08.008.
- SEKIMOTO, T. et al., 2017. Copy Number Loss in the Region of the ASPN Gene in Patients with Acetabular Dysplasia: ASPN CNV in Acetabular Dysplasia. *Bone & Joint Research* [en ligne]. Vol. 6, no. 7, p. 439–445 [visité le 2024-02-01]. ISSN 2046-3758. Disp. à l'adr. DOI: 10.1302/2046-3758.67.BJR-2016-0094.R1.
- SEKIMOTO, T. et al., 2013. Segmental Copy Number Loss in the Region of Semaphorin 4D Gene in Patients with Acetabular Dysplasia. *Journal of Orthopaedic Research* [en ligne]. Vol. 31, no. 6, p. 957–961 [visité le 2024-02-02]. ISSN 1554-527X. Disp. à l'adr. DOI: 10.1002/jor.22310.

- SHARP, I. K., 1961. Acetabular Dysplasia: The Acetabular Angle. *The Journal of Bone and Joint Surgery. British volume* [en ligne]. Vol. 43-B, no. 2, p. 268–272 [visit  le 2022-12-12]. ISSN 0301-620X, ISSN 2044-5377. Disp.   l'adr. DOI: 10.1302/0301-620X.43B2.268.
- SHI, D. et al., 2011. Association of the D Repeat Polymorphism in the ASPN gene with Developmental Dysplasia of the Hip: A Case-Control Study in Han Chinese. *Arthritis Research & Therapy* [en ligne]. Vol. 13, no. 1, R27 [visit  le 2022-10-28]. ISSN 1478-6362. Disp.   l'adr. DOI: 10.1186/ar3252.
- SHI, D. et al., 2014. A Replication Study for the Association of Rs726252 in PAPP2 with Developmental Dysplasia of the Hip in Chinese Han Population. *BioMed research international*. Vol. 2014, p. 979520. Disp.   l'adr. DOI: 10.1155/2014/979520.
- SHORTER, D. et al., 2011. Screening Programmes for Developmental Dysplasia of the Hip in Newborn Infants. *The Cochrane Database of Systematic Reviews* [en ligne]. Vol. 2011, no. 9, p. CD004595 [visit  le 2024-01-04]. ISSN 1469-493X. Disp.   l'adr. DOI: 10.1002/14651858.CD004595.pub2.
- SILVA, M. A. da et al., 2009. Cellular and Epigenetic Features of a Young Healthy and a Young Osteoarthritic Cartilage Compared with Aged Control and OA Cartilage. *Journal of Orthopaedic Research* [en ligne]. Vol. 27, no. 5, p. 593–601 [visit  le 2024-02-01]. ISSN 1554-527X. Disp.   l'adr. DOI: 10.1002/jor.20799.
- SKIRVING, A. P. & SCADDEN, W. J., 1979. The African Neonatal Hip and Its Immunity from Congenital Dislocation. *The Journal of Bone and Joint Surgery. British Volume*. Vol. 61-B, no. 3, p. 339–341. ISSN 0301-620X. Disp.   l'adr. DOI: 10.1302/0301-620X.61B3.479257.
- SLOCUM, B. & DEVINE, T., 1990. Dorsal Acetabular Rim Radiographic View for Evaluation of the Canine Hip. *Journal of The American Animal Hospital Association* [en ligne] [visit  le 2023-12-11].
- SMITH, G. K. et al., 1990. New Concepts of Coxofemoral Joint Stability and the Development of a Clinical Stress-Radiographic Method for Quantitating Hip Joint Laxity in the Dog. *Journal of the American Veterinary Medical Association*. Vol. 196, no. 1, p. 59–70.
- SMITH, G., 1998. Canine Hip Dysplasia: Pathogenesis, Diagnosis, and Genetic Control. *Veterinary Quarterly* [en ligne]. Vol. 20, S22–S24 [visit  le 2023-12-08]. ISSN 0165-2176, ISSN 1875-5941. Disp.   l'adr. DOI: 10.1080/01652176.1998.10807390.
- SMITH, G. K. et al., 2001. Evaluation of Risk Factors for Degenerative Joint Disease Associated with Hip Dysplasia in German Shepherd Dogs, Golden Retrievers, Labrador Retrievers, and Rottweilers. *Journal of the American Veterinary Medical Association* [en ligne]. Vol. 219, no. 12, p. 1719–1724 [visit  le 2021-11-24]. Disp.   l'adr. DOI: 10.2460/javma.2001.219.1719.
- SOLLAZZO, V. et al., 2000. A Two-Locus Model for Non-Syndromic Congenital Dysplasia of the Hip (CDH). *Annals of Human Genetics* [en ligne]. Vol. 64, no. 1, p. 51–59 [visit  le 2022-11-04]. ISSN 1469-1809. Disp.   l'adr. DOI: 10.1046/j.1469-1809.2000.6410051.x.
- SPAIN, C. V. et al., 2004. Long-Term Risks and Benefits of Early-Age Gonadectomy in Dogs. *Journal of the American Veterinary Medical Association*. Vol. 224, no. 3, p. 380–387.
- STEIN-ZAMIR, C. et al., 2008. Developmental Dysplasia of the Hip: Risk Markers, Clinical Screening and Outcome. *Pediatrics International* [en ligne]. Vol. 50, no. 3, p. 341–345 [visit  le 2022-11-04]. ISSN 1442-200X. Disp.   l'adr. DOI: 10.1111/j.1442-200X.2008.02575.x.
- STEINETZ, B. G. et al., 1987. Plasma Relaxin Levels in Pregnant and Lactating Dogs. *Biology of reproduction*. Vol. 37, no. 3, p. 719–725.

- STEINETZ, B. G. et al., 2008. Transmission of Relaxin and Estrogens to Suckling Canine Pups via Milk and Possible Association with Hip Joint Laxity. *American Journal of Veterinary Research* [en ligne]. Vol. 69, no. 1, p. 59–67 [visité le 2020-08-06]. ISSN 0002-9645. Disp. à l'adr. DOI: 10.2460/ajvr.69.1.59.
- SUN, Y. et al., 2015. A Common Variant Of Ubiquinol-Cytochrome c Reductase Complex Is Associated with DDH. *PLOS ONE* [en ligne]. Vol. 10, no. 4, e0120212 [visité le 2022-11-04]. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0120212.
- SUN, Y. et al., 2019a. 3D-bioprinting a Genetically Inspired Cartilage Scaffold with GDF5-conjugated BMSC-laden Hydrogel and Polymer for Cartilage Repair. *Theranostics* [en ligne]. Vol. 9, no. 23, p. 6949–6961 [visité le 2025-01-24]. ISSN 1838-7640. Disp. à l'adr. DOI: 10.7150/thno.38061.
- SUN, Y. et al., 2019b. Genetic Variant of WIF1 Gene Is Functionally Associated with Developmental Dysplasia of the Hip in Han Chinese Population. *Scientific Reports* [en ligne]. Vol. 9, no. 1, p. 285 [visité le 2024-02-06]. ISSN 2045-2322. Disp. à l'adr. DOI: 10.1038/s41598-018-36532-8.
- TANIMOTO, K. et al., 2010. Modulation of Hyaluronan Catabolism in Chondrocytes by Mechanical Stimuli. *Journal of Biomedical Materials Research Part A* [en ligne]. Vol. 93A, no. 1, p. 373–380 [visité le 2020-08-07]. ISSN 1552-4965. Disp. à l'adr. DOI: 10.1002/jbm.a.32540.
- TERJESEN, T., 1996. Ultrasound as the Primary Imaging Method in the Diagnosis of Hip Dysplasia in Children Aged < 2 Years. *Journal of Pediatric Orthopedics. Part B*. Vol. 5, no. 2, p. 123–128. ISSN 1060-152X. Disp. à l'adr. DOI: 10.1097/01202412-199605020-00013.
- TIAN, W. et al., 2012. Association Analysis between HOXD9 Genes and the Development of Developmental Dysplasia of the Hip in Chinese Female Han Population. *BMC Musculoskeletal Disorders* [en ligne]. Vol. 13, no. 1, p. 59 [visité le 2024-01-26]. ISSN 1471-2474. Disp. à l'adr. DOI: 10.1186/1471-2474-13-59.
- TODHUNTER, R. J. et al., 2005. Quantitative Trait Loci for Hip Dysplasia in a Cross-Breed Canine Pedigree. *Mammalian genome : official journal of the International Mammalian Genome Society*. Vol. 16, no. 9, p. 720–730. ISSN 0938-8990. Disp. à l'adr. DOI: 10.1007/s00335-005-0004-4.
- TODHUNTER, R. J. et al., 2019. Gene Expression in Hip Soft Tissues in Incipient Canine Hip Dysplasia and Osteoarthritis. *Journal of Orthopaedic Research®*. Vol. 37, no. 2, p. 313–324.
- TOMÉ, I. et al., 2023. Mechanical Osteoarthritis of the Hip in a One Medicine Concept: A Narrative Review. *BMC Veterinary Research* [en ligne]. Vol. 19, no. 1, p. 222 [visité le 2023-11-09]. ISSN 1746-6148. Disp. à l'adr. DOI: 10.1186/s12917-023-03777-z.
- TOMLINSON, J. L. & JOHNSON, J. C., 2000. Quantification of Measurement of Femoral Head Coverage and Norberg Angle within and among Four Breeds of Dogs. *American Journal of Veterinary Research* [en ligne]. Vol. 61, no. 12, p. 1492–1500 [visité le 2020-08-05]. ISSN 0002-9645. Disp. à l'adr. DOI: 10.2460/ajvr.2000.61.1492.
- TÖNNIS, D., 1976. Normal Values of the Hip Joint for the Evaluation of X-rays in Children and Adults. *Clinical Orthopaedics and Related Research*. No. 119, p. 39–47. ISSN 0009-921X. Disp. à l'adr. PMID: 954321.
- TOPAK, D. et al., 2021. Are Vitamin D and Vitamin D Receptor Levels Different in Children with Developmental Dysplasia of the Hip? *Journal of Orthopaedic Surgery and Research* [en ligne]. Vol. 16, no. 1, p. 24 [visité le 2021-10-25]. ISSN 1749-799X. Disp. à l'adr. DOI: 10.1186/s13018-020-02162-y.

- TOPPARMAK, E. et al., 2025. The Association between Canine Hip Dysplasia and CHST14 Pseudogene Polymorphisms. Vol. 95, p. 211–218. Disp. à l'adr. DOI: 10.24099/vet.arhiv.2485.
- TORISU, T. et al., 1993. Association of Hla-Dr and Hla-Dq Antigens with Congenital Dislocation and Dysplastic Osteoarthritis of the Hip Joints in Japanese People. *Arthritis & Rheumatism* [en ligne]. Vol. 36, no. 6, p. 815–818 [visité le 2024-02-01]. ISSN 1529-0131. Disp. à l'adr. DOI: 10.1002/art.1780360612.
- TORRES DE LA RIVA, G. et al., 2013. Neutering Dogs: Effects on Joint Disorders and Cancers in Golden Retrievers. *PloS One*. Vol. 8, no. 2, e55937. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0055937.
- TRÉGUIER, C. et al., 2013. Pubo-Femoral Distance: An Easy Sonographic Screening Test to Avoid Late Diagnosis of Developmental Dysplasia of the Hip. *European Radiology*. Vol. 23, no. 3, p. 836–844. ISSN 1432-1084. Disp. à l'adr. DOI: 10.1007/s00330-012-2635-7.
- TSAI, K. L. & MURPHY, K. E., 2006. Clinical and Genetic Assessments of Hip Joint Laxity in the Boykin Spaniel. *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire*. Vol. 70, no. 2, p. 148–150. ISSN 0830-9000. Disp. à l'adr. PMID: 16639948.
- VAUGHAN-SCOTT, T. & TAYLOR, J. H., 1997. The Pathophysiology and Medical Management of Canine Osteoarthritis. *Journal of the South African Veterinary Association*. Vol. 68, no. 1, p. 21–25. ISSN 1019-9128. Disp. à l'adr. DOI: 10.4102/jsava.v68i1.861.
- VERHOEVEN, G. et al., 2008. Interobserver Agreement in the Diagnosis of Canine Hip Dysplasia Using the Standard Ventrodorsal Hip-Extended Radiographic Method. *The European Journal of Companion Animal Practice (EJCAP)*. Vol. 18, p. 149.
- VERHOEVEN, G. E. et al., 2010. The Effect of a Technical Quality Assessment of Hip-Extended Radiographs on Interobserver Agreement in the Diagnosis of Canine Hip Dysplasia. *Veterinary Radiology & Ultrasound*. Vol. 51, no. 5, p. 498–503.
- VITALE, M. G. & SKAGGS, D. L., 2001. Developmental Dysplasia of the Hip from Six Months to Four Years of Age. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*. Vol. 9, no. 6, p. 401–411.
- WALKER, J. M. & GOLDSMITH, C. H., 1981. Morphometric Study of the Fetal Development of the Human Hip Joint: Significance for Congenital Hip Disease. *The Yale Journal of Biology and Medicine* [en ligne]. Vol. 54, no. 6, p. 411–437 [visité le 2023-12-08]. ISSN 0044-0086. Disp. à l'adr. PMID: 7342490.
- WANG, E.-B. et al., 2008. [Expression of COL1a1 and COL3a1 in the Capsule of Children with Developmental Dislocation of the Hip]. *Zhongguo Dang Dai Er Ke Za Zhi = Chinese Journal of Contemporary Pediatrics*. Vol. 10, no. 4, p. 493–496. ISSN 1008-8830. Disp. à l'adr. PMID: 18706171.
- WANG, C.-L. et al., 2017. Cyclic Compressive Stress-Induced Scinderin Regulates Progress of Developmental Dysplasia of the Hip. *Biochemical and Biophysical Research Communications* [en ligne]. Vol. 485, no. 2, p. 400–408 [visité le 2024-02-02]. ISSN 0006-291X. Disp. à l'adr. DOI: 10.1016/j.bbrc.2017.02.065.
- WANG, C.-L. et al., 2020. The Long Noncoding RNA H19 Attenuates Force-Driven Cartilage Degeneration via miR-483-5p/Dusp5. *Biochemical and Biophysical Research Communications*. Vol. 529, no. 2, p. 210–217. ISSN 1090-2104. Disp. à l'adr. DOI: 10.1016/j.bbrc.2020.05.180.

- WANG, J. et al., 2022. MicroRNA-34a-5p Promotes the Progression of Osteoarthritis Secondary to Developmental Dysplasia of the Hip by Restraining SESN2-induced Autophagy.
- WANG, K. et al., 2010. Association of a Single Nucleotide Polymorphism in Tbx4 with Developmental Dysplasia of the Hip: A Case-Control Study. *Osteoarthritis and Cartilage* [en ligne]. Vol. 18, no. 12, p. 1592–1595 [visité le 2022-11-04]. ISSN 1063-4584. Disp. à l'adr. DOI: 10.1016/j.joca.2010.09.008.
- WANG, S. et al., 2021. Genome-Wide Association Studies for Canine Hip Dysplasia in Single and Multiple Populations – Implications and Potential Novel Risk Loci. *BMC Genomics* [en ligne]. Vol. 22, no. 1, p. 636 [visité le 2021-09-06]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/s12864-021-07945-z.
- WANG, X. et al., 1999. Analysis of Randomly Amplified Polymorphic DNA (RAPD) for Identifying Genetic Markers Associated with Canine Hip Dysplasia. *The Journal of heredity*. Vol. 90, no. 1, p. 99–103. ISSN 0022-1503. Disp. à l'adr. DOI: 10.1093/jhered/90.1.99.
- WATSON, C. M. et al., 2015. Identification of a Mutation in the Ubiquitin-Fold Modifier 1-Specific Peptidase 2 Gene, UFSP2, in an Extended South African Family with Beukes Hip Dysplasia. *South African Medical Journal* [en ligne]. Vol. 105, no. 7, p. 558–563 [visité le 2022-01-07]. ISSN 2078-5135. Disp. à l'adr. DOI: 10.7196/SAMJnew.7917.
- WIBERG, G., 1939. Studies on dysplastic acetabula and congenital subluxation of the hip joint with special reference to the complication of osteoarthritis. Stockholm: Exp., Norstedt. *Acta chirurgica Scandinavica*.
- WILLEMSSEN, K. et al., 2021. Comparing Hip Dysplasia in Dogs and Humans: A Review. *Frontiers in Veterinary Science* [en ligne]. Vol. 8 [visité le 2022-10-27]. ISSN 2297-1769.
- WILLIS, M. B., 1997. A Review of the Progress in Canine Hip Dysplasia Control in Britain. *Journal of the American Veterinary Medical Association*. Vol. 210, no. 10, p. 1480–1482. ISSN 0003-1488. Disp. à l'adr. PMID: 9154201.
- WYATT, M. et al., 2017. The Femoro-Epiphyseal Acetabular Roof (FEAR) Index: A New Measurement Associated With Instability in Borderline Hip Dysplasia? *Clinical Orthopaedics and Related Research®* [en ligne]. Vol. 475, no. 3, p. 861 [visité le 2023-12-14]. ISSN 0009-921X. Disp. à l'adr. DOI: 10.1007/s11999-016-5137-0.
- XU, J. et al., 2022a. WNT1 Expression Influences the Development of Dysplasia of the Hip via Regulating RBPMS2/NOG-BMP2/4-GDF5- WISP2 Pathway. *Nucleosides, Nucleotides & Nucleic Acids* [en ligne]. Vol. 0, no. 0, p. 1–13 [visité le 2022-06-20]. ISSN 1525-7770. Disp. à l'adr. DOI: 10.1080/15257770.2022.2081337.
- XU, R. et al., 2020. Genetic Variant of COL11A2 Gene Is Functionally Associated with Developmental Dysplasia of the Hip in Chinese Han Population. *AGING-US* [en ligne]. Vol. 12, no. 9, p. 7694–7703 [visité le 2021-10-25]. ISSN 19454589. Disp. à l'adr. DOI: 10.18632/aging.103040.
- XU, R. et al., 2021a. Secreted Frizzled-Related Protein 3 Was Genetically and Functionally Associated with Developmental Dysplasia of the Hip. *Aging*. Vol. 13, no. 8, p. 11281–11295. ISSN 1945-4589. Disp. à l'adr. DOI: 10.18632/aging.202815.
- XU, X. et al., 2016. Lack of Association of a Single Nucleotide Polymorphism in SMOC1 with Developmental Dysplasia of the Hip: A Case-Control Study. *International Journal of Clinical and Experimental Pathology*. Vol. 9, p. 10734–10739.

- XU, X. et al., 2021b. Replicative Verification of Susceptibility Genes Previously Identified from Families with Segregating Developmental Dysplasia of the Hip. *Italian Journal of Pediatrics* [en ligne]. Vol. 47, no. 1, p. 140 [visité le 2024-02-06]. ISSN 1824-7288. Disp. à l'adr. DOI: 10.1186/s13052-021-01087-4.
- XU, X. et al., 2022b. Identification of KANSL1 as a Novel Pathogenic Gene for Developmental Dysplasia of the Hip. *Journal of Molecular Medicine*, p. 1–10.
- YAMANAKA, M. et al., 2009. Association of Oestrogen Receptor Gene Polymorphism with the Long-Term Results of Rotational Acetabular Osteotomy. *International Orthopaedics* [en ligne]. Vol. 33, no. 4, p. 1155–1164 [visité le 2024-02-02]. ISSN 1432-5195. Disp. à l'adr. DOI: 10.1007/s00264-009-0730-4.
- YAN, W. et al., 2019. A Genome-Wide Association Study Identifies New Genes Associated with Developmental Dysplasia of the Hip. *Clinical Genetics* [en ligne]. Vol. 95, no. 3, p. 345–355 [visité le 2024-02-01]. ISSN 1399-0004. Disp. à l'adr. DOI: 10.1111/cge.13483.
- YAN, W. et al., 2022. Heterozygous LRP1 Deficiency Causes Developmental Dysplasia of the Hip by Impairing Triradiate Chondrocytes Differentiation Due to Inhibition of Autophagy. *Proceedings of the National Academy of Sciences* [en ligne]. Vol. 119, no. 37, e2203557119 [visité le 2022-09-12]. Disp. à l'adr. DOI: 10.1073/pnas.2203557119.
- YANG, W. et al., 2022. Comprehensive Bioinformatics Analysis of Susceptibility Genes for Developmental Dysplasia of the Hip. *Intractable & Rare Diseases Research*.
- YILMAZ, A. et al., 2015. The Role of ADAMTS9 Gene Promoter (Ca) Repeat in Developmental Hip Dysplasia. *Clinical and investigative medicine. Medecine clinique et experimentale*.
- ZHANG, J. et al., 2018. Association Analysis on Polymorphisms in WISP3 Gene and Developmental Dysplasia of the Hip in Han Chinese Population: A Case-Control Study. *Gene* [en ligne]. Vol. 664, p. 192–195 [visité le 2024-01-26]. ISSN 0378-1119. Disp. à l'adr. DOI: 10.1016/j.gene.2018.04.020.
- ZHANG, L. et al., 2020. Mapping of Developmental Dysplasia of the Hip to Two Novel Regions at 8q23-Q24 and 12p12. *Experimental and Therapeutic Medicine*. Vol. 19, no. 4, p. 2799–2803. ISSN 1792-0981. Disp. à l'adr. DOI: 10.3892/etm.2020.8513.
- ZHAO, L. et al., 2017. A Recurrent Mutation in Bone Morphogenetic Proteins-2-Inducible Kinase Gene Is Associated with Developmental Dysplasia of the Hip. *Experimental and Therapeutic Medicine*. Vol. 13, no. 5, p. 1773–1778. ISSN 1792-0981. Disp. à l'adr. DOI: 10.3892/etm.2017.4191.
- ZHAO, L. et al., 2013a. Two Single Nucleotide Polymorphisms in the GDF5 Gene Are Associated with Development Dysplasia of the Hip in Chinese Female Population. *Science China Life Sciences* [en ligne]. Vol. 56, no. 11, p. 1063–1065 [visité le 2024-02-06]. ISSN 1869-1889. Disp. à l'adr. DOI: 10.1007/s11427-013-4514-0.
- ZHAO, L. et al., 2013b. Variations of the COL1A1 Gene Promoter and the Relation to Developmental Dysplasia of the Hip. *Genetic Testing and Molecular Biomarkers* [en ligne]. Vol. 17, no. 11, p. 840–843 [visité le 2024-01-26]. ISSN 1945-0265. Disp. à l'adr. DOI: 10.1089/gtmb.2013.0179.
- ZHOU, Z. et al., 2010. Differential Genetic Regulation of Canine Hip Dysplasia and Osteoarthritis. *PloS one*. Vol. 5, no. 10, e13219.
- ZHU, L. et al., 2008. Single Nucleotide Polymorphisms Refine QTL Intervals for Hip Joint Laxity in Dogs. *Animal Genetics*. Vol. 39, no. 2, p. 141–146.

- ZHU, L. et al., 2012. Identification of Quantitative Trait Loci for Canine Hip Dysplasia by Two Sequential Multipoint Linkage Analyses. *Journal of Applied Statistics* [en ligne]. Vol. 39, no. 8, p. 1719–1731 [visité le 2022-11-10]. ISSN 0266-4763. Disp. à l'adr. DOI: 10.1080/02664763.2012.673121.
- ZHU, L.-Q. et al., 2019. Whole Genome Sequencing of Pairwise Human Subjects Reveals DNA Mutations Specific to Developmental Dysplasia of the Hip. *Genomics*. Vol. 111, no. 3, p. 320–326. ISSN 1089-8646. Disp. à l'adr. DOI: 10.1016/j.ygeno.2018.02.006.
- ZHU, L. et al., 2011. Lack of Evidence for Association between DVWA Gene Polymorphisms and Developmental Dysplasia of the Hip in Chinese Han Population. *Rheumatology International*. Vol. 31, no. 7, p. 883–887. ISSN 1437-160X. Disp. à l'adr. DOI: 10.1007/s00296-010-1410-9.
- ZIMMERER, A. et al., 2021. Defining the Gothic Arch Angle (GAA) as a Radiographic Diagnostic Tool for Instability in Hip Dysplasia. *Scientific Reports* [en ligne]. Vol. 11, no. 1, p. 19531 [visité le 2021-10-04]. ISSN 2045-2322. Disp. à l'adr. DOI: 10.1038/s41598-021-99011-7.

DÉVELOPPEMENT D'OUTILS DE BIO-INFORMATIQUES ET DE GÉNÉTIQUE COMPUTATIONNELLE

A Développement d'un outil de concaténation de données hétérogènes

1 Introduction

L'un des premiers défis rencontrés au cours de cette thèse a été la consolidation d'un corpus de données fragmenté, hétérogène et évolutif. En effet, les données disponibles sur les chiens étudiés et collectés par l'équipe depuis les années 2008, étaient réparties dans plus de 70 de fichiers *Excel* et *CSV* distincts, issus de différentes sources (e.g., ACGAO, lecteurs officiels, clubs de race, etc.). A cela s'ajoute la base de données du CRB Cani-DNA dans laquelle étaient rentrés tous les chiens pour lesquels il y avait eu un prélèvement. Ces fichiers contiennent des données cliniques, généalogiques, morphologiques et diagnostiques et étaient continuellement enrichis par de nouvelles informations parfois redondantes, souvent incomplètes ou saisies selon des conventions variables, rendant toute exploitation directe fastidieuse voire impossible.

L'objectif principal a donc été de créer une base de données unique, structurée et normalisée, regroupant l'ensemble des informations disponibles pour chaque individu. Cette base doit permettre d'identifier de manière fiable les individus, de croiser les diagnostics de dysplasie coxo-fémorale issus de différents âges, méthodes et lecteurs et d'intégrer des informations additionnelles (e.g., adresse du vétérinaire, suivi santé, réforme, etc.). Elle doit en outre permettre la mise à jour automatique à partir de nouveaux fichiers, tout en assurant la traçabilité des modifications, la reproductibilité du processus de fusion et une visibilité immédiate sur les éventuelles erreurs ou contradictions entre sources.

2 Développement

Le travail de structuration et d'intégration des données s'est déroulé en deux grandes phases, répondant à l'évolution progressive des besoins et à l'augmentation du volume de données à traiter.

a Phase 1 : Consolidation manuelle et macros *Excel*

Dans un premier temps j'ai utilisé une solution de structuration semi-automatisée sous *Excel*. Cette solution consistait à créer un tableau de référence rassemblant les noms d'individus sans doublons et un ensemble de feuilles satellites correspondant à chaque fichier source. Pour chaque source, les données étaient recopiées dans une feuille brute, puis transformées selon le modèle de référence dans une feuille normalisée via des formules de recherche verticale.

Plusieurs macros ont été développées pour automatiser les points suivants :

- propagation des formules sur l'ensemble de la feuille ;
- ajout des lignes et colonnes de contrôle pour comptabiliser les données présentes ;
- standardisation de l'écriture des colonnes du genre, dates (e.g., de naissance, de prélèvement) et de la race des individus.

Des feuilles de contrôle permettait alors d'identifier automatiquement les incohérences entre les différentes sources et de résumer l'ensemble des données. Cette solution a permis de poser les bases d'un tableau consolidé ainsi que de définir les clés d'identifications multiples. Cependant elle s'est révélée rapidement limitée en termes de réactivité, de mise à l'échelle et de robustesse. La taille croissante des fichiers (jusqu'à 90 Mo) et la complexité des fonctions utilisées ont rapidement rendu impossible le maintien de cette solution.

b Phase 2 : Automatisation par script et base de données

Pour pallier les limites rencontrées avec les fichiers *Excel*, une seconde phase de travail s'est concentrée sur le développement d'un script automatisant l'agrégation et la normalisation des fichiers de données. Ce script repose sur un fichier CSV de configuration listant les sources à intégrer. Il est capable de parcourir les différents fichiers, d'en extraire les informations pertinentes, de les transformer selon des règles prédéfinies, puis de les injecter dans une base de données centralisée.

Une première version de cet outil a été développé en utilisant *MongoDB* (i.e., base de données *NoSQL*) (Győrödi et al., 2015) couplé à une interface en *tkinter* (voir figure 13). Bien que fonctionnelle, cette solution s'est révélée complexe à maintenir et relativement lente pour compiler les données, en particulier avec de gros volumes. Je l'ai donc remplacée par une version entièrement développée en *Python*, s'appuyant sur le package *pandas* (The pandas development team, 2025). Cette nouvelle approche, pensée pour maximiser la vectorisation des opérations, a significativement amélioré les performances en termes de vitesse de traitement et de gestion de la mémoire.

Le système ainsi mis en place permet désormais, à partir d'une solide base de données dédiées au projet DCF, une mise à jour simple, rapide et reproductible à chaque ajout

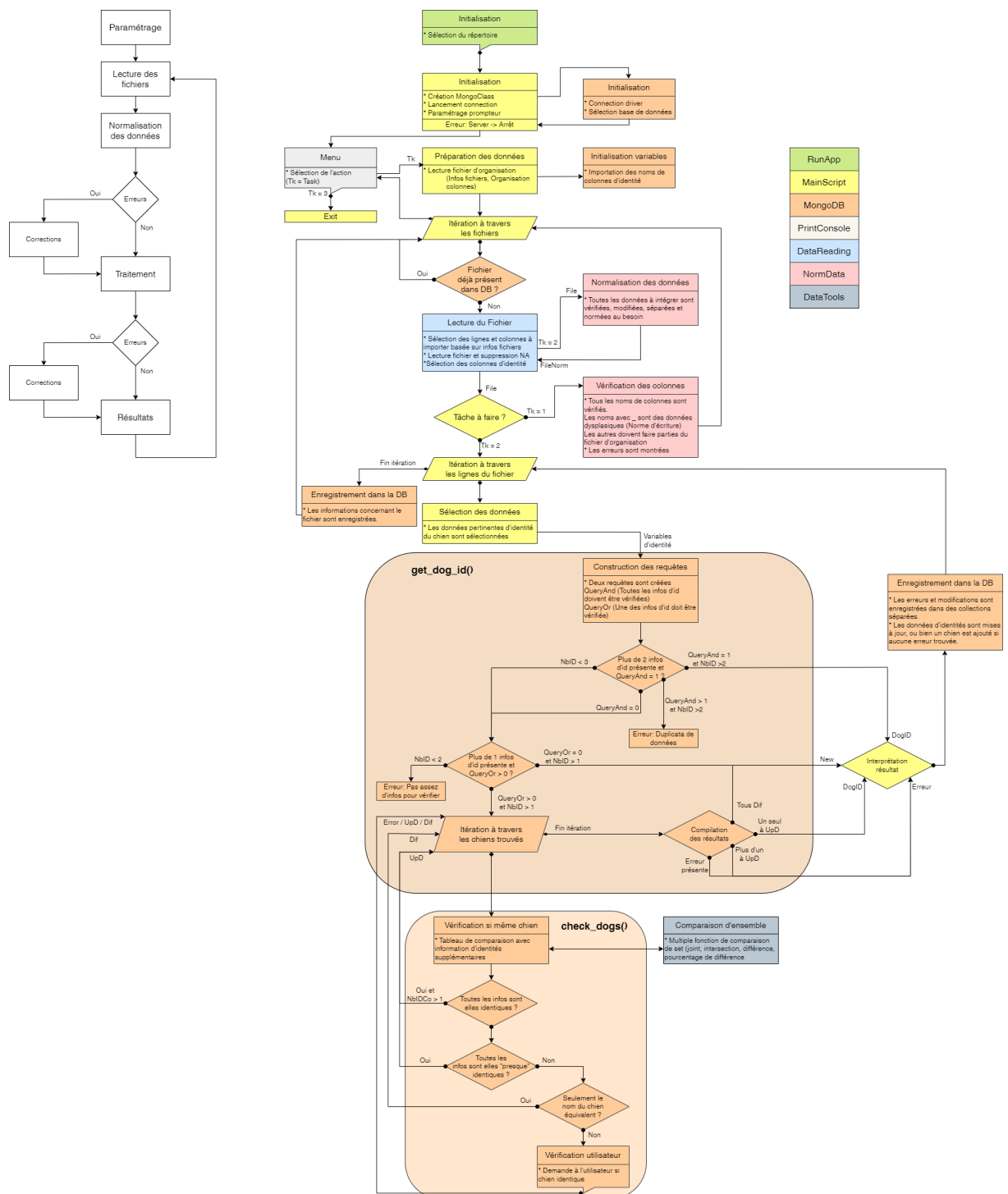


FIGURE 13 – Diagramme fonctionnel d'agrégation des données par le biais d'un script Python et de la base NoSQL MongoDB.

de nouvelles données. Les formats des champs (e.g., dates, noms, races, sexe, etc.) sont automatiquement contrôlés et standardisés.

Cette étape de développement a permis de disposer d'une base de données robuste, évolutive et interopérable, capable de traiter efficacement de larges volumes de données hétérogènes issues de sources multiples. Cette étape constitue désormais le maillon initial de la préparation des données, essentielle aux analyses génétiques et épidémiologiques menées dans le cadre de cette thèse. À terme, l'objectif est de rendre cet outil configurable de manière plus souple, via un simple fichier de configuration, afin de le rendre adaptable à d'autres projets de recherche ou contextes d'analyse.

3 Données d'entrées et de sorties

Le fichier principal contient la liste des fichiers de données à traiter et leurs métadonnées (e.g., chemin d'accès, nom de la feuille à traiter, lignes et colonnes de départ et de fin, origine des données, etc.). Le fichier d'organisation, permet de lister la structure des fichiers, les règles de transformation et les relations entre les différentes variables. Tous les fichiers listés dans le fichier principal doivent être accessibles. Ils représentent les fichiers de données réelles (e.g., *Excel*, *CSV*, etc.) à traiter. Chaque fichier contient des données cliniques, généalogiques ou d'identification. Ils seront itérativement agrégés avant que chacune des variables ne soit transformées et vérifiées.

L'outil de traitement que j'ai développé repose sur l'utilisation de plusieurs fichiers d'entrée distincts, chacun jouant un rôle pour la préparation des données.

Le fichier principal liste l'ensemble des fichiers de données à intégrer, accompagnés de leurs métadonnées indispensables : chemin d'accès, nom de la feuille à traiter pour les fichiers *Excel*, coordonnées précises de la zone de lecture (i.e., lignes et colonnes de départ et de fin), ainsi que l'origine des données (e.g., ACGAO, lecteurs, clubs de races, *Antagene*, *Cani-DNA*, etc.). Cette structure permet d'assurer une flexibilité maximale et une interopérabilité avec des fichiers hétérogènes.

Un second fichier, qualifié de fichier d'organisation, définit quant à lui la structure logique des données à intégrer. Il liste les règles de transformation, les relations entre les variables et les conventions de nommage standardisé à appliquer. Ce fichier joue un rôle fondamental dans la normalisation automatique des valeurs (e.g., codage du sexe, formatage des dates, harmonisation des noms de race), garantissant ainsi la cohérence du jeu de données final.

Les fichiers de données eux-mêmes, référencés dans le fichier principal, sont issus de sources diverses. Ils peuvent prendre la forme de fichiers *Excel* (souvent multi-feuilles) ou de fichiers *CSV*. Ils contiennent des informations cliniques (e.g., résultats de dépistage de la dysplasie coxo-fémorale, évaluations FCI, angles de Norberg, indices de distraction, etc.), des informations d'identification (e.g., nom usuel et officiel,

numéro de puce ou de tatouage, date de naissance, sexe, race, etc.) et des données généalogiques (e.g., noms des parents, affixe, filiation).

Avant intégration, l'outil procède à une lecture itérative de l'ensemble de ces fichiers, puis applique les règles de transformation définies. Chaque variable est alors contrôlée, normalisée en fonction des formats attendus. Ce processus garantit non seulement la qualité des données consolidées mais permet aussi une traçabilité des erreurs et des corrections sans modification des données sources.

Une fois ce processus terminé, l'outil exporte un fichier *CSV* consolidé, unique, contenant l'ensemble des informations agrégées. Chaque ligne correspond à une observation par chien de chaque fichier source, permettant de conserver la granularité des données. En parallèle, un fichier de suivi des erreurs est généré. Il recense l'ensemble des incohérences détectées au cours du traitement (e.g., format non conforme, doublons, données contradictoires, identifiants ambigus, etc.). Certaines erreurs peuvent être corrigées automatiquement via des règles de normalisation ; d'autres, irréductibles, nécessitent une intervention manuelle. Leur traçabilité est assurée par ce fichier, facilitant leur repérage et leur résolution rapide par l'utilisateur.

4 Résultats

L'application de l'outil de concaténation et de normalisation développé en *Python* sur l'ensemble des fichiers collectés dans le cadre du projet dysplasie, provenant de multiples sources partenaires, a permis de générer une base de données consolidée de plus de 88 000 lignes, réparties sur plus de 400 colonnes, pour un total de plus de 3 millions de données renseignées. Une étape complémentaire a ensuite été réalisée sous *R*, via un script dédié, afin d'assurer l'identification unique de chaque individu. Ce traitement a permis de détecter et de corriger de nombreuses erreurs d'entrée, notamment des fautes de frappe dans les numéros d'identification (i.e., puce ou tatouage), des variations d'orthographe dans les noms (e.g., Freesbee *versus* Frisbee) ou encore des doublons issus de sources différentes.

Un second traitement a ensuite été effectué pour consolider l'historique diagnostique de chaque individu à travers les différentes évaluations disponibles (e.g., à différents âges ou par différents opérateurs). La première étape consiste à reconstruire l'historique complet des diagnostics pour les différentes méthodes d'évaluation (e.g., notation FCI, angle de Norberg-Olsson, indice de distraction) et pour chaque date d'examen identifiée. Cette première synthèse permet de mieux appréhender l'évolution du diagnostic au fil du temps et de détecter des discordances d'évaluation. Une deuxième étape s'ensuit, afin de calculer un score global de sévérité de la dysplasie pour chaque individu. Elle filtre les données, puis convertit les différentes mesures en équivalent FCI avant de compiler une note globale cohérente.

À l'issue de ces phases d'identification et de compilation des diagnostics, plus de 42 000 individus uniques ont été identifiés, représentant 784 000 points de données exploitables. De ce nombre considérable doit être sélectionner les individus ayant à la fois un diagnostic de dysplasie ($N = 11\,000$) et ceux ayant un prélèvement ADN disponible ($N = 20\,000$). Nous obtenons ainsi un total exploitable de 4 800 individus avec de l'ADN disponible à Cani-DNA et un diagnostic de dysplasie, toutes races confondues.

Dans le cadre de l'étude des bases génétiques de la dysplasie coxo-fémorale, seuls les individus de races Labrador et Golden Retriever, croisés Labrador x Golden, Cane Corso, Dogue de Bordeaux et Bouvier Bernois ayant des valeurs de phénotypes extrêmes ont été sélectionnés (i.e., FCI A *versus* FCI D ou E). Le tableau 3 répertorie l'ensemble des chiens suivant ces critères et dont un prélèvement était disponible en mai 2024.

A partir des individus présentés dans ce tableau nous avons sélectionnés ceux qui seront utilisés dans les analyses génétiques, présentées ci-après. L'objectif d'une centaine d'individus indemnes de la DCF a pu être atteint pour l'ensemble des races sauf les croisés Labrador et Golden retriever. En revanche il nous manque encore des chiens atteints pour les races des retrievers. Ce manque peut s'expliquer par le fait que depuis le début de ce projet en 2008, les chiens guides atteints étaient encore nombreux, mais les écoles de chiens guides ont fait de gros efforts de sélection ces 10 dernières années, diminuant ainsi le nombre de retrievers sévèrement atteints. Pour le projet, nous analyserons ces individus ensembles (i.e., Golden, Labrador et croisés) en tant que groupe des retrievers afin d'atteindre l'objectif d'une centaine d'individus atteints et une centaine indemne par race ou groupe de race.

TABLE 3 – Bilan global des prélèvements disponibles en mai 2024 dans les races d'intérêt

Race	A	B/C	D/E	Total
Labrador Retriever	316	154	62	532
Golden Retriever	119	89	47	255
Labrador x Golden	29	22	15	66
Total Retrievers	464	265	124	853
Cane Corso	135	68	79	282
Dogue de Bordeaux	394	135	197	726
Bouvier Bernois	1409	182	93	1684
Total Mastiff	1938	385	369	2692
Total	2402	650	493	3545

DÉVELOPPEMENT D'UN PACKAGE *R* : *PEDIXPLORER*

B Développement d'un package *R Bioconductor* de construction de pedigree : *Pedixplorer*

1 Présentation

Dans le cadre du projet de recherche présenté ici, un des objectifs initiaux était de procéder à des analyses de liaison génétique en sélectionnant des familles dites informatives, c'est-à-dire comprenant à la fois des individus atteints et des individus sains sur plusieurs générations et pour lesquels un échantillon d'ADN était disponible. Ces analyses nécessitent de sélectionner des familles informatives au sein de pedigrees canin souvent complexes. Ceux-ci sont en effet constitués de plusieurs centaines voire milliers d'individus, présentent des boucles de consanguinité ainsi que des croisements intergénérationnels. La manipulation efficace de telles structures nécessite des outils capables de gérer des volumes de données importants, tout en offrant des possibilités d'exploration et de visualisation simplifiées, précises et reproductibles.

Face à cette problématique, le développement d'un outil permettant de créer, filtrer, explorer et visualiser de très grands pedigrees s'est avéré nécessaire, par rapport aux logiciels existants et d'ailleurs précédemment utilisés par l'équipe comme *Cyrillic 3* (Cyril Chapman et al., 1999). L'objectif était d'identifier plus aisément des sous-ensembles familiaux pertinents pour les analyses génétiques, tout en assurant une représentation rigoureuse des structures généalogiques. Bien que des solutions logicielles préexistantes existent, telles que *Madelaine2* (Trager et al., 2007) ou le package *kinship2* (Sinnwell et al., 2014) en langage *R* (Team, 2025), celles-ci montraient des limites importantes en termes de performance, de flexibilité graphique et de prise en charge des pedigrees complexes.

Dans ce contexte j'ai conçu *Pedixplorer*, un package *R* développé sous *Bioconductor*, spécifiquement dédié à la gestion et à l'exploration de pedigrees complexes. Il repose sur une refonte complète du package *R kinship2* (Sinnwell et al., 2014). Il utilise les standards modernes de programmation du langage *R*, notamment les objets *S4*. Cette architecture robuste permet une extensibilité du code et une meilleure compatibilité avec les standards *Bioconductor*. Ce nouveau package propose donc un ensemble de fonctions pour importer, nettoyer, filtrer, trier, simplifier et visualiser des pedigrees de grande taille et comportant de nombreuses « boucles » due à la consanguinité.

La fonction centrale `Pedigree()` permet de créer un objet *S4* de type `Pedigree`. Cet objet repose sur une structure de vecteurs parallèles, dans laquelle chaque individu est défini par un identifiant unique, ainsi que par les identifiants de ses parents (i.e., père et mère). De nombreuses métadonnées peuvent être associées à chaque individu, telles que la disponibilité d'un échantillon biologique, l'existence d'une relation gémellaire ou une adoption. Certaines informations sont automatiquement calculées

par le package, comme le nombre d'enfants ou la distance généalogique par rapport à un ensemble d'individus dits informatifs. L'un des atouts de ce package réside dans sa capacité à représenter simultanément plusieurs phénotypes, y compris des variables quantitatives, en utilisant des gradients de couleurs. Les paramètres de visualisation, notamment la couleur de remplissage et celle de la bordure de chaque symbole, sont stockés dans des objets distincts, permettant une personnalisation fine et modulaire de la représentation graphique des pedigrees.

La fonction `generate_colors()` permet de générer automatiquement une palette de couleurs associée aux différents statuts d'affection (i.e., atteint, porteur, indemne), enrichissant ainsi la visualisation. L'utilisateur peut également définir manuellement une échelle spécifique de couleurs.

La méthode `S4 plot()` permet de représenter visuellement la structure familiale sous forme d'un arbre généalogique. Cette fonction opère en deux étapes : la conversion de l'objet `Pedigree` en une table de données via la fonction `ped_to_plotdf()`, puis sa représentation graphique à l'aide de `plot_fromdf()`. *Pedixplorer* utilise les packages *ggplot2* et *plotly* pour permettre une personnalisation poussée du rendu et une visualisation interactive.

Le package inclut également la fonction `kinship()` pour calculer la matrice de parenté. Celle-ci prend en compte les relations complexes, les boucles de consanguinité, les jumeaux monozygotes, ainsi que les particularités des chromosomes sexuels (i.e., X et Y).

Afin de faciliter l'utilisation du package par un public non spécialiste de *R*, une application *shiny* (Chang et al., 2024) interactive a été développée. Elle permet d'importer des données, de les normaliser, de sélectionner des familles, d'appliquer des filtres et de visualiser le pedigree résultant. Les résultats peuvent être exportés sous forme de fichiers texte ou d'images.

Cette application est accessible en ligne via une machine virtuelle de l'université de Rennes : <https://pedixplorer.univ-rennes.fr> .

Parmi les autres fonctions notables, on peut citer :

- `useful_inds()`, qui identifie automatiquement les individus proches d'un groupe d'intérêt, facilitant la création de sous-pedigrees plus lisibles ;
- `shrink()`, qui réduit un pedigree à une taille donnée en conservant les individus les plus informatifs selon une variable d'affection ;
- `fix_parents()`, qui complète automatiquement les informations parentales manquantes et corrige les erreurs de sexe, afin de normaliser les données avant la création d'un objet `Pedigree`.

Le code dans la figure 14 représente un exemple d'utilisation de *Pedixplorer*. Un objet *Pedigree* est tout d'abord créé à partir des données tabulaires. Des données phénotypiques additionnelles sont ensuite ajoutées à l'objet, puis le pedigree est filtré à partir des individus considérés comme informatifs, qui sont eux même marqués comme le premier individu diagnostiqué (i.e., *proband*). Le graphique est ensuite généré avec la légende. On obtient alors le pedigree représenté en figure 15.

Le package *Pedixplorer* est disponible sous sa version 1.4.0 via *Bioconductor* 3.21. La documentation complète est accessible à l'adresse suivante : <https://louislenezet.github.io/Pedixplorer> .

Enfin, bien que les analyses de liaison initialement prévues dans le cadre de cette thèse n'aient pas pu être menées en raison de l'indisponibilité des données généalogiques de chiens d'intérêt, le développement de *Pedixplorer* constitue une contribution autonome à la recherche en génétique.

Il a été utilisé dans le cadre d'autres projets de l'équipe avec des familles complexes, mal ou difficilement représentées par les logiciels classiquement utilisés.

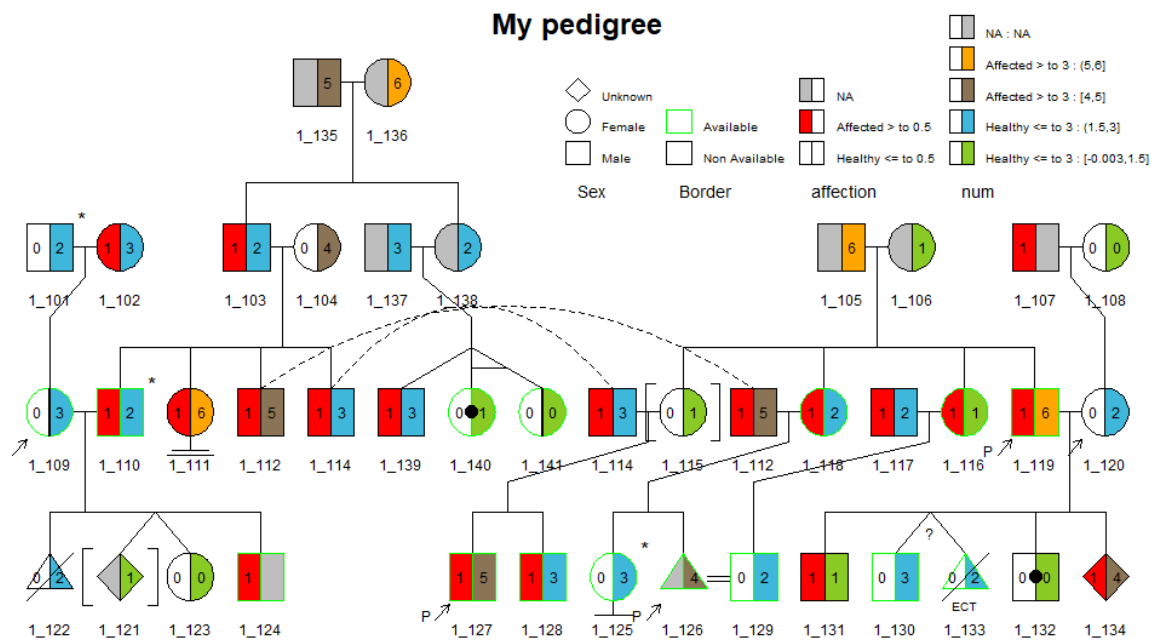
Cet outil peut être utilisé dans divers contextes, aussi bien en génétique animale qu'humaine. Il vise à démocratiser l'analyse de grandes structures généalogiques en combinant rigueur scientifique, performances informatiques et accessibilité via des interfaces conviviales et documentées.

FIGURE 14 – Exemple de code utilisant le package *Pedexplorer*

```

1 library(Pedexplorer)
2 library(dplyr)
3 data("sampleped")
4 data("relped")
5
6 # Create the Pedigree object
7 pedi <- Pedigree(sampleped, relped, missid = NA) %>%
8   generate_colors( # Add a new affection information
9     col_aff = "num", is_num = TRUE,
10     keep_full_scale = TRUE, breaks = 2,
11     threshold = 3,
12     colors_aff = c("#8B7355", "#FFA500"),
13     colors_unaff = c("#8aca25", "#3fb7db")
14   ) %>%
15   is_informative( # Set which individuals are informative
16     col_aff = "num", informative = "AvAf"
17   ) %>%
18   useful_inds(
19     keep_infos = TRUE, # Keep available or affected parents
20     max_dist = 2 # Maximum distance from informative individuals
21   )
22
23 proband(ped(pedi)) <- isinf(ped(pedi)) # Set informative individuals as proband
24
25 png("MyPedigree.png", width = 1000, height = 600)
26 plot_list <- plot(
27   pedi,
28   symbolsize = 1.5, # Increase the symbole size
29   title = "My pedigree", # Add a title
30   legend = TRUE, # Add the legend
31   leg_symbolsize = 0.02, # Set the symbole size of the legend
32   leg_loc = c(0.5, 0.9, 0.8, 1.1), # Specify the legend location
33   lwd = 0.5, # Set the line width
34   ggplot_gen = TRUE, # Use ggplot2 to draw the Pedigree
35   tips = c(
36     "id", "avail",
37     "affection",
38     "num", "dateofbirth"
39   ) # Add some information in the tooltip
40 )
41 dev.off()
42
43 # Plot the Pedigree with plotly to have an interactive plot
44 plotly::ggplotly(
45   plot_list$ggplot,
46   tooltip = "text"
47 ) %>%
48   plotly::layout(hoverlabel = list(bgcolor = "darkgrey"))
49
50

```

FIGURE 15 – Arbre généalogique obtenu en utilisant le package *Pedexplorer*


ARTICLE 3

“ Pedixplorer: a Bioconductor package to streamline pedigree design and visualization ”

AUTEURS : Louis Le Nézet, Jason Sinnwell, Anna Letko,
Pascale Quignon, Catherine André

STATUT : Publié

JOURNAL : *Bioinformatics*, Volume 41, Issue 6, June 2025,
btaf329

<https://doi.org/10.1093/bioinformatics/btaf329>

Pedixplorer : a *Bioconductor* package to streamline pedigree design and visualization

Louis Le Nézet^{1,*}, Jason Sinnwell², Anna Letko³, Catherine André¹, Pascale Quignon¹

¹ UMR-6290, IGDR (Institute Genetics and Development of Rennes), University of Rennes, CNRS, Rennes, France

² Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA

³ Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

* Corresponding author. louislenezet@gmail.com

1 Abstract

Motivation : Understanding kinship relationships is fundamental to genetic research, particularly in the context of genetic linkage studies and population genetics. Pedigree design and analysis are a prerequisite for these investigations. The legacy kinship2 CRAN package has been a cornerstone in this area; however, the need for handling larger and more complex datasets necessitates an updated, flexible, and user-friendly toolset. To address this issue, we present *Pedixplorer*, a novel Bioconductor package designed to enhance kinship analyses with modern functionality and usability, especially in large multigeneration complex pedigrees with inbreeding loops, which are frequently seen in domestic animal breeding.

Results : *Pedixplorer* builds upon the robust foundation of kinship2, integrating Bioconductor standards and most recent programming practices. Its core component is the S4 Pedigree object, facilitating efficient representation of complex pedigrees. The new functions enable automatic querying, filtering, and trimming of large pedigrees, while the graphical functions have been rewritten for better customization in pedigree visualizations. Additionally, *Pedixplorer* offers a comprehensive shiny application, accessible both locally and via a dedicated website, allowing non-R users to easily create, filter, and customize pedigrees.

Availability : The *Pedixplorer* package is freely available at : <https://www.bioconductor.org/packages/release/bioc/html/Pedixplorer.html> with additional documentation at : <https://louislenezet.github.io/Pedixplorer> . A user-friendly web application is available at : <https://pedixplorer.univ-rennes.fr> .

Supplementary information : are available at *Bioinformatics* online.

Keywords Pedigree, Visualization, Genetics, R, Bioconductor

2 Introduction

Pedigree analysis has historically been a cornerstone of genetic research, offering critical insights into inheritance patterns, familial aggregation of diseases, and population genetics. While the rise of high-throughput sequencing technologies and genome-wide association studies (GWAS) has led to a decline in traditional pedigree-based studies, linkage analysis remains highly relevant (Ott et al., 2015). It offers a powerful approach for detecting genetic loci with strong effects in studies with small sample sizes. Furthermore, it facilitates the discovery of rare genetic variants inherited within families (Bailey-Wilson & Wilson, 2011), and novel methods are being proposed to leverage pedigree information within whole-genome sequencing data (Ott et al., 2015). Large-scale pedigree analysis is also essential in studies of isolated human populations, such as those conducted in Quebec, where extensive and complex family structures are used to elucidate genetic structure (Roy-Gagnon et al., 2011). In addition, pedigree analysis allows for better management of animal breeding practices, resulting in healthier animal populations, and remains a fundamental component of genetic counseling in human medicine (Alliance & Collaborative, 2010).

Pedigrees are crucial in various fields, including medical and forensic genetics, animal breeding, and ecology. Researchers in these areas need tools for visualizing complex genealogical relationships,

managing large datasets, and calculating kinship coefficients.

Since the early 2000s, more than 25 software programs (e.g. *Haploforge* (Tekman et al., 2017), *HaploPainter* (Thiele & Nürnberg, 2005), *Pelican* (Dudbridge et al., 2004)) have been created to visualize and manage pedigrees. However, most are now deprecated (see Table 4, available as supplementary data at *Bioinformatics* online) even if some software packages remain actively maintained. Standalone software includes *Madeline2.0 PDE* (Trager et al., 2007), *Cranefoot3.2* (Mäkinen et al., 2005), *DrawPed* (Schönberger et al., 2024), *pedigreejs* (Carver et al., 2018), and *OpenPedigree* (Buske et al., 2019). Additionally, package-based solutions include *QuickPed* (Vigeland, 2022), which is built on *kinship2* (Sinnwell et al., 2014). Each program has unique features but adheres to the recommendations of the Pedigree Standardization Task Force (PSTF) (Bennett et al., 2008) to varying degrees. The PSTF was created to create uniform graphical and semantic conventions for pedigree representation in clinical and research settings. Their guidelines are necessary to avoid discrepancies in layout, symbol usage, or metadata that could lead to confusion and hinder collaboration.

One of the main challenges in pedigree visualizations is to properly align individuals. Existing software often struggles with intricate structures such as cross-generational matings, multiple complex and/or repeated loops, monozygotic

twins, as well as large litters and extremely large pedigrees. *Madeline 2.0*, *DrawPed*, and *pedigreejs* struggle with cross-generational matings, and their alignment is disrupted by consanguinity. In contrast, the alignment algorithms of *kinship2* and *OpenPedigree* remain accurate in these cases. *QuickPed* requires extensive manual data entry for each individual, making it impractical for large-scale pedigrees. As for *OpenPedigree*, it can display only one affection status at a time and does not support visualization of twins nor spouses without children. These limitations highlight the need for a more flexible and user-friendly alternative capable of accommodating these complex scenarios.

The *kinship2* package (Sinnwell et al., 2014) written in the R programming language (Team, 2025) was a significant advancement in pedigree alignment and visualization. Originally developed to complement the *coxme* package for mixed-effects Cox models (Therneau & Grambsch, 2000), *kinship2* was the first Comprehensive R Archive Network (CRAN) (Team, 2025) package available for pedigree drawing, and is used by other packages such as *FamAgg* (Rainer et al., 2016), *Familias* (Simonsson & Mostad, 2016), *QuickPed* (Vigeland, 2022), or *GESE* (Qiao et al., 2017) for the pedigrees' alignment and drawing. However, its maintenance has ceased, and the landscape of genetic epidemiology demands a modern, robust, and scalable solution to meet the needs of complex, large-scale pedigree studies.

To address this gap, we introduce *Pedixplorer*, an up-to-date R package designed for handling complex pedigrees, released on the open-source Bioconductor platform (Gentleman et al., 2004). R is widely used for statistical computing and bioinformatics, with CRAN serving as its primary repository for general-purpose packages, while Bioconductor specializes in bioinformatics and computational biology tools.

Pedixplorer builds upon the existing codebase of *kinship2*, refactoring it to work seamlessly within R's S4 object-oriented framework and aligning it with current Bioconductor standards. It provides a comprehensive suite of tools for pedigree creation, validation, visualization, and analysis, ensuring compatibility with Bioconductor's bioinformatics ecosystem. Moreover, *Pedixplorer* adheres to Findable, Accessible, Interoperable, and Reusable (FAIR) data principles, promoting reproducibility and accessibility in genetic research.

In addition to all the previous functions available in *kinship2*, *Pedixplorer* offers several new features tailored to current genetic research :

- Enhanced data normalization and validation workflows to ensure pedigree integrity ;
- Comprehensive pedigree visualization with customizable parameters ;
- Informative metrics from proband selection to automatically refine large pedigrees ;

- Comprehensive *shiny* (Chang et al., 2024) web application for inexperienced R users.

In this article, we outline *Pedixplorer*'s core functionalities, its applications in genetic research and breeding problems, and discuss the ways in which it addresses the limitations of existing tools. *Pedixplorer* aims to streamline pedigree-based workflows and facilitate new discoveries in genetic epidemiology, population genetics, and molecular genetics by empowering users with an open and flexible solution.

3 Results

3.1 Enhanced visualization and customization

The main graphical functions have been rewritten and restructured into a two-step process :

1. Data preparation : Converts a Pedigree object into a structured data frame containing all graphical elements needed for plotting.
2. Plot generation : Uses the prepared data frame, along with user-defined settings, to render the pedigree visualization.

This modular approach simplifies debugging by allowing users to inspect and modify the intermediate data frame prior to rendering the final plot, making it easier to identify and resolve issues in the visualization process. Furthermore, it enhances flexibility for advanced users, who can fine-tune the pedigree representation by adjusting colours, text

sizes, or repositioning labels to highlight specific individuals or improve readability according to their needs.

Additional enhancements include :

- Gradient colouring : using the `generate_colors()` function allows the user to specify the continuous trait of an affection to accurately represent quantitative data by colour gradients ;
- Interactive plots : the pedigree can additionally be drawn using the *ggplot2* package (Wickham, 2016). Using the *plotly* package (Sievert, 2020) allows to zoom and to display information by hovering over specific individuals ;
- Improved support for complex layouts : in complex pedigrees where individuals span multiple generations, the tool can now bypass spouse alignment to allow the pedigree rendering. A comparative example is provided in Figure S1, available as supplementary data at *Bioinformatics* online.

To contextualize *Pedixplorer*'s strengths, we compared its features with those of existing pedigree visualization tools (see Table S2, available as supplementary data at *Bioinformatics* online). While several tools support basic pedigree drawing, *Pedixplorer* is the only one able to support the 17 listed features of the PSTF. Notably, only 6 out of 16 tools support multiple affection status visualization, and none offer gradient-based coloring, a key feature for representing quantitative traits or risk scores.

3.2 Shiny web application

To facilitate usage for non-programmers, *Pedexplorer* includes a *shiny* (Chang et al., 2024) application accessible via `ped_shiny()` or via a dedicated web server hosted by the University of Rennes (<https://pedexplorer.univ-rennes.fr/>). This application has a user-friendly design and visualization of the pedigrees. The application is built using a modular design where each component operates as a stand-alone *shiny* module, communicating with other modules through reactive objects. This modular approach provides several key advantages : it simplifies maintenance, promotes code reusability and enables separate documentation and unit testing for each module. The main components of the application include :

- Data Upload : Handles importing pedigree datasets in various formats (e.g., csv, tsv, xlsx, rda) ;
- Health and Family Selection : Allows users to select a family and to choose which health condition to use as affection status ;
- Filtering : Provides dynamic filtering options to subset pedigree data ;
- Plot Customization and Rendering : Enables users to fine-tune visual elements like colours and information to display upon hovering over an individual before generating the final pedigree plot ;
- Data and Plot Download : Facilitates the export of processed data and the customization of plots in different formats.

This modular design approach ensures that each component can be developed, tested, and documented independently. This enhances the maintainability and adaptability of the codebase to future improvements. The use of reactive objects for communication between modules facilitates the propagation of modifications across modules while reducing the computational cost by only processing the new data when necessary. This structure also makes it easier to extend the application with new features and to integrate it with other tools, thereby providing a flexible and robust solution for pedigree design and analysis.

3.3 Refining of large pedigrees

Handling large pedigrees can be cumbersome, making it necessary to extract and focus on relevant subsets for clearer visualization. *Pedexplorer* introduces the new `useful_inds()` function, which identifies informative individuals based on specified probands and a maximum kinship-based distance. The distance $D_{i,j}$ is derived from the kinship coefficient $K_{i,j}$ using the formula :

$$D_{i,j} = \log_2(1/K_{i,j}) \quad (II.1)$$

This transformation makes kinship values more intuitive by representing them as discrete steps in the pedigree. For example, siblings (with $K = 0.25$) have a distance of $D = 2$, while cousins (with $K = 0.0625$) have a distance of $D = 4$. By setting a threshold on D , users can dynamically filter pedigrees to retain only

individuals closely related to the probands, effectively reducing visual complexity while preserving key familial relationships. Parents who are needed to keep the structure of the pedigree will also be kept. This metric allows researchers to focus on the most relevant portion of the pedigree without manually tracing connections. Consequently, the interpretation of large and complex family trees becomes faster, easier, and more straightforward.

3.4 Pedigree S4 Object

One of the major improvements in the *Pedixplorer* package is the handling of pedigree data. In contrast to the *kinship2* management of pedigree data through an S3 object, our implementation utilizes a structured S4 object. S3 objects, commonly used in base R, offer a lightweight and flexible approach, relying on generic functions and implicit class structures. However, this flexibility comes at the cost of weaker validation and less formalized data integrity, which can lead to inconsistencies and harder-to-track modifications. In contrast, the formal definition of S4 classes ensures that data structures are rigorously defined, reducing the likelihood of errors and making interactions more predictable for users. Additionally, S4 objects support inheritance, allowing other packages to extend the Pedigree class and integrate seamlessly with *Pedixplorer*. This design choice aligns with Bioconductor standards, promoting interoperability within the bioinformatics community and facilitating collaborative developments.

Although S4 objects require more careful planning and upfront design, this investment ultimately results in code that is unambiguous, easily maintained, and consistent throughout the package. The Pedigree object contains four slots, each storing a different S4 object for a specific type of information essential for the pedigree construction (details are available in Figure S2, available as supplementary data at Bioinformatics online).

- `ped` (Ped object) : Stores basic pedigree information (i.e., identity, disease, proband, other information);
- `rel` (Rel object) : Describes special relationships that cannot be captured in the `ped` slot (i.e., twins, spouse with no child);
- `scales` (Scales object) : Provides plotting information (i.e., filling and bordering of the symbols);
- `hints` (Hints object) : Provides horizontal ordering between spouses and siblings.

This structure ensures data integrity by allowing validation rules and interconnection between the different objects. For example, all individuals referenced in the `rel` object must exist in the `ped` object, and monozygotic twins must share the same sex.

To facilitate the use of this new data structure and to ensure a smooth transition from *kinship2*, multiple constructors have been added and are interoperable with the previous *kinship2* version. Standardization steps have also been incorporated

to automatically detect and correct errors, such as the mis-gendering of a parent or loop detection when one individual is its own ancestor. The latter was not checked in *kinship2* and the consequent errors were complicated to understand and identify.

3.5 Kinship matrix analysis

Pedixplorer also provides functionality for computing kinship matrices. The `kinship()` function from *kinship2* based on (Lange, 1997) has been rewritten as an S4 generic method, allowing it to operate directly on Pedigree objects while accounting for monozygotic twins present in the `Rel` object. To maintain backward compatibility, it also supports input as a `Ped` object or a sequence of character vectors. Kinship analysis can be performed for autosomal chromosomes or restricted to sex chromosomes, providing flexibility for different genetic studies. The output is a sparse block-diagonal matrix, structured by family units to optimize storage and computation.

3.6 Documentation and unit-testing

All *kinship2* package documentation and vignettes have been reviewed, improved, and expanded. Function documentation has been rewritten using the *roxygen2* package (Wickham et al., 2024), enabling in-place documentation and simplifying updates. The revised documentation is now accessible via a dedicated website (<https://louislenezet.github.io/Pedixplorer/>), facilitating user navigation and retrieval of information.

Testing is now implemented using the *test-th* (Wickham, 2011) and *shinytest2* packages (Schloerke, 2024), achieving over 90 % code coverage. This robust testing framework ensures better tracking of modifications and enhances code stability.

4 Discussion

We developed the *Pedixplorer* package in response to the lack of accessible, user-friendly pedigree software and the lack of maintenance of the *kinship2* package. In accordance with the Bioconductor community guidelines, *Pedixplorer* is based on the *kinship2* package, which has been thoroughly revisited and modernized using S4 objects and a *shiny* application to design and visualize complex pedigrees. This architecture offers a robust foundation for future extensions that bridge pedigree visualization with genomic analysis. A natural evolution of the package would involve interoperability with tools that provide genetic data annotations, enabling users to overlay key genetic information, such as carrier status for pathogenic variants, allelic segregation patterns, or polygenic risk scores directly onto pedigree plots.

Pedixplorer's intuitive interface and extensive documentation make it suitable for both novice and experienced genetic researchers. The inclusion of an *R shiny* application further enhances usability by providing a graphical interface for users who may not be familiar with *R* programming language. This application can be accessed either locally or via a dedicated web server, thus offering

flexibility in terms of usage. By providing *Pedixplorer*, we aim to foster new collaborations. While the package has primarily been used within our genetic research teams, we invite feedback from the community to further refine and improve its use and functionality. Contributions and feedback can be shared via the dedicated GitHub repository (<https://github.com/LouisLeNezet/Pedixplorer>).

To ensure the long-term maintenance of *Pedixplorer*, the first author is committed to actively developing and updating the package, addressing issues and feature requests through the *GitHub* issue page. Additionally, the web server hosting the *shiny* application is expected to remain operational, as there are no anticipated issues with the maintenance of the virtual machine at the University of Rennes.

5 Acknowledgements

The authors are grateful to the CaniDNA BRC, hosted at Biosit (<https://biosit.univ-rennes.fr/centre-de-ressources-biologiques-crb-cani-dna>), for providing genealogies of dogs in the frame of several genetic research projects, as well as to the GenOuest bioinformatics core facility (<https://www.genouest.org>) for providing the computing infrastructure.

L.L.N. did the experimental work, by coding and testing the package, managed the collaboration with J.S., and wrote the draft paper. P.Q. managed the whole project, reviewed and edited the manuscript. C.A. reviewed and edited the manuscript. J.S. and A.L. provided feedback on code improvements, tested the package, and

contributed to the review and editing of the manuscript.

6 Funding

This work was performed in the frame of L.L.N.'s PhD research, as a part of a project on the comparative genetics of developmental genetic diseases in dogs and humans, funded by the « Fondation pour la Recherche Médicale » and the French Guide Dogs Association (Association les Chiens Guides d'Aveugles de l'Ouest (AC-GAO)) as well as the French foundation Visio, which supports visually impaired children and adults.

7 Competing interests

No competing interest is declared.

8 Author contributions statement

Louis Le Nézet (Conceptualization [lead], data curation [equal], formal analysis [lead], investigation [lead], methodology [lead], software [lead], validation [equal], visualization [lead], writing—original draft [lead], writing—review & editing [equal]), Jason P. Sinnwell (conceptualization [equal], formal analysis [equal], methodology [equal], software [equal], validation [equal], writing—review & editing [equal]), Anna Letko (investigation [equal], methodology [equal], software [equal], validation [equal], writing—review & editing [equal]), Catherine Andrée (funding acquisition [lead], project administration [equal], resources [equal], supervision [equal], writing—review & editing [equal]), and Pascale Quignon (funding acquisition [equal], project administration [equal],

resources [supporting], supervision [lead], validation [equal], visualization [supporting], writing—original draft [supporting], writing—review & editing [equal])

9 Bibliography

- ALLIANCE, G. & COLLABORATIVE, N., 2010. Pedigree and Family History Taking. In: Understanding Genetics: A New England Guide for Patients and Health Professionals [en ligne]. Washington USA: Genetic Alliance, p. 15–19 [visité le 2025-02-04].
- BAILEY-WILSON, J. E. & WILSON, A. F., 2011. Linkage Analysis in the Next-Generation Sequencing Era. *Human Heredity* [en ligne]. Vol. 72, no. 4, p. 228–236 [visité le 2025-05-22]. ISSN 0001-5652. Disp. à l'adr. DOI: 10.1159/000334381.
- BENNETT, R. L. et al., 2008. Standardized Human Pedigree Nomenclature: Update and Assessment of the Recommendations of the National Society of Genetic Counselors. *Journal of Genetic Counseling* [en ligne]. Vol. 17, no. 5, p. 424–433 [visité le 2024-08-26]. ISSN 1573-3599. Disp. à l'adr. DOI: 10.1007/s10897-008-9169-9.
- BUSKE, O. et al., 2019. Phenotips/Open-Pedigree [en ligne]. PhenoTips. [visité le 2025-04-16]. Disp. à l'adr.: <https://deepwiki.com/phenotips/open-pedigree>.
- CARVER, T. et al., 2018. Pedigreejs: A Web-Based Graphical Pedigree Editor. *Bioinformatics* [en ligne]. Vol. 34, no. 6, p. 1069–1071 [visité le 2025-02-06]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/btx705.
- CHANG, W. et al., 2024. Shiny: Web Application Framework for R. Version 1.10.0. Aussi disponible à l'adresse: <https://shiny.posit.co/>.
- DUDBRIDGE, F. et al., 2004. Pelican: Pedigree Editor for Linkage Computer Analysis. *Bioinformatics* (Oxford, England). Vol. 20, no. 14, p. 2327–2328. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/bth231.
- GENTLEMAN, R. C. et al., 2004. Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biology* [en ligne]. Vol. 5, no. 10, R80 [visité le 2025-05-23]. ISSN 1474-760X. Disp. à l'adr. DOI: 10.1186/gb-2004-5-10-r80.
- LANGE, K., 1997. Mathematical and Statistical Methods for Genetic Analysis [en ligne]. New York : Springer [visité le 2025-04-16]. ISBN 978-0-387-94909-3.
- MÄKINEN, V.-P. et al., 2005. High-Throughput Pedigree Drawing. *European Journal of Human Genetics* [en ligne]. Vol. 13, no. 8, p. 987–989 [visité le 2024-08-26]. ISSN 1476-5438. Disp. à l'adr. DOI: 10.1038/sj.ejhg.5201430.
- OTT, J. et al., 2015. Genetic Linkage Analysis in the Age of Whole-Genome Sequencing. *Nature reviews. Genetics* [en ligne]. Vol. 16, no. 5, p. 275–284 [visité le 2024-12-04]. ISSN 1471-0056. Disp. à l'adr. DOI: 10.1038/nrg3908.
- QIAO, D. et al., 2017. Gene-Based Segregation Method for Identifying Rare Variants in Family-Based Sequencing Studies. *Genetic Epidemiology* [en ligne]. Vol. 41, no. 4, p. 309–319 [visité le 2024-12-09]. ISSN 1098-2272. Disp. à l'adr. DOI: 10.1002/gepi.22037.
- RAINER, J. et al., 2016. FamAgg: An R Package to Evaluate Familial Aggregation of Traits in Large Pedigrees. *Bioinformatics* [en ligne]. Vol. 32, no. 10, p. 1583–1585 [visité le 2024-12-09]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/btw019.
- ROY-GAGNON, M.-H. et al., 2011. Genomic and Genealogical Investigation of the French Canadian Founder Population Structure. *Human Genetics* [en ligne]. Vol. 129, no. 5, p. 521–531 [visité le 2025-04-23]. ISSN 1432-1203. Disp. à l'adr. DOI: 10.1007/s00439-010-0945-x.
- SCHLOERKE, B., 2024. Shinytest2: Testing for Shiny Applications. Version 0.3.2.9000. Aussi disponible à l'adresse: <https://rstudio.github.io/shinytest2/>.

- SCHÖNBERGER, J. et al., 2024. Drawing Human Pedigree Charts with DrawPed. *Nucleic Acids Research* [en ligne]. Vol. 52, no. W1, W61–W64 [visité le 2025-02-06]. ISSN 0305-1048. Disp. à l'adr. DOI: 10.1093/nar/gkae336.
- SIEVERT, C., 2020. Interactive Web-Based Data Visualization with R, Plotly, and Shiny. Boca Raton, FL: CRC Press, Taylor and Francis Group. ISBN 978-0-429-82420-3 978-0-429-44727-3.
- SIMONSSON, I. & MOSTAD, P., 2016. Stationary Mutation Models. *Forensic Science International: Genetics* [en ligne]. Vol. 23, p. 217–225 [visité le 2024-12-09]. ISSN 1872-4973. Disp. à l'adr. DOI: 10.1016/j.fsigen.2016.04.005.
- SINNWELL, J. P. et al., 2014. The Kinship2 R Package for Pedigree Data. *Human heredity* [en ligne]. Vol. 78, no. 2, p. 91–93 [visité le 2024-08-26]. ISSN 0001-5652. Disp. à l'adr. DOI: 10.1159/000363105.
- TEAM, R. C., 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- TEKMAN, M. et al., 2017. HaploForge: A Comprehensive Pedigree Drawing and Haplotype Visualization Web Application. *Bioinformatics* [en ligne]. Vol. 33, no. 24, p. 3871–3877 [visité le 2024-08-26]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/btx510.
- THERNEAU, T. M. & GRAMBSCH, P. M., 2000. Modeling Survival Data: Extending the Cox Model [en ligne]. Red. by DIETZ, K. et al. New York, NY: Springer [visité le 2024-12-09]. Statistics for Biology and Health. ISBN 978-1-4419-3161-0 978-1-4757-3294-8. Disp. à l'adr. DOI: 10.1007/978-1-4757-3294-8.
- THIELE, H. & NÜRNBERG, P., 2005. HaploPainter: A Tool for Drawing Pedigrees with Complex Haplotypes. *Bioinformatics* [en ligne]. Vol. 21, no. 8, p. 1730–1732 [visité le 2024-08-26]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/bth488.
- TRAGER, E. H. et al., 2007. Madeline 2.0 PDE: A New Program for Local and Web-Based Pedigree Drawing. *Bioinformatics* [en ligne]. Vol. 23, no. 14, p. 1854–1856 [visité le 2024-08-26]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/btm242.
- VIGELAND, M. D., 2022. QuickPed: An Online Tool for Drawing Pedigrees and Analysing Relatedness. *BMC Bioinformatics* [en ligne]. Vol. 23, no. 1, p. 220 [visité le 2025-02-06]. ISSN 1471-2105. Disp. à l'adr. DOI: 10.1186/s12859-022-04759-y.
- WICKHAM, H., 2011. Testthat: Get Started with Testing. *The R Journal* [en ligne]. Vol. 3, no. 1, p. 5–10 [visité le 2024-12-13]. ISSN 2073-4859.
- WICKHAM, H., 2016. Ggplot2: Elegant Graphics for Data Analysis. Second edition. Cham: Springer international publishing. Use R! ISBN 978-3-319-24277-4.
- WICKHAM, H. et al., 2024. Roxygen2: In-Line Documentation for R. Version 7.3.2. Aussi disponible à l'adresse: <https://roxygen2.r-lib.org/>.

10 Supplementary data

TABLE S1 - Pedigree software compared by source, license, input formats, and activity.

Name	DOI	Code Source	Info & Web Interface	License	Last Update	Latest Version	System	Code	Input Files	PSTF	Free	Open	Act.	Cat
Cranefoot	10.1038/sj.ejhg.5201430	github.com/caot/CraneFoot	finndiane.fi/software/cranefoot		2019.12.04	3.2	Windows Linux	C++	flat file	yes	yes	yes	yes	1
Madeline	10.1093/bioinformatics/btm242	github.com/piratical/Madeline_2.0_PDE		GPLv2	2023.02.10	2.1.1	Any	C++	flat, XML-based files	yes	yes	yes	yes	1
QuickPed	10.1186/s12859-022-04759-y	github.com/magnusdv/pedsuite	magnusdv.github.io	GPLv3	2025.01.22	4.2.0	Any	R	ped, flat file	yes	yes	yes	yes	1
DrawPed	10.1093/nar/gkae336	git-ext.charite.de/genecascade/drawped	genecascade.org	CC BY-SA 4.0	2024.01.01		Web	HTML perl javascript	ped, drawped	yes	yes	yes	yes	1
pedigreejs	10.1093/bioinformatics/btx705	github.com/CCGE-BOADICEA/pedigreejs	ccge-boadicea.github.io	GPLv3	2022.08.31	2.1.0	Any	javascript	ped, canrisk, pedigree	yes	yes	yes	yes	1
Pedixplorer		github.com/LouisLeNezet/Pedixplorer	louislenezet.github.io/Pedixplorer	GPLv2	2024.10.09	1.1.5	Any	R	flat file	yes	yes	yes	yes	1
Open Pedigree		github.com/phenotips/open-pedigree	phenotips.github.io/open-pedigree	LGPL 2.1	2022.06.30	1.0.0	Any	javascript	ped, GEDCOM, BOADICEA, GA4GH	yes	yes	yes	yes	1

TABLE S1 - Pedigree software compared by source, license, input formats, and activity.

Name	DOI	Code Source	Info & Web Interface	License	Last Update	Latest Version	System	Code	Input Files	PSTF	Free	Open	Act.	Cat
ped_draw	10.1186/s12859-020-03917-4	github.com/mvelinder/ped_draw	peddraw.github.io	MIT	2022.02.10		Any	python	ped	no	yes	yes	yes	2
PEDhunter	10.1186/1471-2350-11-68	ftp.ncbi.nih.gov/pub/agarwala/pedhunter	ncbi.nlm.nih.gov/research/staff/schaffer/pedhunter			2.0	Windows Linux	C	ASCII, Sybase database	no	yes	yes	yes	2
PedGraph		animalgene.umn.edu/download-pedigraph	animalgene.umn.edu/pedigraph	(C) 2004		2.4	Windows		flat file	no	yes	yes	yes	2
HaploForge	10.1093/bioinformatics/btx510	gitlab.com/mtekman/HaploForge	mtekman.github.io/haploforge	GPLv3	2021.02.02		Web	HTML5 javascript	By hand, ped	yes	yes	yes	no	3
Kinship2	10.1159/000363105	github.com/mayoverse/kinship2	github.com/mayoverse/kinship2		2023.06.08	1.9.7	Any	R	flat file	yes	yes	yes	no	3
Pedraw		davecurtis.net/software/dcurtis	davecurtis.net/dcurtis/software.html		2014.09.14	2.5	Windows	C++	ped, ppd, sdf	yes	yes	yes	no	3
HaploPainter	10.1093/bioinformatics/bth488	sourceforge.net/projects/haplopainter	sourceforge.net/projects/haplopainter		2015.10.08	1.043	Windows	Perl	linkage, csv, database, map	yes	yes	yes	no	3
Pelican	10.1093/bioinformatics/bth231	rfcgr.mrc.ac.uk/Software/PELICAN	rfcgr.mrc.ac.uk/Software/PELICAN		2004	1.1.0	Any	java	ped	yes	yes	yes	no	3

TABLE S1 - Pedigree software compared by source, license, input formats, and activity.

Name	DOI	Code Source	Info & Web Interface	License	Last Update	Latest Version	System	Code	Input Files	PSTF	Free	Open	Act.	Cat
Pedimap	10.1093/jhered/ess060	plantbreeding.wur.nl/UK	plantbreeding.wur.nl/UK		2005	1.1	Windows			no	yes	yes	no	3
PediTree	10.1093/jhered/esi059	dpw.wau.nl/pv/pub/Peditree	dpw.wau.nl/pv/pub/peditree		2007					no	yes	yes	no	3
wpdraw		davecurtis.net/software/dcurtis	davecurtis.net/dcurtis/software.html			1.4.1	Windows	C++	ped, ppd, sdf		yes	yes	no	3
Family Genome Browser	10.1093/bioinformatics/btv151		mlg.hit.edu.cn/FGB				Web		ped		yes		no	3
famgenix			famgenix.com							yes	no	no	yes	4
PhenoTips			docs.phenotips.com/docs/pedigree							yes	no	no	yes	4
PPB			progenygenetics.com							yes	no	no	yes	4
Ebrida			agripartner.com/nl/producten/ebrida							no	no	no	yes	4
Cryllic3		apbenson.com/cyrillic-downloads	apbenson.com/about-cyrillic-3			3	Windows			yes	no	no	no	4
PedigreeDraw			pedigreedraw.com		2014					yes	no	no	no	4

TABLE S1 - Pedigree software compared by source, license, input formats, and activity.

Name	DOI	Code Source	Info & Web Interface	License	Last Update	Latest Version	System	Code	Input Files	PSTF	Free	Open	Act.	Cat
PED6		medgen.de/ped	en.bio-soft.net/ other/PED.html		2014.12	6.0.2	Windows				no	no	no	4

Legend:

- **PSTF:** Pedigree Standardization Task Force guidelines
- **Free:** Free of charge
- **Open:** Open source
- **Act.:** Active (updated $\leq 2y$)
- **Cat:** Category
 - 1: Open, active, PSTF
 - 2: Open, active, not PSTF
 - 3: Discontinued
 - 4: Not free
- **PPB:** Progeny Pedigree Builder

Notes:

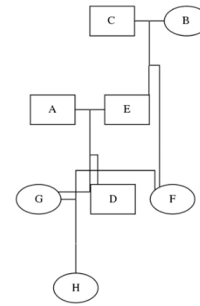
- **QuickPed:** Uses kinship2
- **OpenPedigree:** No documentation
- **PEDhunter 2.0:** *NCBI* tool, standalone, no plotting
- **PedGraph:** Graph output
- **Pelican:** Only .jar via WayBackMachine
- **Pedimap/PediTree:** Not available, plant, graph output
- **Wpdraw:** Like peddraw, not on Win11
- **Family Genome Browser:** Unavailable
- **PhenoTips:** Uses Open Pedigree
- **Ebrida:** Plant, graph output

Ped file used :

```

A A 0 0 1
A B 0 0 2
A C 0 0 1
A D A B 2
A E C B 2
A F C B 1
A G A E 2
A H F G 1

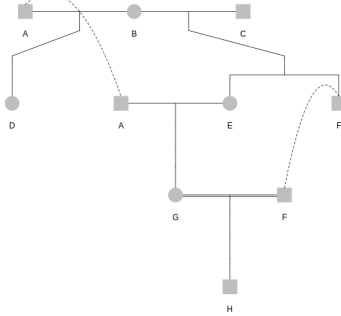
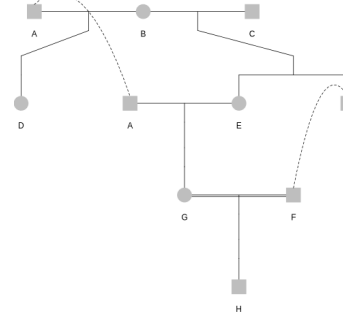
```

Peddraw :**Pedixplorer :**

```

1 library(pedixplorer)
2 # First pedigree with no reordering
3 pedi <- plink_to_pedigree("test.ped")
4 plot(pedi, align_parents = FALSE, title = "1) Without reordering")
5 # Second pedigree with G and F spouse reordered
6 hints(pedi) <- auto_hint(pedi, align_parents = FALSE)
7 spouse(hints(pedi)) <- data.frame(idl = "G", idr = "F", anchor = "left")
8 plot(pedi, align_parents = FALSE, title = "2) With ordering")

```

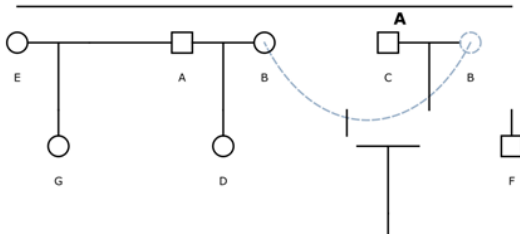
1) Without reordering**2) With ordering****Madeline2.0 :**

Madeline web interface is no more working. The local installation was complex but did work on linux.

```

1 # Launch Madeline2.0 with ped file
2 madeline2 test2.data

```

**Other tested software :**

- **DrawPed** throw « error 64 » ;
- **QuickPed** and **pedigreejs** doesn't allow to use the same individual in multiple relationship. Therefore this pedigree cannot be plot.

FIGURE S1 - Multi-marriage and generational spanning complex pedigree

TABLE S2 - Comparison of pedigree visualization software features. (Part 1/3)

Name	Complexity					Symbols shape					
	Multiple families on same plot	Generation spanning mate	Multiple marriage	Matings with no offspring	Consanguinous	Twins	Unknown sex	Deceased	Miscarriage	Infertility	Adopted
Cranefoot	yes			no		no	yes	yes	no	no	no
DrawPed	no	no	yes	no	yes	no	yes	yes	yes	no	no
QuickPed	no	yes	yes	no	yes	manual	manual	manual	manual	no	no
pedigreejs	no	no	yes	no	yes	yes	yes	yes	yes	yes	yes
OpenPedigree	no	yes	yes	no	yes	no	yes	yes	yes	yes	yes
Madeline	yes	no	yes	yes	sometimes	yes	yes	yes	yes	no	yes
Pedexplorer	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
PedGraph	yes	yes	yes	no	no	no	yes	no	no	no	no
ped_draw	yes	no	no	no	no	no	yes	no	no	no	no
pedraw	no	yes	yes	no	no	no	yes	yes	no	no	no
kinship2	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no
HaploPainter	yes	no	yes	no	manual	yes	yes	yes	yes	no	yes
HaploForge	yes	no	yes	yes	yes	no	yes	no	no	no	no
Pelican	yes	yes	yes	no	no	no	yes	yes	no	no	no
Pedigreedraw	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
PPB	yes	no	no	no	no	yes	yes	yes	yes	yes	yes

TABLE S2 - Comparison of pedigree visualization software features. (Part 2/3)

Name	Additional informations									Affection		Customization	
	Date of birth	Date of death	Proband	Consultand	Carrier	A/Pre symptomatic	Evaluation	Sampled	Haplotype	Multiple affection	Color gradient	Custom font	Custom ordering
Cranefoot	no	no	no	no	no	no	no	no	no	no	no	no	no
DrawPed	no	no	yes	no	no	no	no	no	no	no	no	no	no
QuickPed	no	no	no	no	yes	no	no	no	no	no	no	no	no
pedigreejs	yes	no	no	no	no	no	no	no	no	yes	no	no	no
OpenPedigree	yes	no	yes	yes	yes	yes	yes	no	no	no	no	no	no
Madeline	yes	no	yes	yes	yes	no	no	yes	no	yes	manual	no	no
Pedexplorer	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes
PedGraph	no	no	no	no	no	no	no	no	no	no	no	no	no
ped_draw	no	no	no	no	no	no	no	no	no	no	no	no	no
pedraw	no	no	no	no	no	no	no	no	no	no	no	no	yes
kinship2	yes	no	no	no	no	no	no	yes	no	yes	no	yes	yes
HaploPainter	no	no	yes	no	no	no	no	no	yes	no	no	no	no
HaploForge	no	no	no	no	no	no	no	no	yes	no	no	no	yes
Pelican	no	no	manual	no	no	no	no	manual	yes	no	no	no	yes
Pedigreedraw	yes	no	yes	no	yes	no	no		no	yes			
PPB	no	no	yes	no	no	no	yes	yes	no	yes	no	no	yes

TABLE S2 - Comparison of pedigree visualization software features. (Part 3/3)

Name	Interoperability		Data input / output			Notes
	Command line	Graphical interface	Ped file	Other format	Output file	
Cranefoot	yes	no	no	flat file with header	ps	Postscript file can be generated but no pedigree drawn in
DrawPed	no	yes	yes	ped, drawped	svg, png, ped	Alignment not optimal with consanguinity no difference between unknown status and intact
QuickPed	no	yes	yes	flat file with header	png, pdf	Most characteristics by hand
pedigreejs	no	yes	yes	canrisk ped	svg, png, pdf	Alignment quickly thrown off with complex pedigree
OpenPedigree	no	yes	yes	ped, LINKAGE, GEDCOM, BOADICEA, GA4GH Pedigree	ped, GA4GH Pedigree, pdf, svg	
Madeline	yes	yes	no	tab-delimited, xml, html, ods	svg	Web interface did not work with input file but Docker container did. Inter-generational marriage wrongly represented
Pedexplorer	yes	yes	yes	flat file with header	image, flat file	
PedGraph	yes	no	no	tab-delimited	ps, png, jpg	Graph based output
ped_draw	yes	yes	yes	ped	png, svg, pdf	Alignment quickly thrown off with complex pedigree
pedraw	yes	yes	yes	sdf, ppd, dat, GEDCOM	dat, ps, bmp, xml	Example data works and manual addition to
kinship2	yes	no	yes	flat file with header	image, flat file	
HaploPainter	no	yes	no	LINKAGE, csv, database, map	pdf, svg, png, csv, ps	
HaploForge	no	yes	no	no	not working	No file import possible
Pelican	no	yes	yes	LINKAGE, ped, slink	png, jpeg, ps	Need to realign manually
Pedigreedraw		yes			png, svg, pdf	
PPB	no	yes	no	no	png	Relationship quite limited

Notes

- **Cat:** Category
 - **1** Open active and follow PSTF
 - **2** Open and active but do not follow PSTF guidelines
 - **3** Discontinued
 - **4** Not free
- **PSTF:** Respect the Pedigree Standardization Task Force guidelines
- **PPB:** Progeny Pedigree Builder

- **ped (Ped object)** : Stores basic pedigree information with the following slots :
 - **Identifiers** : id, dadid, momid, famid (i.e., individual, his father, mother and family)
 - **Information influencing the shape of the symbols** :
 - sex (“male” < “female” < “unknown”)
 - miscarriage (TOP = Termination of Pregnancy, SAB = Spontaneous Abortion, ECT = Ectopic Pregnancy, FALSE = no miscarriage)
 - **Information influencing the colors of the symbols** :
 - avail : availability status of the individual (TRUE, FALSE, NA)
 - affected (TRUE, FALSE, NA)
 - **Additional information plotted next or upon the symbols** :
 - evaluated : is documented evaluation of the individual available (TRUE, FALSE)
 - consultand : is the individual seeking genetic counseling/testing (TRUE, FALSE)
 - proband : is the individual an affected family member coming to medical attention independent of other family members (TRUE, FALSE)
 - carrier : does the individual has the genetic trait but is not likely to manifest the disease regardless of inheritance pattern (TRUE, FALSE, NA)
 - asymptomatic : is the individual clinically unaffected at this time but could later exhibit symptoms (TRUE, FALSE, NA)
 - fertility : is the individual fertile, infertile or infertile by choice or unknown reason : infertile_-choice_na
 - deceased (TRUE, FALSE)
 - adopted (TRUE, FALSE)
 - dateofbirth
 - dateofdeath
- **rel (Rel object)** : Describes special relationships that cannot be captured in the ped slot, with the following slots :
 - id1, id2 : identifier of the two individuals
 - code : relationship type MZ twin (monozygotic twins) < DZ twin (dizygotic twins) < UZ twin (unknown zygosity twins) < Spouse (spouse with no child)
- **scales (Scales object)** : Provides plotting information, including :
 - fill : data frame specifying how each status (e.g., affected) present in the Ped object should be visually represented in the pedigree. This includes colors, angles, and density patterns used for shading the individual's box.
 - border : data frame specifying the visual attributes of the box border (e.g., to indicate sample availability)
- **hints (Hints object)** : Provides information for the alignment of the individuals :
 - horder : numeric vector to order individuals horizontally
 - spouse : matrix to horizontally order spouses

FIGURE S2 - Pedigree S4 Object structure

DÉVELOPPEMENT D'UN PIPELINE D'IMPUTATION : *NF-CORE/PHASEIMPUTE*

C Développement d'un pipeline d'imputation : *nf-core/phaseimpute*

Dans le cadre de l'étude des bases génétiques de la dysplasie coxo-fémorale chez le chien (DCF), nous avons choisi de recourir à une stratégie de séquençage à faible couverture, dite "*Low-Pass Whole Genome Sequencing*" (LP-WGS). Cette approche, généralement conduite à des profondeurs comprises entre 0,1× et 2×, constitue une alternative économique aux méthodes classiques de séquençage haut débit ou de génotypage par puce SNP, tout en permettant une couverture homogène du génome. Si une telle profondeur ne permet pas d'appeler directement les génotypes avec une confiance élevée, l'association de techniques d'alignement des haplotypes et d'imputation permet de reconstruire les génotypes manquants avec une précision pouvant dépasser les 95 % pour les variants communs (Buckley et al., 2021).

L'un des principaux atouts de cette stratégie réside dans la possibilité de réduire considérablement les coûts par échantillon, ce qui permet d'envisager l'analyse de cohortes de plus grande taille que dans la stratégie de séquençage WGS à 30×. Étudier un grand nombre d'individus est un facteur crucial pour garantir la puissance statistique des études d'association pangénomiques (GWAS) dans le cadre de maladies complexes tel que la DCF. Par ailleurs, le LP-WGS est compatible avec des échantillons présentant une faible concentration d'ADN ou une qualité altérée. En outre, cette approche permet une détection plus exhaustive de la variation génétique, y compris les variants rares ou structuraux qui échappent aux technologies de génotypage standard.

Des études récentes menées chez les animaux de compagnie, notamment chez le chien (Buckley et al., 2021), ont confirmé l'efficacité du couplage LP-WGS et imputation pour la reconstitution haplotypique et l'augmentation de la puissance statistique des GWAS. Ces résultats ont été obtenus même en l'absence de panels de référence de très grande taille, tant que la diversité allélique de la population d'étude est conservée. En revanche, l'imputation des variants rares, définis par une fréquence de l'allèle mineur (MAF) inférieure à 1 %, demeure délicate. À de faibles profondeurs, proches de 1×, la précision de leur imputation peut se révéler insuffisante pour certaines analyses, cependant dans le cadre de GWAS, ces variants sont bien souvent filtrés par un seuil de $MAF \leq 5\%$ et donc ce biais n'impacte pas vraiment les analyses GWAS (Wragg et al., 2024).

Dans le cadre de mon travail de thèse, l'objectif était de séquençer en LP-WGS un total de 100 chiens atteints et 100 chiens indemnes par race et ceci pour 5 races. Nous avons fait réaliser ce séquençage en 3 séries d'échantillons par la société Psomagen (USA). Pour la phase d'imputation, dans un premier temps, nous avons confié l'imputation des données du premier lot d'échantillons, début 2023 à la société *Gencove*, principal prestataire proposant ce service à grande échelle. Toutefois, nous avons rapidement

identifié plusieurs limites. D'une part, *Gencove* ne proposait pas l'imputation sur la dernière version du génome canin, *canFam4* et exigeait un surcoût de plus de 20 000 dollars pour intégrer un panel de référence pourtant déjà disponible dans la communauté scientifique (Meadows et al., 2023). Par ailleurs, les étapes précises du processus, les filtres appliqués aux variants et les paramètres d'imputation utilisés n'étaient ni détaillés ni accessibles, rendant la compréhension et la reproductibilité des analyses difficile, voire impossible.

Ces limitations nous ont conduits à développer une solution ouverte et reproductible. Avec les conseils du Dr Antony Herzig (laboratoire INSERM GGB à Brest), puis en collaboration avec la bio-informaticienne argentine Anabella Trigilla (*ZS discovery*, Buenos Aires), confrontée à des difficultés similaires avec *Gencove*, j'ai conçu un pipeline d'imputation génétique au sein de la communauté *nf-core*, baptisé *nf-core/phaseimpute* (voir figure 16). Ce pipeline a été conçu pour répondre aux contraintes spécifiques posées par les données génétiques de faible densité, pour le moment focalisé sur les données de séquençage à faible couverture.

Une version stable est désormais disponible via le dépôt *nf-core* à l'adresse suivante : <https://github.com/nf-core/phaseimpute> . Le détail de son utilisation et des paramètres d'entrées est visible sur le site de *nf-core* à l'adresse : <https://nf-co.re/phaseimpute/1.0.0/> .



FIGURE 16 – Logo du pipeline *nf-core/phaseimpute*

Le pipeline couvre l'ensemble des étapes, depuis l'alignement des haplotypes du panel de référence jusqu'à l'imputation proprement dite, tout en intégrant des modules de simulation et de validation de la qualité d'imputation. Il intègre plusieurs logiciels d'imputation parmi les plus performants actuellement disponibles, notamment les versions 1 et 2 de *Glimpse*, ainsi que *Stitch* et *Quilt* (Davies et al., 2016 ; Davies et al., 2021 ; Rubinacci et al., 2023). Ce pipeline met à disposition des utilisateurs l'ensemble des paramètres critiques de l'imputation, tels que la taille des fenêtres d'analyse, la taille effective de la population ou encore le nombre d'itérations, permettant d'optimiser les performances, en particulier dans le cas d'espèces non modèles comme le chien.

Un soin particulier a été apporté à la modularisation des différentes étapes et à la compatibilité avec différents formats et types de données, cela afin d'être adaptable à une plus grande variété de projet de recherche. Conçu dans le respect des principes FAIR (i.e., *Findable, Accessible, Interoperable & Reusable*) le pipeline garantit une standardisation rigoureuse, une traçabilité complète des analyses et une facilité de partage entre chercheurs. Par cette initiative, nous souhaitons contribuer à la démocratisation de l'imputation génétique dans les études de génétique vétérinaire, en fournissant à la communauté scientifique un outil robuste, pérenne et totalement *open-source*, inscrit dans une dynamique collaborative et transparente.

ARTICLE 4

“ nf-core/phaseimpute: A pipeline for genetic imputation ”

AUTEURS : Louis Le Nézet, Anabella Trigila, communauté *nf-core*, Thomas Derrien, Catherine André, Anthony Herzig, Pascale Quignon

STATUT : En préparation et en relecture par les co-auteurs

JOURNAL : Publication en tant que *proceeding* à JOBIM 2024 et prévu en tant qu'article dans *Bioinformatics*

nf-core/phaseimpute: A pipeline for genetic imputation

Louis Le Nézet^{1,+,*}, Anabella Trigila^{2,+,*}, *nf-core* community, Catherine André¹, Anthony Herzig³, Pascale Quignon¹

¹ UMR-6290, IGDR (Institute Genetics and Development of Rennes), University of Rennes, CNRS, Rennes, France

² ZS Discovery, Buenos Aires, Argentina

³ Inserm UMR-1078, CHRU Brest, University of Brest, Brest, France

⁺ These authors contributed equally to this work.

^{*} Corresponding author. louislenezet@gmail.com anatrighila@gmail.com

1 Abstract

Motivation: Genotype and low-coverage sequencing data provide cost-effective avenues for genomic research, but they inherently exhibit the limitations of being of low resolution and low quality, respectively. This sparsity in genetic data poses challenges, particularly in non-model organisms lacking species-specific phased panels, necessitating accurate haplotype phasing for effective downstream analyses. Genetic imputation serves as a valuable tool to supplement these methods by filling in missing data and assigning probabilities to variant observations, thereby enhancing data resolution and quality. However, ensuring the reliability of imputed genotypes requires rigorous tool validation and thorough exploration of parameter impacts (e.g., sliding window size, effective diploid population size, number of burn-in iterations), critical for maximising the utility of these techniques in genomic studies.

Results: To tackle these challenges, we introduce an *nf-core* compliant pipeline tailored for phasing, imputation, and validation of the imputed data. Our pipeline equips users with advanced genomic analysis tools for phasing and imputation analysis. This enables them to leverage the full potential of their genetic data by filling in missing information, harmonising datasets, and enhancing the resolution of genetic analyses. By adhering to *nf-core* guidelines, our pipeline ensures users have access to a suite of high-standard, versioned, and rigorously tested up-to-date tools, developed within a large and helpful community - a crucial aspect for FAIR (i.e., Findable, Accessible, Interoperable, and Reusable) analyses. Our specialised *nf-core* pipeline offers a complete genomic imputation solution, integrating phasing, target file imputation, and quality validation. This pipeline employs advanced tools to provide reliable imputation outcomes for both model and non-model organisms. Additionally, by adhering to FAIR principles, it fosters standardised, reproducible, and community-supported genomic analyses.

Availability: freely available at: <https://github.com/nf-core/phaseimpute> .

Keywords New Generation Sequencing (NGS), Genetic imputation, Nextflow, *nf-core*

2 Introduction

Sparse genetic data pose both opportunities and obstacles in genomic research. It offers a cost-effective means to study large populations, facilitating broader investigations into genetic diversity and disease susceptibility (Li et al., 2021; Mazzonetto et al., 2024). Genotyping microarray has historically been the primary tool for large genome-wide association studies (Grant & Hakonarson, 2008). Variants studied using those microarrays are limited (i.e., from 240 K to 4 M variants, Verlouw et al. (2021)), but the confidence in the genotype observed is high. On the other hand, low-coverage Whole Genome Sequencing (lc-WGS) has recently emerged as a viable low-cost alternative (Chat et al., 2022; Pasaniuc et al., 2012). However, the increase in the number of variants observed comes at the cost of a reduced confidence in their genotypes. A low resolution as well as poor genotype quality can hinder accurate analysis, particularly for rare variants (Lou et al., 2021).

To overcome these challenges, phasing and imputation techniques are crucial. Phasing determines the parental origin of alleles, reconstructing an individual's haplotypes. This allows for accurate prediction of missing genotypes (imputation) by leveraging correlation patterns between known and unknown variants within haplotypes (Marchini & Howie, 2010). Phasing can be performed *de novo* with high-coverage data or by using an already phased reference panel from multiple individuals (Delaneau et

al., 2008). Importantly, for low-pass sequencing data, phasing and imputation are often not completely separate steps; tools such as Glimpse perform both processes (Rubinacci et al., 2021).

However, using a phased panel is not always feasible for non-model species, and the phasing and pre-processing steps are complicated by the lack of clear best practices. The variability in data quality and format, combined with numerous phasing algorithms and parameters, makes applying phasing techniques difficult. Furthermore, the absence of standardised quality control guidelines and the need for scalability in handling large datasets add further complexity to this step.

Imputation leverages known haplotype information to predict missing genotypes, effectively filling in the gaps in sparse genetic datasets. These techniques, often based on probabilistic approaches using Hidden Markov Models (Marino et al., 2022), provide likelihood scores for imputed variants, necessitating consideration of these posterior probabilities in subsequent analyses (Lou et al., 2021). To address the limitations of genotype likelihood, researchers can adopt a hybrid approach, where a subset of individuals is sequenced at high coverage to ensure reliable genotype data for specific analyses. Meanwhile, the remaining individuals can be sequenced at low coverage, providing supplementary data for comprehensive analysis. This strategy allows for the acquisition of both high-quality genotype information and broader coverage across the sample population.

Implementing phasing and imputation workflows presents unique complexities. Researchers face the challenge of navigating multiple software tools, each with its own dependencies and compatibility requirements. Popular phasing tools include *Shapeit5* (Hofmeister et al., 2023) and *PHASE2* (Stephens & Scheet, 2005) but also *wphase* (Fearnhead & Donnelly, 2001), *HAP2* (Lin et al., 2004), *PL-EM/triple* (Niu et al., 2002) or *Eagle2* (Loh et al., 2016). Imputation software use will depend on the type of input data. For genotyping microarray data, the *Beagle5* (Browning et al., 2018) software is widely used by the community, as well as *Impute5* (Rubinacci et al., 2020) and *Minimac4* (Das et al., 2016); whereas for low-coverage whole genome sequences, data is processed with an available phased panel with *Beagle5* (Browning et al., 2018), *Quilt* (Davies et al., 2021), *Glimpse2* (Rubinacci et al., 2023); or without one with: *Stitch* (Davies et al., 2016). This choice complexity, added to an absence of standardised environments, complicates the installation, integration, and reproducibility of analysis pipelines. Moreover, some of these pieces of software are not user-friendly, and the lack of documentation, the high computational loads, the susceptibility to batch artifacts, and the limited availability of phased panels (Lou et al., 2021) hinder the adoption of phasing and imputation techniques in genomic research, highlighting the need for robust, user-friendly solutions.

Nextflow, a workflow language implemented in *Groovy* designed for scalable and reproducible scientific workflows, emerges as a promising solution to address these challenges. It uses a dataflow programming model that connects process nodes through channels, allowing for easy parallelisation of huge data (Di Tommaso et al., 2017).

With built-in support for containerisation technologies like *Docker*, *Conda*, and *Singularity*, *Nextflow* streamlines the development and deployment of complex genomic workflows. Its compatibility with popular cloud computing platforms allows researchers to leverage scalable computing resources for large-scale genomic analyses. Nonetheless, challenges remain, including the relatively steep learning curve for new users and the difficulty of interpreting error messages in some cases. Additionally, the absence of a central repository of tools, reference vocabulary, easy visualization of the data flow, and a GUI has been described as a major limitation (Cohen-Boulakia et al., 2017). To overcome these drawbacks, the *nf-core* community has made significant progress in recent years (Ewels et al., 2020).

The *nf-core* community plays a pivotal role in advancing genomic analyses through standardised tools, processes, and workflows written in *Nextflow* through collaborative efforts. By adhering to *nf-core* guidelines, researchers benefit from validated, versioned, and community-supported tools and components, ensuring the reliability and reproducibility of

genomic analyses (Ewels et al., 2020). Researchers have developed more than 71 pipelines to date, with most focusing on genetics and an increasing number on other scientific fields. Considering these factors, the development of a dedicated *nf-core* phasing and imputation pipeline, tailored to the unique demands of genomic analyses, became increasingly important. Such a pipeline would integrate state-of-the-art phasing and imputation algorithms, provide seamless compatibility with existing genomic datasets, and adhere to FAIR (i.e., Findable, Accessible, Interoperable, Reusable) principles to promote transparency and reproducibility in genomic research.

3 Materials and Methods

3.1 Workflow overview

The *nf-core/phaseimpute* pipeline was developed using the *nf-core* tooling framework (Ewels et al., 2020) and is implemented in *Nextflow* DSL2, a modular, updated version of the *Nextflow* language (Langer et al., 2025). Built from the *nf-core* template (version 3.1.1), the pipeline benefits from standardized directory structures, automated file generation, and pre-configured support for continuous integration via *GitHub*.

The development process began with a skeleton pipeline and progressively added components: starting with individual modules, followed by subworkflows, and culminating in the complete workflow. Following *nf-core* best practices (<https://nf-co.re/docs/guidelines/>), each module is

designed to wrap a single tool, simplifying maintenance, version control, and reuse across projects. Subworkflows are constructed by linking these modules and can be combined to form larger workflows.

All software tools are encapsulated within the *Bioconda* environment and have their *Singularity* and *Docker* counterparts. This containerization ensures a highly reproducible environment, guaranteeing consistent outputs for a given set of tools and data.

Visual representations of the workflow (i.e., “metro-maps”) were generated with the *draw.io* software (Benson, 2024).

3.2 Unit-testing data

3.2.1 Human dataset

We assessed the performance of all components in the pipeline, using data from the 1000 Genomes Project (Auton et al., 2015) and the *GRCh38* reference genome from the Genome Reference Consortium. Specifically, we used the phased panel “1000G_2504_high_coverage” which includes SNV and *InDels* call sets from 3,202 high-coverage samples Byrska-Bishop et al. (2022). This dataset is publicly available at: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20201028_3202_phased/

To streamline unit testing and reduce computational demands, we constructed a lightweight dataset of variants. This reduced dataset, available at [nf-core/test-datasets/tree/phaseimpute](https://nf-co.re/test-datasets/tree/phaseimpute) (commit hash: `dee41dd`), includes subregions

from chromosomes 21 and 22 spanning 16.57 Mb to 16.61 Mb. These regions were chosen to align with the *nf-core/test-datasets* standard, ensuring consistency with other *nf-core* pipelines while maintaining speed and simplicity in testing. For testing purposes, we selected three individuals from the reference panel: *NA12878* (also featured in the *Glimpse1* tutorial (https://odelaneau.github.io/GLIMPSE/glimpse1/tutorial_hg19.html); Delaneau & Rubinacci (2021)), along with *NA19401* and *NA20359*.

To ensure a fair assessment of imputation performance, all identified relatives (*NA12891*, *NA12892*, and *NA20362*) were excluded from the reference panel, along with the target individuals themselves (*NA12878*, *NA19401*, and *NA20359*), as including closely related samples would bias results by introducing haplotypes that are overly similar to the targets. Using the pedigree file provided by the 1000 Genomes Project ("1kGP.3202_samples.pedigree_info.txt"), we identified *NA12891* and *NA12892* as the parents of *NA12878*, *NA20362* as the child of *NA20359*, and confirmed that *NA19401* had no recorded relatives. These relationships were also verified using *plink* (version 1.90b6.21) (Purcell et al., 2007) with the `--genome` command.

Summary statistics were obtained for the three selected individuals using the command `coverage` of *samtools* (version 1.17) (Danecek et al., 2021).

We then retrieved the high-coverage Compressed Read Alignment Map files (CRAM) for the specified subregions from <ftp://ftp.sra.ebi.ac.uk/vol1/run>. Genotypes were called using the `mpileup` and `call` tools of *bcftools* (version 1.17) (Danecek et al., 2021). To simulate low-pass sequencing data, we down-sampled the original CRAM files to 1× coverage using the `subsample` option of the `view` command of *samtools* (version 1.17) (Danecek et al., 2021).

3.2.2 Non-human dataset

To extend testing beyond humans, we created a lightweight dog dataset using the *canFam3* reference genome from UCSC, a reference panel of 658 individuals from Animal-ImputeDB, and a BAM file for individual 1735 from iDog. This dataset can be accessed with the `--test_dog` profile. As with the human test dataset, only subregions of chromosomes 21 and 22 (16.57-16.61 Mb) were selected to reduce unit test computing time.

3.3 Benchmarking strategy

The benchmarking process was designed to evaluate the accuracy and computational performance of the *nf-core/phaseimpute* pipeline across its primary stages: reference panel pre-processing (`--panelprep`), genotype imputation (`--impute`), and concordance analysis (`--validate`). To ensure reproducibility and manageable runtimes, we restricted benchmarking to chromosomes 1 and 22, as they represent the longest and shortest chromosomes.

3.3.1 Benchmarking test data and setup

For the reference panel pre-processing step, we used the unphased, non-annotated data from the 1000 Genomes Project (see Table 1 for full reference of the source files).

The individual *NA12878* selected for testing was first down-sampled to 0.1, 0.2, 0.5, 2 and 4× coverage using through the *simulate* step of the pipeline. Imputation testing was conducted using phased data from the 1000 Genomes Project as the reference panel while target individual files came from the simulating step. We benchmarked imputation using *Glimpse* (versions 1 and 2) and *Quilt*. The *Stitch* software was not included, as it is designed to operate without a reference panel, and our setup with only three target individuals would not represent typical use cases and would yield unrealistically low accuracy.

3.3.2 Accuracy evaluation

To assess imputation accuracy, results were compared to high-confidence benchmark variants from the Genome In A Bottle (GIAB) initiative. Concordance metrics were obtained for the phasing test by comparing the results to the reference panel phased by the 1000 Genomes Project using the *Whatshap* compare tool (Martin et al., 2016).

3.3.3 Computational performance and carbon footprint

All benchmarking evaluations were conducted on a shared High-Performance

Computing (HPC) cluster using a *Slurm* scheduler on the GenOuest bioinformatic platform (www.genouest.org). The computational environment was configured using *mamba* (version 1.5.6), and the environment specification file is available in the pipeline repository (https://github.com/nf-core/phaseimpute/blob/master/tests/config/env_nf.yml). This environment file includes *openjdk* (version 17.0 or higher), *Nextflow* (version 23.10 or higher), *Singularity* (version 3.8 or higher), *nf-core* (version 3.1 or higher), *prettier* (version 3.0 or higher) and *nf-test* (version 0.8 or higher). All software versions used during computation in the pipeline are made available in an *MultiQC* (Ewels et al., 2016) *html* file. The *Singularity* profile was used for all the tests.

To quantify environmental and computational cost, the computational resources were measured using the *nf-co2footprint* *Nextflow* plugin (*Nextflow-io/Nf-Co2footprint*, 2023) which estimates the carbon footprint using the method developed in the Green Algorithms project (Lannelongue et al., 2021). This allowed for a comparison of the computational cost across different scenarios and also helped identify the most resource-intensive stages of the pipeline.

Time spent, energy consumed, and CO_2 equivalent are the three metrics provided by the *nf-co2footprint* plugin. The computation was done on 12 *Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30 GHz* for the phasing test and on 12 *Intel(R) Xeon(R) Gold 5220 CPU @ 2.20 GHz*

for all the other tests. The default carbon intensity (i.e., 475 CO_2e g/kWh) and power usage effectiveness (i.e., 1.00) were used to be more representative of what the user should expect.

The three metrics are linked through the following formula:

$$CO_2e = t \times (n_c \times p_c \times u_c + n_m \times p_m) \times PUE \times CI$$

Where:

- t is the running time (in hours) ;
- n_c is the number of used cores ;
- p_c is the power draw per core (12.0 W / core) ;
- u_c is the usage factor of the cores (between 0 and 1) ;
- n_m is the memory allocated (in GB) ;
- p_m is the power draw of memory (0.3725 W / Gb) ;
- PUE is the power usage effectiveness of the data centre (1.00 for our tests) ;
- CI is the carbon intensity of the respective energy production (475 CO_2e g / kWh).

The benchmarking was performed using CPU only, with PUE and CI set to their default values to reflect typical user conditions. Under this setup, CO_2 -equivalent emissions are mainly influenced by task duration, number of cores, and core usage, while memory consumption contributes negligibly compared to CPU power draw. Therefore, when the number of cores and their usage remain relatively constant, CO_2e emissions can be reasonably approximated as a linear

function of runtime, scaled by a constant factor ($e_c = n_c \times u_c$). Accordingly, in the following tests, since the effective number of cores remains nearly constant across tasks, the three metrics are presented as a linear scaling.

3.3.4 End-to-end validation

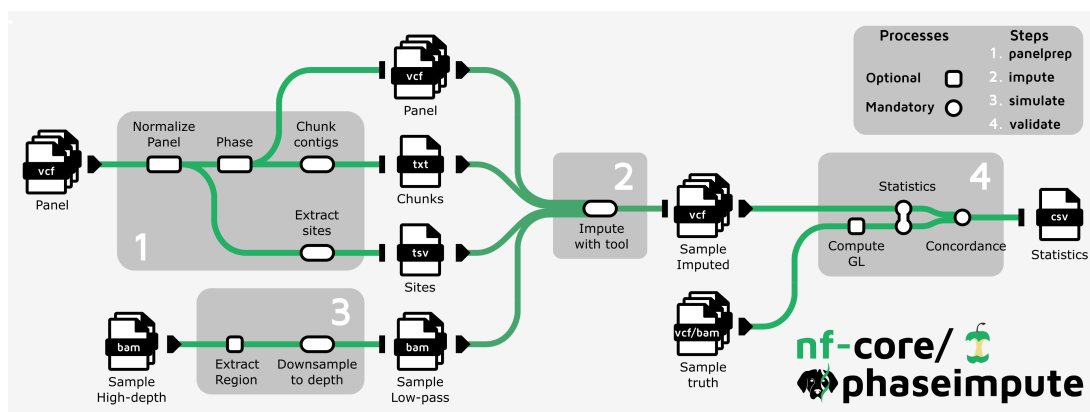
Lastly, the full pipeline, from simulation, panel pre-processing, imputation and concordance, was executed across the entire genome to validate end-to-end functionality to confirm end-to-end functionality and system stability under realistic workloads. The configuration used the *GRCh38* reference from iGenomes, the imputation tool *Glimpse2*, and the `full_test` profile. Key parameters included `chunk_model=sequential`, `ngen=100`, `k_val=2`, `normalize=true`, `phase=false`, `buffer=10 000`, `batch_size=100`, `seed=1`, and `remove_samples=NA12878,NA12891,NA12892`.

The test was performed in a cloud computing environment with *AWS Batch* using *Seqera Cloud* (<https://cloud.seqera.io/>). The compute environment was set with *AWS Batch Forge* mode within the eu-west-1 region, using *Spot* instances. It supported *Wave* containers, *Fusion v2*, and allowed fast instance storage for high I/O performance. The maximum pool of CPUs provisioned was 500.

TABLE 1 - Files used for the full chromosome size benchmarking

The following sites were used for data access *url1000GP*: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/ , *urlGIAB*: <ftp://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release> .

Step	File	URLs to the file
phasing	Unphased and halved unphased panel	url1000GP/20201028_3202_raw_GT_with_annot/20201028_CCDG_14151_B01_GRM_WGS_2020-08-05_chr{1-22}.recalibrated_variants.vcf.gz(tbi)
panelprep	Phased panel	url1000GP/20201028_3202_phased/CCDG_14151_B01_GRM_WGS_2020-08-05_chr{1-22}.filtered.shapeit2-duohmm-phased.vcf.gz(tbi)
simulate	NA12878 alignment file (high depth)	ftp : // ftp . sra . ebi . ac . uk / vol1 / run / ERR323 / ERR3239334/NA12878.final.cram(crai)
impute	Phased panel	Results from the panelprep step
	NA12878 alignment file (low depth)	Results from the simulate test at 5 different depths
validate	Truth file	urlGIAB/NA12878_HG001/NISTv4.2.1/GRCh38/HG001_GRCh38_1_22_v4.2.1_benchmark.vcf.gz(tbi)

FIGURE 1 - Overview of the *nf-core/phaseimpute* pipeline workflow

One box per step of the pipeline. 1: pre-processing of the panel, 2: target files imputation, 3: simulation of low coverage WGS and genotyping array data, 4: validation of imputed data.

modules / subworkflows:

- local / nf-core: # The code can be shared across nf-core
or only be available in this pipeline
 - tool: # Name of the tool (e.g., glimpse2)
 - subtool: # Name of the subtool (e.g., chunk)
 - tests:
 - main.nf.test # Process code for the unit tests
 - # Snapshot of the expected results
- main.nf.test.snap
- nextflow.config # Configuration for the unit tests
 - tags.yml # Tags and path for the unit test
 - environment.yml # Environment file of the tool used by conda
 - main.nf # Modules or subworkflow code
 - meta.yml # Documentation describing the tool
 - subtools.diff # Git diff text to patch modifications

FIGURE 2 - *nf-core* file organisation for modules and subworkflows with *nf-test*

Each module or subworkflow contains all the files needed for its execution and testing.

TABLE 2 - Summary statistics of the three individuals selected

Chr: Chromosomes selected regions are chr21: 16.57-16.61 Mb and chr22: 16.57-16.61 Mb, numreads: Number reads aligned to the region (after filtering), covbases: Number of covered bases with depth ≥ 1 , coverage: Percentage of covered bases [0..100], meandepth: Mean depth of coverage, meanbaseq: Mean baseQ in covered region, meanmapq: Mean mapQ of selected reads.

Sample	Chr	numreads	covbases	coverage	meandepth	meanbaseq	meanmapq
NA12878	21	8475	40001	100.000	31.53	29.6	59.7
	22	8973	40001	100.000	33.20	29.5	57.3
NA20359	21	8143	40000	99.998	30.18	29.3	59.8
	22	8384	40001	100.000	30.91	29.1	57.3
NA19401	21	8741	39996	99.988	32.49	29.6	59.8
	22	8682	40001	100.000	32.07	29.5	57.7

4 Results

The pipeline encompasses four steps: (i) reference panel phasing and preprocessing, (ii) imputation of the target files, (iii) simulation of target files and (iv) concordance analysis between ground truth and imputed files (see Figure 1). Each stage can be executed individually by specifying the desired step with: `--steps <panelprep/impute/simulate/-validate>` argument. The pipeline allows for multiple entry points, allowing users to perform various types of analyses. Each entry point generates a CSV sample sheet that can be passed into subsequent steps. Likewise, the imputation tool can be selected with `--tools <glimpse1/glimpse2/stitch/quilt>`.

The input data comprises a sample sheet in comma-separated value (CSV) format, which specifies the target individuals alongside the paths to their genetic files (e.g., VCF, BCF, BAM, CRAM) and corresponding indices (e.g., TBI, CSI, BAI, CRAI). Depending on the step and the imputation tool, the user may also need to supply a reference genome, making the pipeline species-agnostic. The only limitation is that the species to be imputed needs to be diploid.

Each tool parameter can be easily customized by supplying external arguments through `.config` files. For example, imputation can be fine-tuned by adding `--ne 100` to the `GLIMPSE1_phase` process, which explicitly specifies the effective diploid population size, or by adding `--window-size 200 000` to the

`GLIMPSE1_chunk` process to increase or decrease the window size. Key parameters with greater influence on results are also exposed at the pipeline-level (i.e., prefixed with `--`) and can easily be adjusted from the command line. For example the following command alters the default *Stitch* software main parameters while running the pipeline.

```
nextflow run \
  nf-core/phaseimpute \
  -profile test_stitch,conda \
  --k_val 3 --seed 3 \
  --outdir results
```

4.1 Reference panel preparation and phasing panelprep

The first step of the workflow, referred to as `panelprep`, (see Figure 1, step 1), performs the required preprocessing to prepare the reference panel to be used for downstream analyses. This step is essential when working with imputation tools that require a pre-phased reference panel, such as *Glimpse* or *Quilt*.

The process begins with normalization of the Variant Calling Format (VCF) files, retaining only the biallelic Single Nucleotide Polymorphisms (SNPs). The normalised panel genotypes are then phased into haplotypes, using the *Shapeit5* tools `phase_common` and `ligate` applied within a sliding window generated by `GLIMPSE2_chunk`.

For the phasing step, users must supply the reference genome of the target species and, optionally, a genetic map to better determine the recombination rate

along the chromosomes. The pipeline automatically handles file conversions to generate the required input files for the different imputation software.

Once the panel is phased, the panel's chromosomes are split into smaller genomic intervals (i.e., chunks) using *Glimpse* (versions 1 and 2) to allow high parallelisation in subsequent steps. Users can define custom chunk sizes.

An additional step, site extraction, isolates the variant positions to be used during imputation. These variant site files, along with chunking metadata, form the primary output of the *panelprep* step and are critical for the later imputation stage.

Several options allow for customization during the *panelprep* step: `--phase`: Enables or disables phasing of the reference panel, `--remove_samples`: Accepts a list of samples to exclude, which is useful for removing close relatives based on pedigree data and `--compute_freq`: determines whether to calculate allele counts (AC/AN) per variant.

4.2 Imputation of the target individuals' variants impute

The second step (see Figure 1, step 2), called *impute*, receives input in the form of BAM/CRAM or VCF files listed in a CSV sample sheet and performs genotype imputation using preprocessed reference panel files generated during the *panelprep* step. This step supports four imputation tools: *Glimpse* version 1 and version 2, *Stitch* and *Quilt* each with distinct input requirements and internal algorithms tailored for specific data types

and coverage levels. The modular design of this step allows users to choose the most appropriate tool based on input data type, computational resources, and research goals. Multiple tools can be specified in a single run using `--tools` (e.g., `--tools glimpse1,quilt`), allowing for direct comparison.

To initiate this step, users must specify `--steps impute` and provide the appropriate reference genome via the `--genome` or `--fasta` flag. Depending on the selected imputation tool(s), the workflow may require additional input files, including: `--chunks` (*chunks.csv*), `--posfile` (*posfile.csv*) and/or `--panel` (*panel.csv*). Each of these files are produced during the *panelprep* stage and reused here for imputation.

Upon execution, the workflow imputes missing variants using the selected tool(s), aggregates the imputed data across chromosomes, and returns the imputed VCF for each sample.

4.3 Simulation and concordance analysis simulate and validate

An optional initial stage of the pipeline, the *simulate* step (see Figure 1, step 3) enables the generation of sparse genetic data from high-coverage whole genome sequence data. This simulation process is accomplished by downsampling BAM or CRAM files to a user-defined desired depth via the `--depth` flag. This flag controls the specified mean coverage via the `-s` option in *samtools* view. This process is particularly useful when true low-pass sequencing data is unavailable

but high-quality sequence data exists, allowing one to create controlled test scenarios for imputation evaluation.

The final stage of the pipeline, the validate step (see Figure 1, step 4), provides a framework for assessing imputation performance using the GLIMPSE2-concordance tool to perform statistical analysis. This step requires input from both the ground truth and the imputed data from the simulated data. As VCF files are needed for both inputs, a subworkflow is used to call the variant from whole-genome sequence alignment files. The final output of this validation process is a CSV file summarizing concordance statistics, in particular, the proportion of correctly imputed genotypes relative to the truth data.

4.4 Unit testing

The entire *nf-core* community has recently opted to transition all their *Python* test procedures to *Groovy*-based test procedures available through the *nf-test* (<https://www.nf-test.com>) plugin. Each module and subworkflow undergoes separate testing using the test dataset to ensure that each component operates as intended. This new unit test workflow is structured so that each module includes all the files needed for its execution and testing within its own folder. This organisational approach improves visualisation and code debugging. A module is therefore organised as depicted in Figure 2.

The advantage of such an organisation is the ability to directly test a module or a subworkflow with the simple following command:

```
nf-core <modules/subworkflows> \
  test tool/subtool \
  --profile <docker/conda/...>
```

The pipeline also features multiple tests that can be easily initiated using the `--profile` command:

```
nextflow run \
  nf-core/phaseimpute \
  --profile <docker/conda/...>,
  <test_quilt/test_sim/...> \
  --outdir results
```

Table 2 presents the statistical information obtained by the coverage tool of *samtools* for the three individuals selected as targets for the test dataset.

4.5 Benchmarking results

4.5.1 Test of panel phasing

The phasing test was made on the longest and shortest chromosomes of the human genome and used either the full panel or a smaller version using only half the individuals of the reference panel of the *1000 Genomes Project*. The phasing process took 1 day and 13 h to complete (436 CPU-hours, $e_c = 3.53 \pm 1.99$) for the full panel ($N = 3202$ individuals) and took 21 h 24 m to complete (252 CPU-hours, $e_c = 3.78 \pm 1.78$) for the halved panel ($N = 1601$ individuals). Figure S1 illustrates the repartition of the computational load by task and chromosome.

The concordance between the phased haplotype by our pipeline and the ones available by the *1000GP* was computed for the first ten samples in the dataset using the *WhatsHap* compare tool. This tool provides three different metrics: switch rates (i.e., phased variant pairs that require a switch in haplotype assignment to match the reference), switch and flip rate (i.e., phased variant pairs with either isolated switch or block-level flip) and the Hamming distance rate (i.e., the fraction of heterozygous sites at which the assigned haplotype differs from the reference) for each individual. The Hamming distance provides a global measure of haplotype agreement, and a seemingly high value should not be interpreted as poor genotype accuracy, but rather as a reflection of the difficulty in accurately reconstructing long haplotype blocks from sparse data. The mean and standard deviation for chromosomes 1 and 22 with the full and halved panel are available in Table 3, as well as a graphic representation in Figure 3. Switch and switch-and-flip error rates are of the same order of magnitude as those reported by Hofmeister et al. (2023) for the UK Biobank data.

4.5.2 Test of panel preparation

The panel preparation test was made on the full phased reference panel of the 1000 Genomes Project. The panel preparation process took 7 h and 49 min to complete (88 CPU-hours, $e_c = 1.49 \pm 0.47$). See Figure S2

for the details of the repartition of the computational load by task.

4.5.3 Test of the simulation process

Simulation was done for 6 different target genomic depths (i.e., 0.1, 0.2, 0.5, 1, 2 and 4×) and took 80 ± 18.6 min to complete (4.4 ± 0.5 CPU-hours, $e_c = 0.69 \pm 0.38$). The relationship between the simulated genome coverage depth and the effective CPU time (in minutes) was modeled using a quadratic regression. The best-fit equation was: $y = 80.57 + 4.83 \times x - 0.13 \times x^2$ where y is the effective CPU time (in minutes) is defined as the product of the number of cores used and their usage fraction and x is the target genome coverage depth. This model indicates that CPU usage initially increases nearly linearly with depth and slowly saturates for high depth ($R^2 = 1$).

See Figure S3 for the details of the regression and of the repartition of the computational load by task and target genomic depth.

4.5.4 Test of imputation process

Imputation was then done on the 6 simulated files, and with the panel prepared in the precedent step for imputation tools *Glimpse1*, *Glimpse2* and *Quilt*. It took 2 day 16 h 54 m to complete (4.4 ± 0.5 CPU-hours, $e_c = 0.69 \pm 0.38$). See Figure S4 for the details of the repartition of the computational load by task and imputation tool. Evaluation of imputation accuracy results is performed during the concordance step of the pipeline.

TABLE 3 - Phasing accuracy by chromosomes and panel size (mean $\pm$ standard deviation).

The concordance metrics were computed using the *WhatsHap compare* tool between the phased haplotypes obtained with our pipeline and the ones provided by the 1000 Genome Project. The mean and standard deviation are given for the first ten samples of the dataset for chromosomes 1 and 22 with the full ($N = 3202$ individuals) and halved ($N = 1601$ individuals) panel.

Chr	Panel size	Switch rate	Switch and flip rate	Hamming distance rate
1	Full	0.0094 ± 0.0006	0.0063 ± 0.0004	0.477 ± 0.0124
	Halved	0.0105 ± 0.0008	0.0070 ± 0.0005	0.470 ± 0.0249
22	Full	0.0197 ± 0.0020	0.0133 ± 0.0014	0.443 ± 0.0433
	Halved	0.0200 ± 0.0021	0.0135 ± 0.0014	0.445 ± 0.0434

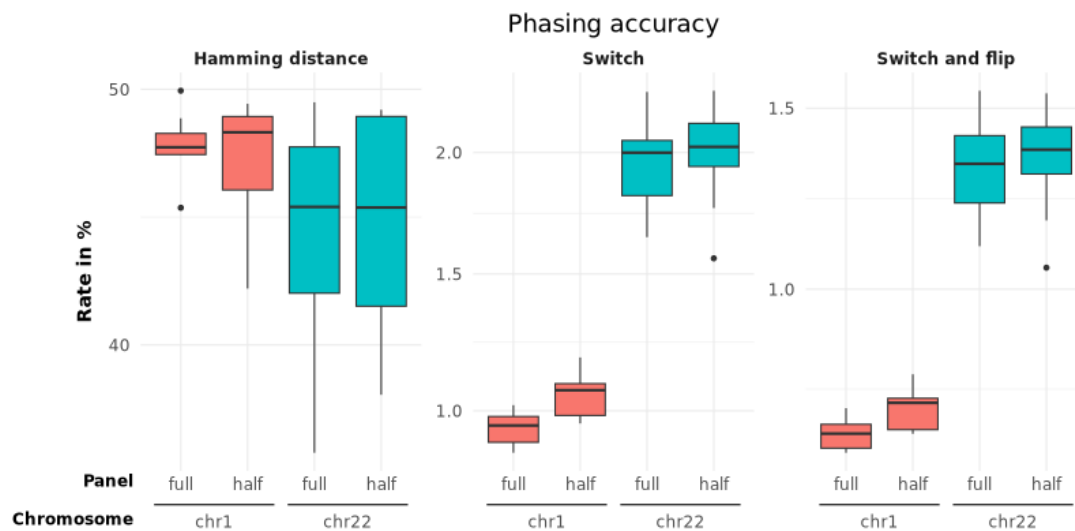


FIGURE 3 - Panel phasing accuracy by metric, chromosome and panel size

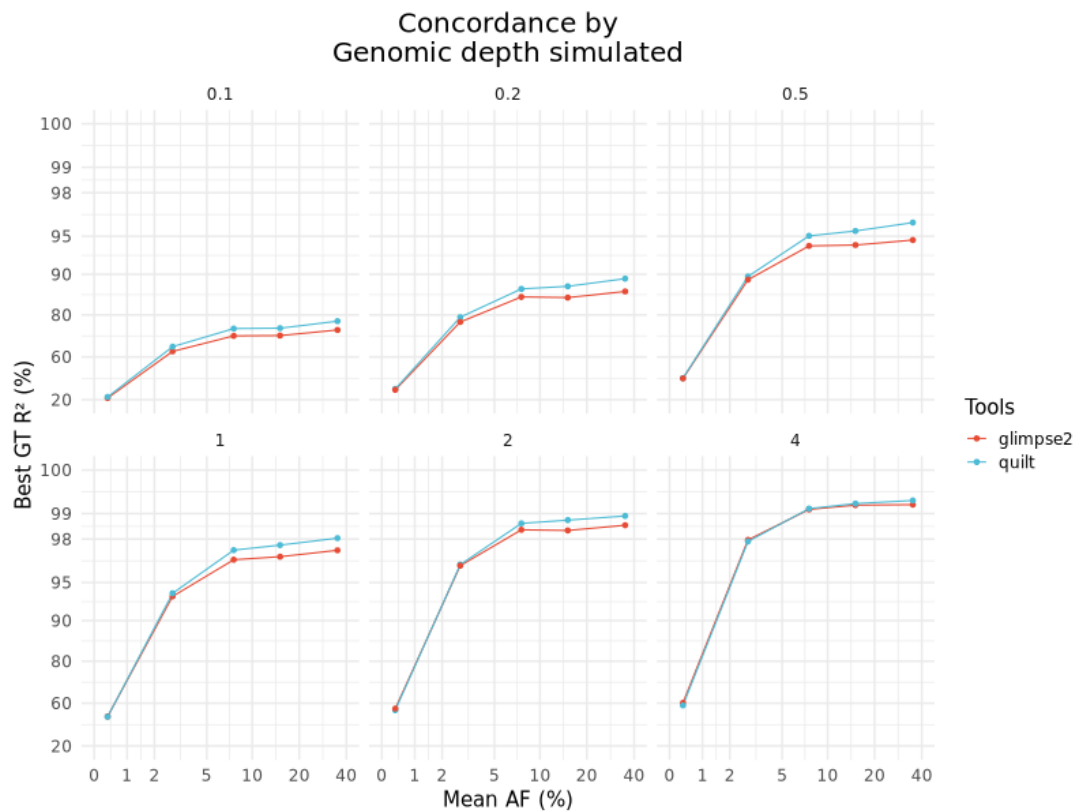


FIGURE 4 - Concordance analysis of *Glimpse2* and *Quilt* software best imputed genotype r-squared rate at different simulated depth.

The concordance metrics were computed using the *Glimpse2* concordance tool between the imputed genotypes and the ground truth provided by GIAB for the individual. Each panel represents the genotype r-squared rate by bins of minor allele frequency (MAF) for the different simulated depth.

4.5.5 Test of concordance process

The accuracy of the imputation on the 6 simulated files was then assessed using the genotype provided by Genome In A Bottle. It took 1 h 19 m to complete (12 CPU-hours, $e_c = 0.93 \pm 0.26$). See Figure S5 for the details of the repartition of the computational load by task and imputation tool.

This step produces a CSV file containing the concordance metrics for each sample by bins of minor allele frequency (MAF). *Glimpse1* is not shown in Figure 4, as it leverages the genotype observed in the joint target VCF file; therefore, the concordance metrics do not accurately reflect the simulated depth. On the other hand, *Quilt* software performs better than *Glimpse2* for high MAF at lower coverage (see Figure 4). The current pipeline configuration shows lower accuracy for rare variants (MAF < 5 %) compared to the results reported by Rubinacci et al. (2020), Rubinacci et al. (2023), and Davies et al. (2021). Fine-tuning parameters and incorporating detailed genomic recombination maps may improve these results and are planned for future updates.

4.5.6 Whole pipeline test

This test is available as a profile named `test_full` and is the one run before each release on the *Seqera* platform. This ensures that the pipeline is able to run seamlessly on a whole genome. These results are available in the *Results* section of the pipeline webpage on the *nf-core* website (<https://nf-co.re/>

`phaseimpute/1.0.0/results/phaseimpute/`). The *nf-core/phaseimpute* run (*Git* commit 61c18f6; Nextflow 24.10.2, run ID `reverent_mandelbrot`) started on 2024-12-06 05:51:22 and finished in 3 h 7 m 43 s (wall), accumulating 129.7 h CPU time, using 1044.93 GB memory, reading 501.50 GB and writing 139.44 GB of data, with an estimated cost of \$4.585. The `multiqc_report.html` file allows the user to screen all the main statistics of the pipeline run in one place.

5 Discussion

The use of a dedicated workflow language offers significant advantages for seamlessly processing high-throughput data while ensuring high reproducibility (Di Tommaso et al., 2017; Mölder et al., 2021). The *nf-core* initiative, which aims to create highly standardised tools, modules, sub-workflows, workflows, and templates, enables users to rapidly develop high-quality pipelines.

Integrating each tool into the *nf-core* repository ensures users access the latest versions and maintain cutting-edge workflows. This inclusion of pre-developed modules also eliminates redundant development. Furthermore, *nf-core*'s guidelines guarantee versatile tool usage through a standardised input-output architecture. The *nf-core* tool and *GitHub* Continuous Integration (CI) facilitate tracking changes, linting, and automatic testing across all processes and workflows. This ensures that the pipeline remains robust and up to date. Archiving each version within *Zenodo*

(Potter & Smith, 2015) and *GitHub* enables reproducibility of analyses even decades later, ensuring proper tool versioning without normally encountering any issues.

The modular design of *nf-core/phaseimpute* makes it highly versatile. We demonstrate the primary configurations a user might encounter, from the simulation of low-depth data, phasing and preparation of the reference panel to the imputation and concordance analysis. Through the use of the `params.yaml` file, it is easy to set the different parameters as well as reuse the same configuration.

Future efforts will focus on integrating new features, refining existing functionalities, and adapting to the evolving needs of the rapidly growing field of genetic imputation. Community engagement will play a crucial role in shaping the pipeline's development trajectory, fostering collaboration from the broader bioinformatics community.

6 Availability and implementation

The stable release of the *nf-core/phaseimpute* pipeline is accessible at <https://nf-co.re/phaseimpute/1.0.0>.

The test dataset is located in the *nf-core/test-datasets* repository, specifically under its dedicated branch, accessible at <https://github.com/nf-core/test-datasets/tree/phaseimpute>.

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.14329226>.

7 Acknowledgements

The authors express their gratitude to the *nf-core* and *Nextflow* community for their efforts in developing the *nf-core* infrastructure and resources for *Nextflow* pipelines. A comprehensive list of *nf-core* community members can be found at <https://nf-co.re/community>. Particularly, the authors deeply thank Eugenia Fontecha, Matias Romero Victorica, Hemanoel Passarelli and Saul Pierotti for their contribution to the pipeline and to Maxime Garcia, Matthias Hörtenhuber, Tomasso Mazza and Sofia Stamouli for their reviews.

The authors extend their thanks and acknowledge the GenOuest bioinformatics core facility (<https://www.genouest.org>) for providing the computing infrastructure.

8 Funding

This work was performed in the frame of L.L.N.'s PhD research, as a part of a project on the comparative genetics of developmental genetic diseases in dogs and humans, funded by the “Fondation pour la Recherche Médicale” and the french guide dogs association: “Association les Chiens Guides d'Aveugles de l'Ouest” (ACGAO) as well as the french foundation “Visio”, which supports visually impaired children and adults.

9 Competing interests

No competing interest is declared.

10 Bibliography

- AUTON, A. et al., 2015. A Global Reference for Human Genetic Variation. *Nature* [en ligne]. Vol. 526, no. 7571, p. 68–74 [visité le 2025-08-27]. ISSN 1476-4687. Disp. à l'adr. DOI: 10.1038/nature15393.
- BENSON, D., 2024. Draw.io a JavaScript, Client-Side Editor for General Diagramming. [en ligne]. Version 24.4.6 [visité le 2024-05-23]. Disp. à l'adr.: <https://www.drawio.com/>.
- BROWNING, B. L. et al., 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics*. Vol. 103, no. 3, p. 338–348. ISSN 1537-6605. Disp. à l'adr. DOI: 10.1016/j.ajhg.2018.07.015.
- BYRSKA-BISHOP, M. et al., 2022. High-Coverage Whole-Genome Sequencing of the Expanded 1000 Genomes Project Cohort Including 602 Trios. *Cell*. Vol. 185, no. 18, p. 3426–3440.e19. ISSN 1097-4172. Disp. à l'adr. DOI: 10.1016/j.cell.2022.08.004.
- CHAT, V. et al., 2022. Ultra Low-Coverage Whole-Genome Sequencing as an Alternative to Genotyping Arrays in Genome-Wide Association Studies. *Frontiers in Genetics* [en ligne]. Vol. 12 [visité le 2024-03-15]. ISSN 1664-8021.
- COHEN-BOULAKIA, S. et al., 2017. Scientific Workflows for Computational Reproducibility in the Life Sciences: Status, Challenges and Opportunities. *Future Generation Computer Systems* [en ligne]. Vol. 75, p. 284–298 [visité le 2024-03-15]. ISSN 0167-739X. Disp. à l'adr. DOI: 10.1016/j.future.2017.01.012.
- DANECEK, P. et al., 2021. Twelve Years of SAMtools and BCFtools. *GigaScience* [en ligne]. Vol. 10, no. 2, giab008 [visité le 2024-03-15]. ISSN 2047-217X. Disp. à l'adr. DOI: 10.1093/gigascience/giab008.
- DAS, S. et al., 2016. Next-Generation Genotype Imputation Service and Methods. *Nature Genetics*. Vol. 48, no. 10, p. 1284–1287. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/ng.3656.
- DAVIES, R. W. et al., 2016. Rapid Genotype Imputation from Sequence without Reference Panels. *Nature Genetics* [en ligne]. Vol. 48, no. 8, p. 965–969 [visité le 2024-03-15]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/ng.3594.
- DAVIES, R. W. et al., 2021. Rapid Genotype Imputation from Sequence with Reference Panels. *Nature Genetics* [en ligne]. Vol. 53, no. 7, p. 1104–1111 [visité le 2024-03-15]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-021-00877-0.
- DELANEAU, O. & RUBINACCI, S., 2021. GLIMPSE Tutorial [en ligne]. Genotype Likelihoods IMputation and PhaSing mEthod (GLIMPSE) tutorial b38, 2021-05-21. [visité le 2024-03-15].
- DELANEAU, O. et al., 2008. Shape-IT: New Rapid and Accurate Algorithm for Haplotype Inference. *BMC Bioinformatics* [en ligne]. Vol. 9, no. 1, p. 540 [visité le 2024-03-15]. ISSN 1471-2105. Disp. à l'adr. DOI: 10.1186/1471-2105-9-540.
- DI TOMMASO, P. et al., 2017. Nextflow Enables Reproducible Computational Workflows. *Nature Biotechnology* [en ligne]. Vol. 35, no. 4, p. 316–319 [visité le 2024-03-15]. ISSN 1546-1696. Disp. à l'adr. DOI: 10.1038/nbt.3820.
- EWELS, P. et al., 2016. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* [en ligne]. Vol. 32, no. 19, p. 3047–3048 [visité le 2024-05-21]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/btw354.
- EWELS, P. A. et al., 2020. The Nf-Core Framework for Community-Curated Bioinformatics Pipelines. *Nature Biotechnology* [en ligne].

- Vol. 38, no. 3, p. 276–278 [visit  le 2023-03-16]. ISSN 1546-1696. Disp.   l'adr. DOI: 10.1038/s41587-020-0439-x.
- FEARNHEAD, P. & DONNELLY, P., 2001. Estimating Recombination Rates from Population Genetic Data. *Genetics*. Vol. 159, no. 3, p. 1299–1318. ISSN 0016-6731. Disp.   l'adr. DOI: 10.1093/genetics/159.3.1299.
- GRANT, S. F. A. & HAKONARSON, H., 2008. Microarray Technology and Applications in the Arena of Genome-Wide Association. *Clinical Chemistry* [en ligne]. Vol. 54, no. 7, p. 1116–1124 [visit  le 2025-09-16]. ISSN 0009-9147. Disp.   l'adr. DOI: 10.1373/clinchem.2008.105395.
- HOFMEISTER, R. J. et al., 2023. Accurate Rare Variant Phasing of Whole-Genome and Whole-Exome Sequencing Data in the UK Biobank. *Nature Genetics* [en ligne], p. 1–7 [visit  le 2023-07-11]. ISSN 1546-1718. Disp.   l'adr. DOI: 10.1038/s41588-023-01415-w.
- LANGER, B. E. et al., 2025. Empowering Bioinformatics Communities with Nextflow and Nf-Core. *Genome Biology* [en ligne]. Vol. 26, no. 1, p. 228 [visit  le 2025-09-29]. ISSN 1474-760X. Disp.   l'adr. DOI: 10.1186/s13059-025-03673-9.
- LANNELONGUE, L. et al., 2021. Green Algorithms: Quantifying the Carbon Footprint of Computation. *Advanced Science* [en ligne]. Vol. 8, no. 12, p. 2100707 [visit  le 2025-04-28]. ISSN 2198-3844. Disp.   l'adr. DOI: 10.1002/advs.202100707.
- LI, J. H. et al., 2021. Low-Pass Sequencing Increases the Power of GWAS and Decreases Measurement Error of Polygenic Risk Scores Compared to Genotyping Arrays. *Genome Research* [en ligne]. Vol. 31, no. 4, p. 529–537 [visit  le 2024-03-15]. ISSN 1088-9051. Disp.   l'adr. DOI: 10.1101/gr.266486.120.
- LIN, S. et al., 2004. Haplotype and Missing Data Inference in Nuclear Families. *Genome Research*. Vol. 14, no. 8, p. 1624–1632. ISSN 1088-9051. Disp.   l'adr. DOI: 10.1101/gr.2204604.
- LOH, P.-R. et al., 2016. Reference-Based Phasing Using the Haplotype Reference Consortium Panel. *Nature Genetics*. Vol. 48, no. 11, p. 1443–1448. ISSN 1546-1718. Disp.   l'adr. DOI: 10.1038/ng.3679.
- LOU, R. N. et al., 2021. A Beginner's Guide to Low-Coverage Whole Genome Sequencing for Population Genomics. *Molecular Ecology* [en ligne]. Vol. 30, no. 23, p. 5966–5993 [visit  le 2024-03-15]. ISSN 1365-294X. Disp.   l'adr. DOI: 10.1111/mec.16077.
- MARCHINI, J. & HOWIE, B., 2010. Genotype Imputation for Genome-Wide Association Studies. *Nature Reviews Genetics* [en ligne]. Vol. 11, no. 7, p. 499–511 [visit  le 2022-11-10]. ISSN 1471-0064. Disp.   l'adr. DOI: 10.1038/nrg2796.
- MARINO, A. D. et al., 2022. A Comparative Analysis of Current Phasing and Imputation Software. *PLOS ONE* [en ligne]. Vol. 17, no. 10, e0260177 [visit  le 2024-03-15]. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0260177.
- MARTIN, M. et al., 2016. WhatsHap: Fast and Accurate Read-Based Phasing [en ligne]. 2016-11-02. [visit  le 2025-04-28]. Disp.   l'adr. DOI: 10.1101/085050.
- MAZZONETTO, P. C. et al., 2024. Low-Pass Whole Genome Sequencing Is a Reliable and Cost-Effective Approach for Copy Number Variant Analysis in the Clinical Setting. *Annals of Human Genetics* [en ligne]. Vol. 88, no. 2, p. 113–125 [visit  le 2024-03-15]. ISSN 1469-1809. Disp.   l'adr. DOI: 10.1111/ahg.12532.

- MÖLDER, F. et al., 2021. Sustainable Data Analysis with Snakemake [en ligne]. No. 10:33 [visité le 2024-03-15]. Disp. à l'adr. DOI: 10.12688/f1000research.29032.2.
- Nextflow-Io/Nf-Co2footprint, 2023 [en ligne]. nextflow-io. [visité le 2025-04-28]. Disp. à l'adr.: <https://nextflow-io.github.io/nf-co2footprint/>.
- NIU, T. et al., 2002. Bayesian Haplotype Inference for Multiple Linked Single-Nucleotide Polymorphisms. *American Journal of Human Genetics*. Vol. 70, no. 1, p. 157–169. ISSN 0002-9297. Disp. à l'adr. DOI: 10.1086/338446.
- PASANIUC, B. et al., 2012. Extremely Low-Coverage Sequencing and Imputation Increases Power for Genome-Wide Association Studies. *Nature genetics* [en ligne]. Vol. 44, no. 6, p. 631–635 [visité le 2024-03-15]. ISSN 1061-4036. Disp. à l'adr. DOI: 10.1038/ng.2283.
- POTTER, M. & SMITH, T., 2015. Making Code Citable with Zenodo and GitHub [en ligne] [visité le 2024-05-21]. Disp. à l'adr. DOI: 10.5281/zenodo.45042.
- PURCELL, S. et al., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* [en ligne]. Vol. 81, no. 3, p. 559–575 [visité le 2022-11-04]. ISSN 0002-9297. Disp. à l'adr. DOI: 10.1086/519795.
- RUBINACCI, S. et al., 2020. Genotype Imputation Using the Positional Burrows Wheeler Transform. *PLOS Genetics* [en ligne]. Vol. 16, no. 11, e1009049 [visité le 2024-03-15]. ISSN 1553-7404. Disp. à l'adr. DOI: 10.1371/journal.pgen.1009049.
- RUBINACCI, S. et al., 2021. Efficient Phasing and Imputation of Low-Coverage Sequencing Data Using Large Reference Panels. *Nature Genetics* [en ligne]. Vol. 53, no. 1, p. 120–126 [visité le 2025-09-24]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-020-00756-0.
- RUBINACCI, S. et al., 2023. Imputation of Low-Coverage Sequencing Data from 150,119 UK Biobank Genomes. *Nature Genetics* [en ligne]. Vol. 55, no. 7, p. 1088–1090 [visité le 2025-09-26]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-023-01438-3.
- STEPHENS, M. & SCHEET, P., 2005. Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. *American Journal of Human Genetics*. Vol. 76, no. 3, p. 449–462. ISSN 0002-9297. Disp. à l'adr. DOI: 10.1086/428594.
- VERLOUW, J. A. M. et al., 2021. A Comparison of Genotyping Arrays. *European Journal of Human Genetics* [en ligne]. Vol. 29, no. 11, p. 1611–1624 [visité le 2024-05-21]. ISSN 1476-5438. Disp. à l'adr. DOI: 10.1038/s41431-021-00917-7.

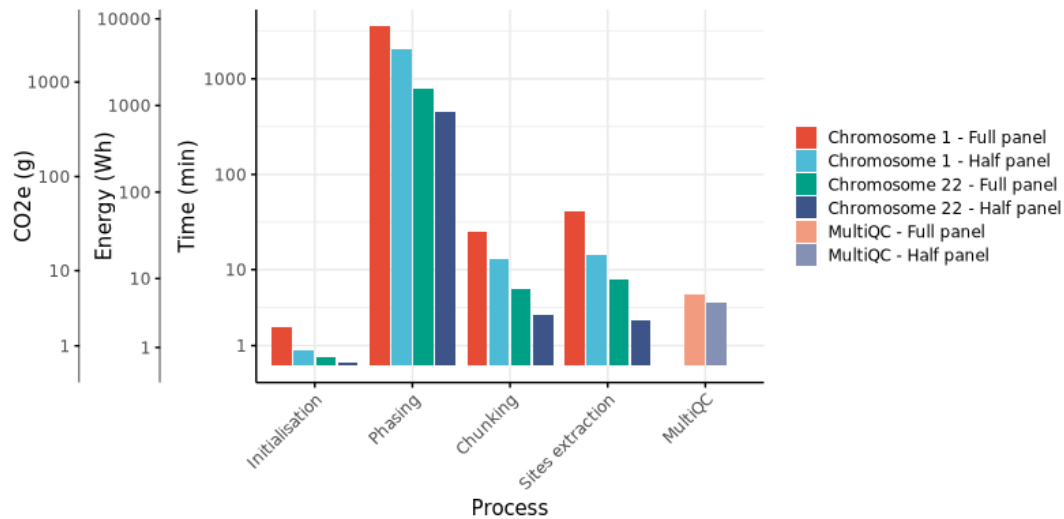


FIGURE S1 - Panel phasing computational load by chromosome and tasks for time (in minutes), energy consumed (in Watts) and CO₂ equivalent (in grams).

The computational load is represented part of the process for chromosome 1 and 22 for the full panel ($N = 3202$ individuals) and the halved panel ($N = 1601$ individuals).

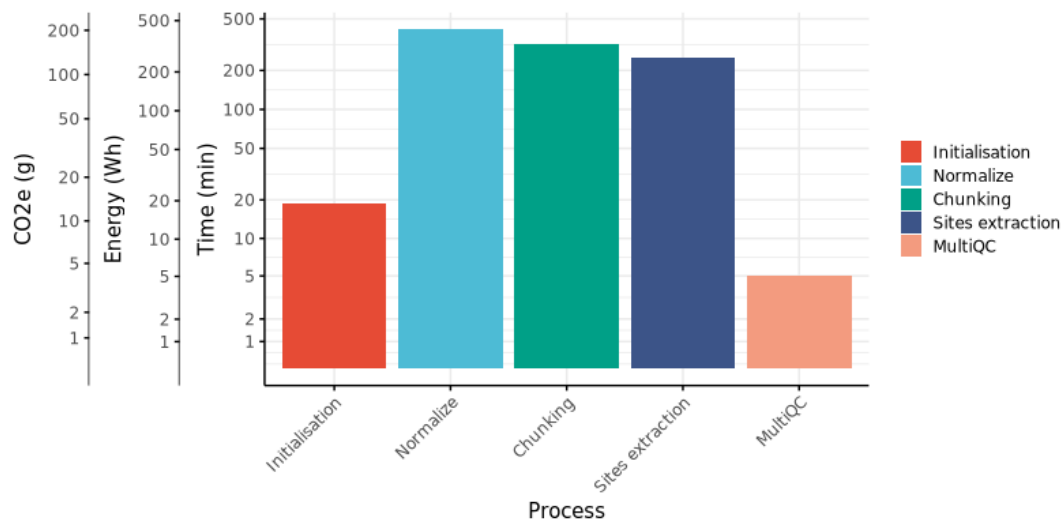


FIGURE S2 - Computational load for the panel preparation test by tasks for time (in minutes), energy consumed (in Watts) and CO₂ equivalent (in grams).

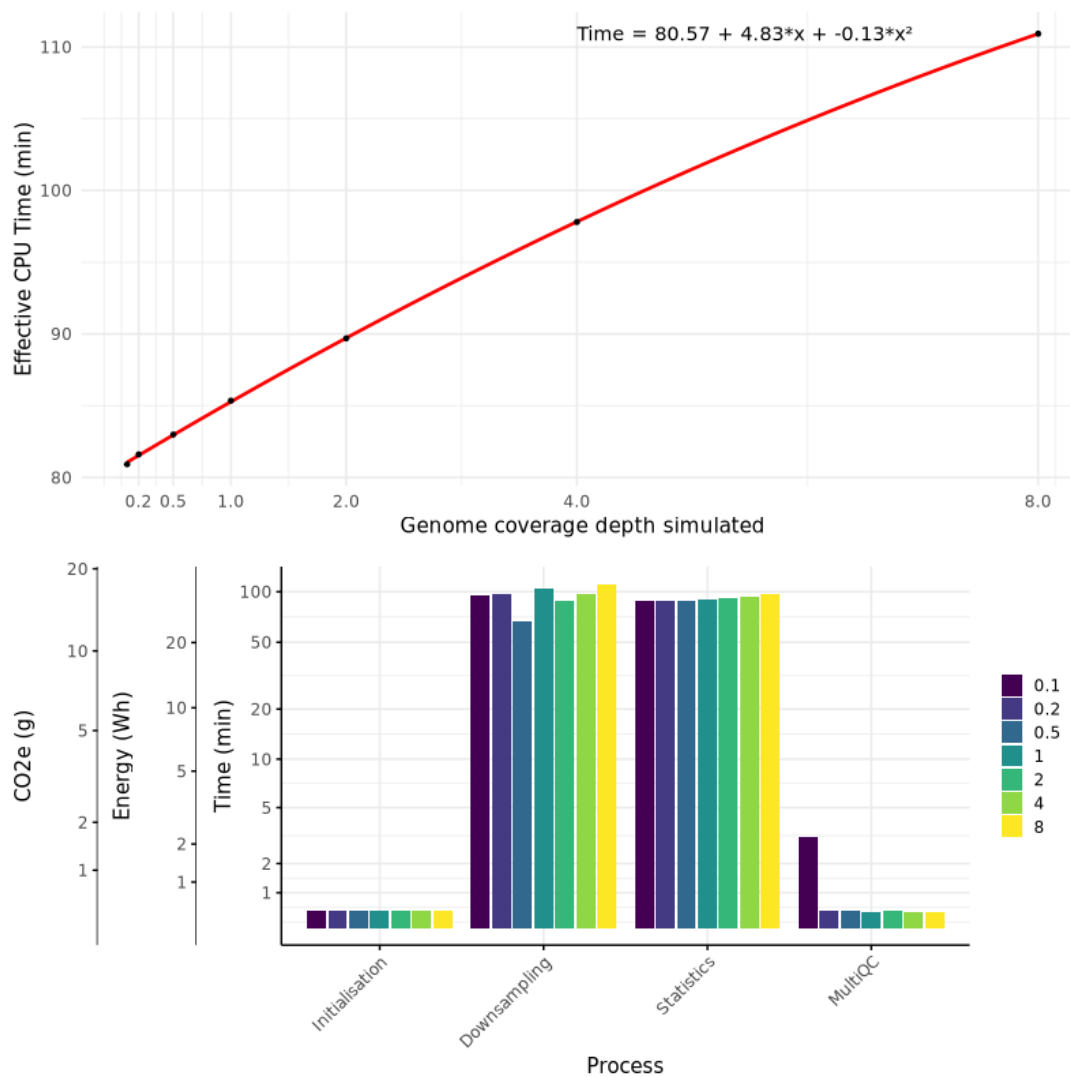


FIGURE S3 - Computational load for the downsampling test by target genomic depth

Top panel: time in minute to run the full test by the target downsampling coverage. Bottom panel: Computational load by tasks and target coverage for time (in minutes), energy consumed (in Watts) and CO₂ equivalent (in grams).

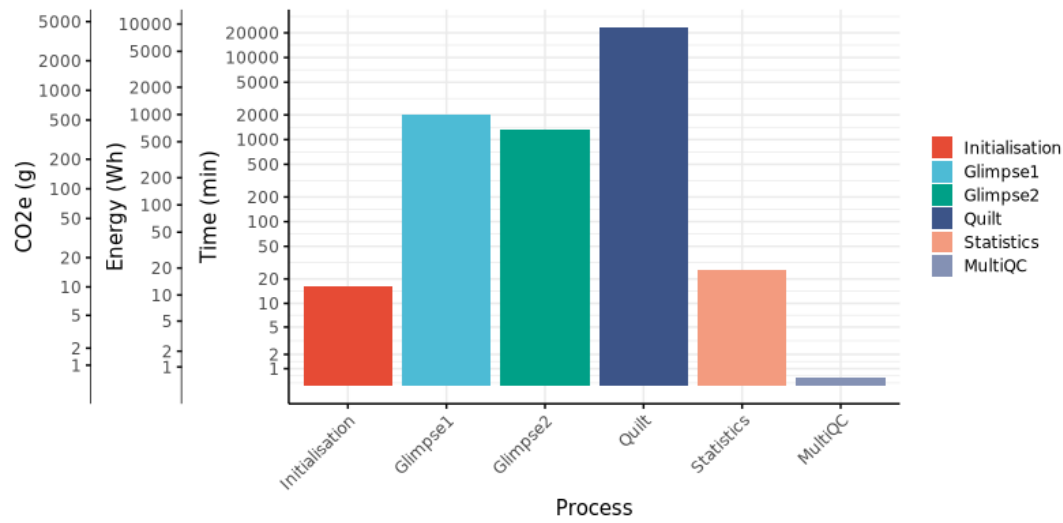


FIGURE S4 - Computational load for the imputation test by tasks for time (in minutes), energy consumed (in Watts) and CO_2 equivalent (in grams).

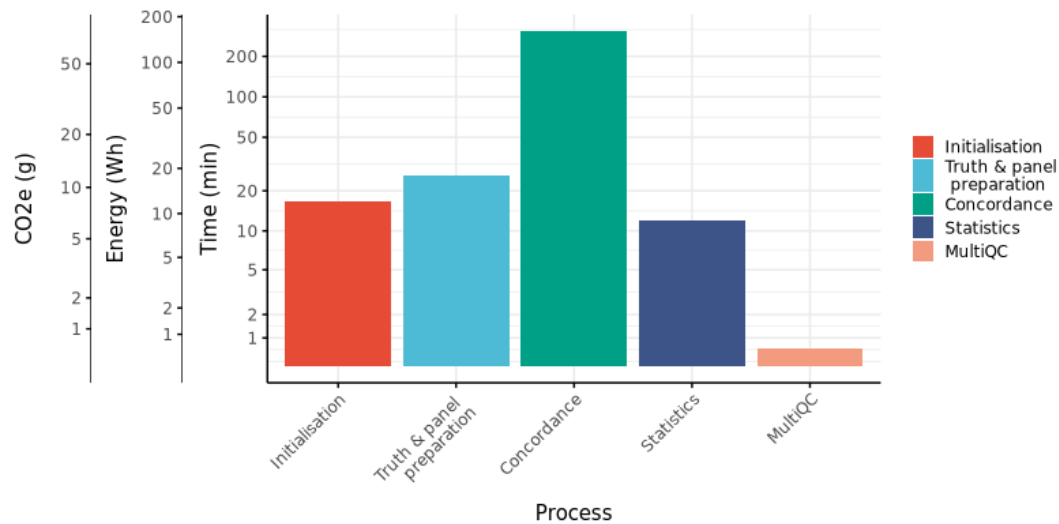


FIGURE S5 - Computational load for the concordance test by tasks for time (in minutes), energy consumed (in Watts) and CO_2 equivalent (in grams).

ÉTUDE GÉNÉTIQUE DE LA DYSPLASIE COXO-FÉMORALE

A Préparation des échantillons et séquençage *low-pass*

À partir de la base de données du projet DCF regroupant plus de 40 000 individus, nous avons sélectionné ceux présentant des données cliniques et un diagnostic de DCF validé et vérifié. Parmi ces individus, nous avons ensuite sélectionné ceux dont un prélèvement (sang dans la majorité des cas) était disponible au sein du CRB-CaniDNA ou au sein de la société *Antagene*, soit 980 échantillons des 5 races d'intérêt (i.e., Labrador et Golden Retriever, Cane Corso, Dogue de Bordeaux et Bouvier Bernois). L'ADN a ensuite été extrait à l'aide du kit *Macherey-Nagel NucleoSpin Blood*. La quantification des ADN a été réalisée par fluorométrie (*Qubit*) par séries de plaques de 96 puits.

Parmi l'ensemble des ADN extraits, 707 échantillons présentant une concentration suffisante ont été retenus. Ceux-ci ont été préparés sous forme d'aliquotes standardisées à 200 ng, lyophilisés, puis envoyées à la société *Psomagen* pour un séquençage *Low-Pass Whole Genome Sequencing* (LP-WGS) ciblant une couverture moyenne de 1× par échantillon.

Les échantillons ont été expédiés en trois lots distincts. Chaque plaque de 96 puits a été constituée en équilibrant au mieux les statuts cliniques (i.e., atteint *versus* indemne) pour limiter les biais expérimentaux. Lorsque le nombre d'échantillons le permettait, les races ont été réparties sur des plaques distinctes afin de limiter les effets de *batch*.

- le premier lot contenait une plaque unique composée de 96 Labrador (atteints et indemnes) ;
- le deuxième lot comprenait trois plaques : une contenant 96 Dogue de Bordeaux, une autre contenant 77 Cane Corso, complétée à 96 par des échantillons d'un autre projet et une troisième avec 87 Golden Retriever, également complétée de la même manière ;
- le troisième lot incluait quatre plaques : une avec 72 Bouvier Bernois et 24 Dogue de Bordeaux, une autre avec 96 Cane Corso, une troisième avec 51 Golden Retriever et 45 Labrador Retriever et enfin une dernière regroupant 21 Labrador Retriever et 42 chiens croisés Labrador x Golden.

A la réception des données de séquençage réalisées par la société *Psomagen*, malgré la standardisation des quantités d'ADN (200 ng par échantillon) et les contrôles qualité en amont (concentration par picogreen et qualité par migration des ADN bruts sur gel d'agarose 1 %), une hétérogénéité importante des couvertures génomiques a été observée entre les échantillons.

B Alignement des séquences et imputation sur le génome de référence *canFam4*

L'imputation des données issues des deux premières séries de séquençage a d'abord été confiée à la société *Gencove*, principal prestataire proposant ce service. Toutefois, cette plateforme ne permettait l'imputation que sur la version *canFam3* du génome de référence, alors même que *canFam4*, plus récent et mieux annoté, s'était déjà imposé comme le standard dans la communauté scientifique canine.

Entre-temps, nous avons développé notre propre pipeline d'imputation, *nf-core/phaseimpute* (section précédent). Ce développement nous a permis d'effectuer l'imputation directement sur *canFam4* sur le serveur de calcul de la plateforme *GenOuest* (i.e., service commun de l'Université de Rennes proposant des serveurs de stockage et de calculs : <https://www.genouest.org/>). J'ai donc réalisé l'imputation de nos données LP-WGS pour l'ensemble des individus à notre disposition pour ce projet (N = 707) sur le nouveau génome de référence, *canFam4*, permettant de ce fait d'accéder à un panel de référence plus complet ainsi qu'à une meilleure annotation du génome canin.

Les analyses présentées ci-dessous s'appuient exclusivement sur les données de séquençage générées spécifiquement dans le cadre de ce projet. L'alignement des séquences LP-WGS a été réalisé à l'aide du pipeline *nf-core/sarek* sur la version *canFam4* du génome de référence. L'imputation a ensuite été effectuée à l'aide de *nf-core/phaseimpute*, en utilisant le panel de référence du projet international *Dog 10 000 genomes* (N = 1929) (Meadows et al., 2023), comportant de nombreuses races, dont celles étudiées dans ce travail.

C Contrôles qualités des séquences LP-WGS obtenues

Les contrôles qualité initiaux obtenus via *nf-core/sarek* ont mis en évidence un déséquilibre dans la proportion des nucléotides au niveau des dix premières bases de chaque séquence. Plus précisément, les 10 premières bases ont montré une distribution non aléatoire des nucléotides qui diffère de la composition uniforme attendue (28 % A/T et 22 % G/C) (voir figure 17).

Après échanges avec l'entreprise de séquençage *Psomagen*, ce biais a été attribué à l'étape de fragmentation et d'insertion des barcodes réalisée à l'aide d'une enzyme de type transposase (kit *SeqWell LP384*). Cette enzyme induit des préférences d'insertion

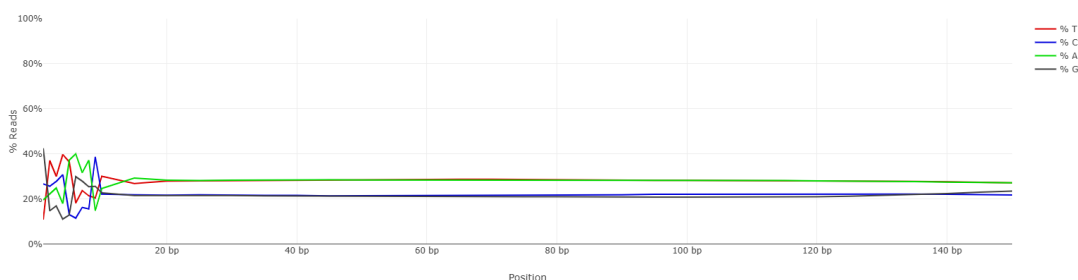


FIGURE 17 – Proportion de chaque nucléotide pour chaque position le long de la séquence lue

Les 10 premières bases montrent un biais systématique dans la composition en nucléotides, avec une distribution aléatoire des bases A, T, G et C.

spécifiques, entraînant un biais systématique au niveau des extrémités 5' des lectures. Ce profil, bien que résultant d'un artefact technique, ne correspond ni à une contamination par des adaptateurs ni à une erreur de séquençage : les bases concernées proviennent bien du génome de l'individu. Il n'a donc pas été jugé nécessaire de les retirer lors du traitement des données.

Nous avons ensuite vérifié que la couverture moyenne du génome était globalement homogène entre les individus atteints et indemnes, excluant ainsi un éventuel biais de séquençage lié au statut phénotypique. La concordance entre sexe déclaré et sexe génomique a également été contrôlée grâce à la méthode développée par le Dr Benoît Hédan, ingénieur de recherche dans l'équipe. Cette méthode est basée sur un indice de couverture normalisé du chromosome X. Cette étape a permis d'identifier six individus pour lesquels le sexe phénotypique renseigné ne correspondait pas au profil génomique observé. Ces données ont été vérifiées et les individus présentant une discordance confirmée ont été éliminés.

Des filtres de qualité ont ensuite été appliqués au niveau des variants mis en évidence et des individus, en utilisant les outils *bcftools* et *plink*. Après filtrage, le jeu de données final comprenait 7 534 827 variants et 663 individus, dont 287 cas de dysplasie coxo-fémorale et 376 contrôles (voir figure 18).

À partir de ce jeu de données (séquences LP-WGS de 663 chiens de 5 races), nous avons réalisé plusieurs analyses de structure génétique. Une analyse de type MDS (*Multidimensional Scaling*) a permis de vérifier la cohérence entre les données génétiques et les races déclarées (voir figure 19). Par ailleurs, un dendrogramme circulaire basé sur la matrice de proximité génomique entre tous les chiens séquencés (voir figure 20) a confirmé que les cas et les contrôles étaient bien répartis de manière homogène au sein de chaque race, limitant ainsi les risques de biais liés à la stratification génétique.

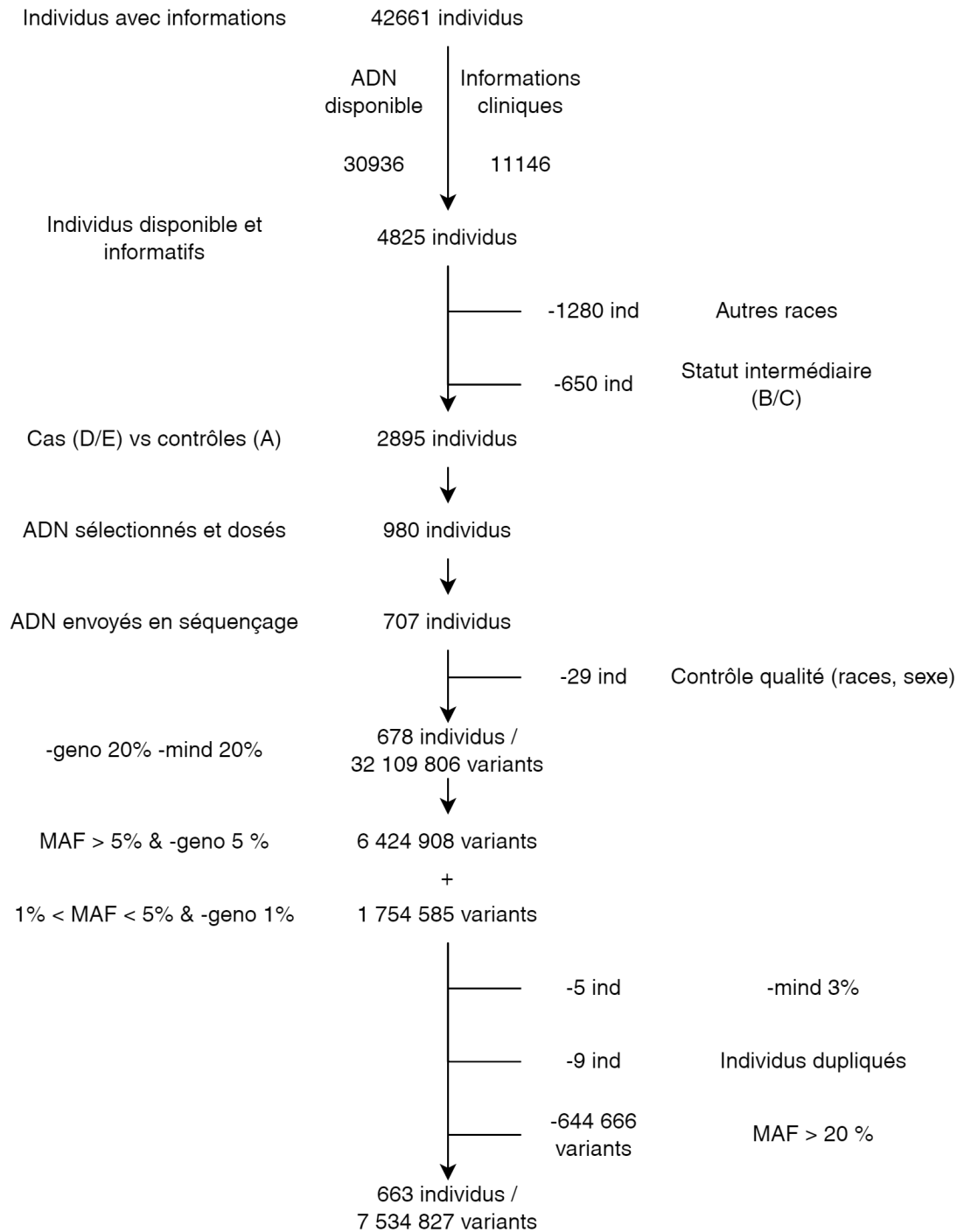


FIGURE 18 – Logigramme décisionnel de la sélection des individus à partir des échantillons du CRB Cani-DNA et de la société Antagene et des variants dans le cadre du projet DCF.

Abréviations : chiens A, B, C, D, E : Indices du statut dysplasique de la FCI (Fédération Cynologique Internationale); ind : individus; -geno et -mind : commande du logiciel plink filtrant respectivement les variants et les individus dont les taux d'appel manquants dépassent la valeur fournie; MAF : Minor Allele Frequency.

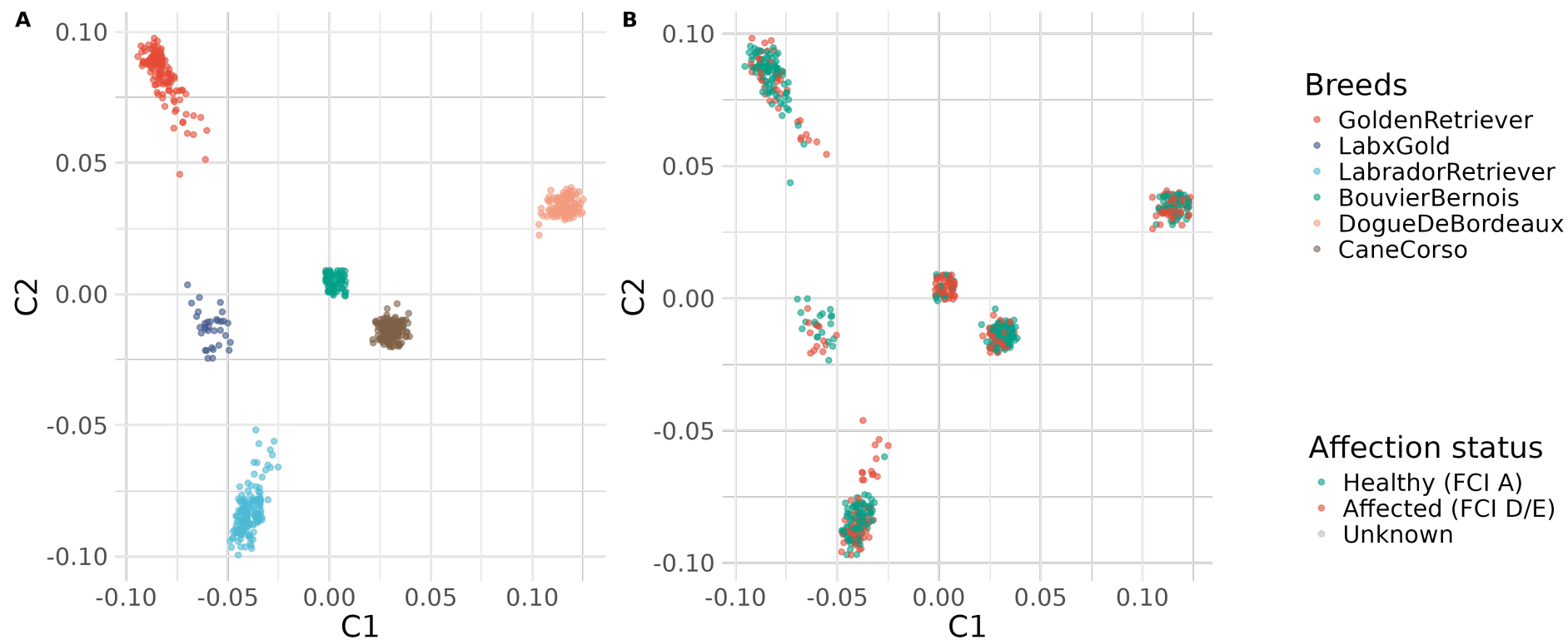


FIGURE 19 – Graphique MDS (*Multi-Dimensional Scaling*) de l'ensemble des échantillons par race et statut dysplasique comme défini par la Fédération Cynologique Internationale (FCI) pour les dimensions 1 et 2 et un ensemble de variants indépendants obtenus via le logiciel *plink* v1.9.

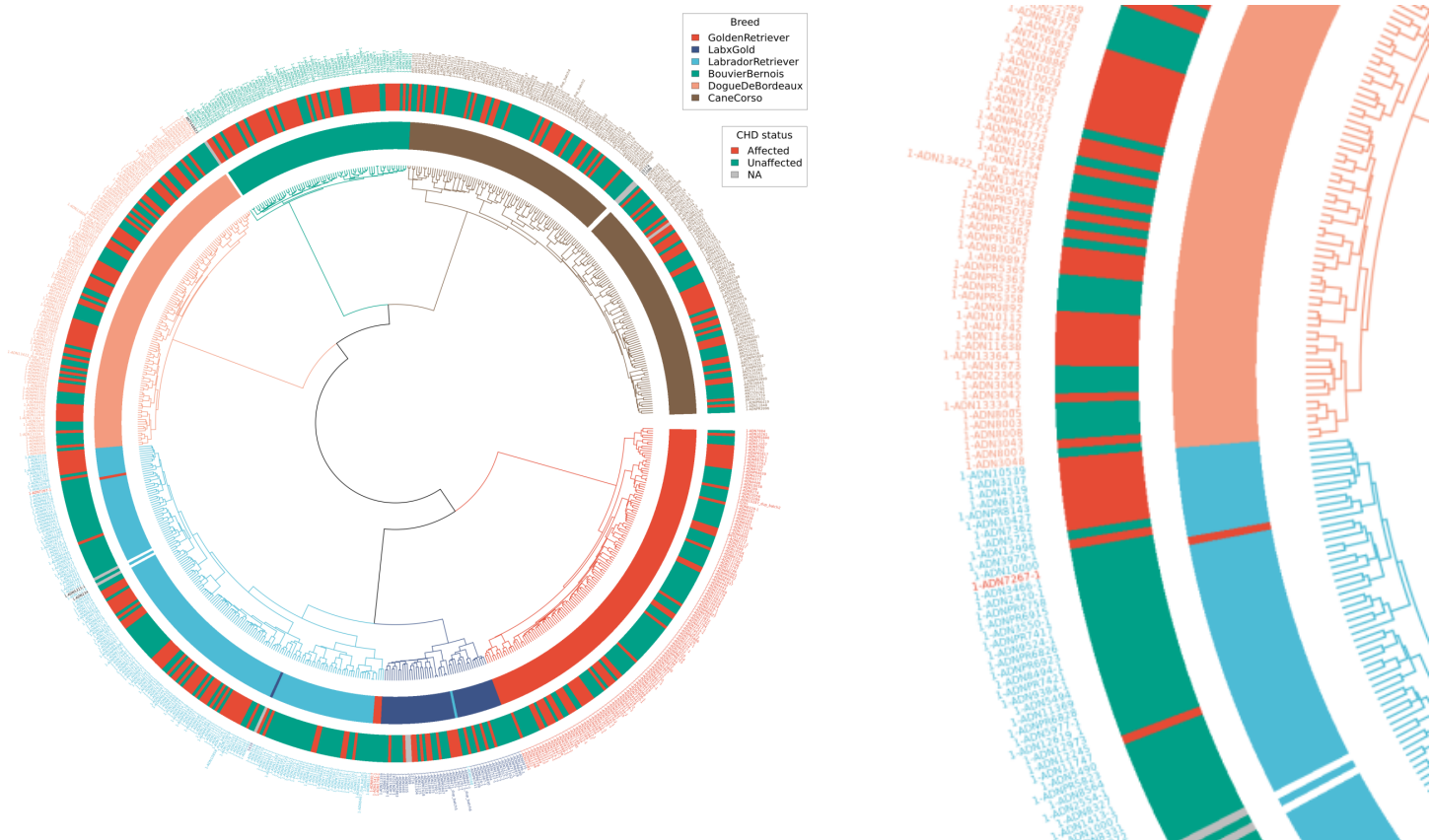


FIGURE 20 – Dendrogramme circulaire de proximité génomique entre les individus des différentes races (cercle intérieur) et des statuts cliniques (cercle extérieur).

Golden en rouge, mix Labrador x Golden en violet, Labrador en bleu, Bouvier Bernois en vert, Dogue de Bordeaux en orange et Cane Corso en marron. Le statut dysplasique des individus est représenté sur le cercle extérieur : rouge pour les atteints, vert pour les indemnes et gris pour ceux sans statut.

D Analyses d'association pangénomiques (GWAS, *Genome-Wide Association Study*)

Notre hypothèse de travail repose sur l'idée que les combinaisons de variants associés à la dysplasie coxo-fémorale (DCF) varient selon les races, bien que certaines races phylogénétiquement proches puissent partager certains variants communs. Toutefois, la taille d'effet de ces variants peut différer d'une race à l'autre, en fonction du contexte génétique propre à chaque population. Ainsi, un variant ayant un impact important dans une race donnée pourrait n'avoir qu'un effet modeste dans une autre, bien que présent. Cette hypothèse se base sur la forte structuration génétique des populations canines, conséquence directe de la sélection artificielle et des goulets d'étranglement spécifiques à chaque lignée (Quignon et al., 2007).

En conséquence, nous avons adopté une stratégie d'analyse en plusieurs étapes : nous avons d'abord réalisé des analyses par race, puis par groupes de races apparentées, en distinguant notamment les retrievers des molossoïdes, avant de procéder à une analyse globale regroupant l'ensemble des individus. Les Bouvier Bernois, dont nous ne disposons que de 72 échantillons, ont dans un premier temps été exclus de ces analyses. En effet, leur faible effectif, couplé à une structure de population très marquée, introduisait une variabilité excessive, réduisant la puissance et la robustesse des analyses inter-races.

Un des défis majeurs des études d'association pangénomiques est le choix du modèle statistique approprié. Celui-ci doit à la fois corriger la structure de population, source de faux positifs, et maintenir une puissance statistique suffisante pour détecter des associations réelles. Un indicateur clé pour évaluer la validité d'un GWAS est le facteur d'inflation génomique, noté λ_{GC} , défini comme le rapport entre la médiane de la distribution empirique des statistiques de test et la médiane attendue sous l'hypothèse nulle. Ce facteur permet d'estimer l'ampleur de l'inflation due à la structure génétique sous-jacente et le taux de faux positifs associé.

Pour contrôler cette inflation, plusieurs approches sont envisageables. La méthode la plus répandue consiste à introduire de façon incrémentielle des composantes principales (issues d'une ACP sur la matrice génomique) en tant que covariables, jusqu'à atteindre un λ_{GC} satisfaisant. D'autres approches consistent à intégrer directement des covariables catégorielles comme la race ou bien la matrice de proximité génomique (GRM, *Genome Relationship Matrix*).

Cependant, ces méthodes, bien que robustes en génétique humaine, peuvent s'avérer excessivement contraignantes dans des populations isolées comme celles des races canines. L'ajustement peut alors devenir trop strict, réduisant significativement la puissance de détection. Il est donc crucial d'adapter le modèle au cas par cas, en tenant

compte du déséquilibre de liaison, de la diversité intrarace et du degré d'apparentement des individus inclus dans chaque analyse.

1 Première approche

Dans un premier temps, notre objectif était de sélectionner, pour chaque combinaison de races, le modèle d'association le plus adapté tout en conservant une approche aussi objective que possible. Plutôt que de nous fonder sur une inspection visuelle des Manhattan plots, méthode subjective et sensible aux biais d'interprétation, nous avons comparé de manière systématique un ensemble de modèles statistiques couvrant un large spectre de complexité.

Pour chaque groupe (i.e., race seule ou combinaison de races), nous avons testé les modèles suivants :

- un test d'association allélique classique par chi-deux (χ^2) ;
- une régression logistique simple, sans covariable ;
- des modèles de type *Generalized Linear Model* (GLM) incluant entre 1 et 10 composantes principales (PC) pour corriger la structure de population ;
- un modèle GLM intégrant la race comme covariable fixe (lorsque plusieurs races étaient analysées ensemble) ;
- un modèle linéaire mixte (*Mixed Linear Model Association*, MLMA) implémenté dans l'outil *GCTA* (Yang et al., 2011) avec une matrice de proximité génétique (*Genomic Relationship Matrix*, GRM) comme effet aléatoire ;
- un modèle multi-locus (*Multi-Locus Mixed Model*, MLM) (Segura et al., 2012), utilisant également une GRM en covariable, qui ajoute itérativement les loci les plus significatifs comme covariables jusqu'à minimisation du critère bayésien étendu (eBIC).

Pour chacun de ces modèles et combinaisons, nous avons mesuré le facteur d'inflation génomique (λ_{GC}) et comptabilisé le nombre de SNPs significatifs après correction de Bonferroni (i.e., ajustement par le nombre de tests indépendants). Cette approche permettait d'objectiver le compromis entre contrôle des faux positifs et perte de puissance, sans se fonder uniquement sur l'apparence graphique des résultats. L'évolution du λ_{GC} en fonction de la complexité des modèles testés est illustrée dans la figure 21.

Les modèles basés sur la GRM, bien que très efficaces pour contrôler la structure de population, se sont révélés trop conservateurs dans plusieurs cas, filtrant potentiellement des signaux d'association réels. La GRM expliquant une large part de la variance génétique dans des populations aussi structurées que les races canines, les signaux d'association pouvaient être artificiellement atténués, voire complètement masqués.

Pour conserver une cohérence d'ensemble et un niveau d'interprétabilité suffisant, nous avons retenu un critère pragmatique : sélectionner, pour chaque combinaison de races, le modèle le plus simple dont l'inflation génomique restait inférieure à un seuil tolérable. Alors qu'un seuil de 1,05 est classiquement recommandé en génétique humaine, nous avons fixé un seuil plus permissif de 1,6 dans ce contexte, afin de ne pas exclure trop rapidement des signaux d'intérêt dans une affection potentiellement polygénique et dans un cadre de puissance limitée par le nombre de chiens analysés.

Les modèles retenus ont ensuite été soumis à une extraction des loci d'intérêt à l'aide de *Manhattan Harvester* avec les paramètres suivants : -inlimit 0.01 -peak-limit 6 -dots 10. Le paramètre -inlimit 0.01 applique un filtre préliminaire ne conservant que les SNPs présentant une p-valeur ajustée inférieure à 0,01, afin de réduire la charge de calcul sans influencer directement sur le seuil de significativité. Le paramètre -peak-limit 6 impose que les signaux retenus atteignent une hauteur minimale de 6 sur l'échelle $-\log_{10}(p)$ (soit $p \leq 10^{-6}$). Enfin, chaque pic devait être constitué d'au moins dix SNPs (-dots 10). Seuls les dix pics les plus significatifs par GWAS ont été retenus pour les analyses ultérieures.

Cet exercice a permis d'obtenir 18 loci différents répartis sur 10 chromosomes sur l'ensemble des 7 GWAS réalisés. Seul deux loci sur le chromosome 22 et 34 sont présents dans plus d'un GWAS, indiquant une très forte spécificité des régions identifiées (voir figure 22). Les SNPs les plus significatifs de chaque locus ont ensuite été annotés via la base *Ensembl v111*. Cette stratégie a permis d'identifier plusieurs régions génomiques contenant au total 68 gènes dont 15 ont déjà été identifiés comme associés à la dysplasie de la hanche chez l'humain ou chez le chien, dans de précédentes études (voir tableau 4).

Toutefois, cette approche de sélection modèle-par-modèle a montré ses limites. D'une part, elle introduisait une hétérogénéité dans les conditions d'analyse, rendant les résultats moins comparables entre races. D'autre part, les loci mis en évidence restaient largement stables entre modèles et seules les p-values et le bruit de fond variaient. Cela nous a conduit à adopter, dans une seconde approche, un modèle unique standardisé utilisé pour l'ensemble des analyses.

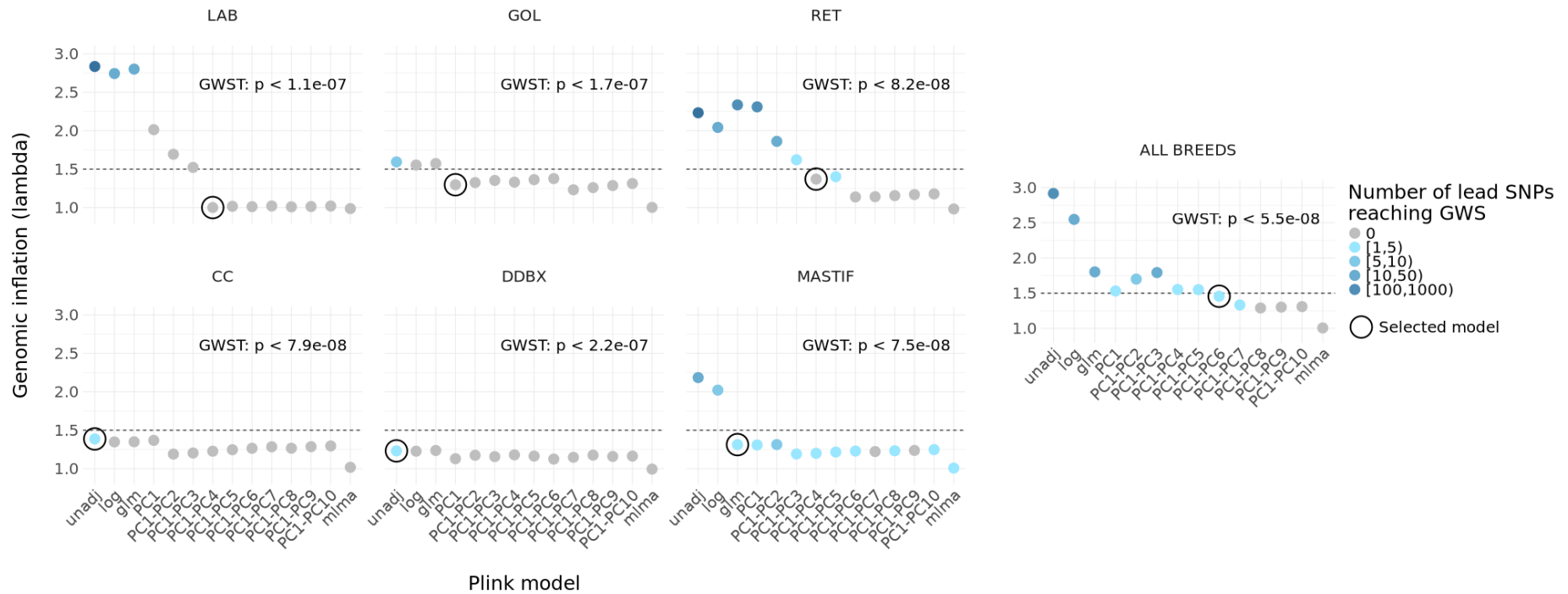


FIGURE 21 – Évolution de l'inflation génomique et du nombre de SNPs significatifs en fonction du modèle statistique utilisé pour chaque race et combinaison de races.

En abscisse : unadj = chi-deux allélique à 1 degré de liberté, log = régression logistique simple, glm = régression logistique avec la race des individus en covariable, PCX1-PCX2 = régression logistique avec en covariable les dimensions X1 à X2 de l'Analyse en Composante Principale, mlma = modèle linéaire mixte avec la matrice de relation génomique en covariable ; LAB = Labrador, GOL = Golden, RET= LAB, GOL et mix Labrador x Golden, CC= Cane Corso, DDBX = Dogue de Bordeaux, MASTIF = DDBX et CC, All breeds = toutes races confondues. Le modèle entouré pour chaque GWAS correspond au modèle le plus simple ayant une inflation génomique λ_{GC} en dessous du seuil fixé à 1,5 (ligne en pointillée). La couleur de chaque point correspond aux nombres des variants associés indépendants (i.e., SNPs sentinelles) sur une fenêtre de 100 000 bp dépassant le seuil de significativité génomique (i.e., Genome Wide Significance threshold, GWST).

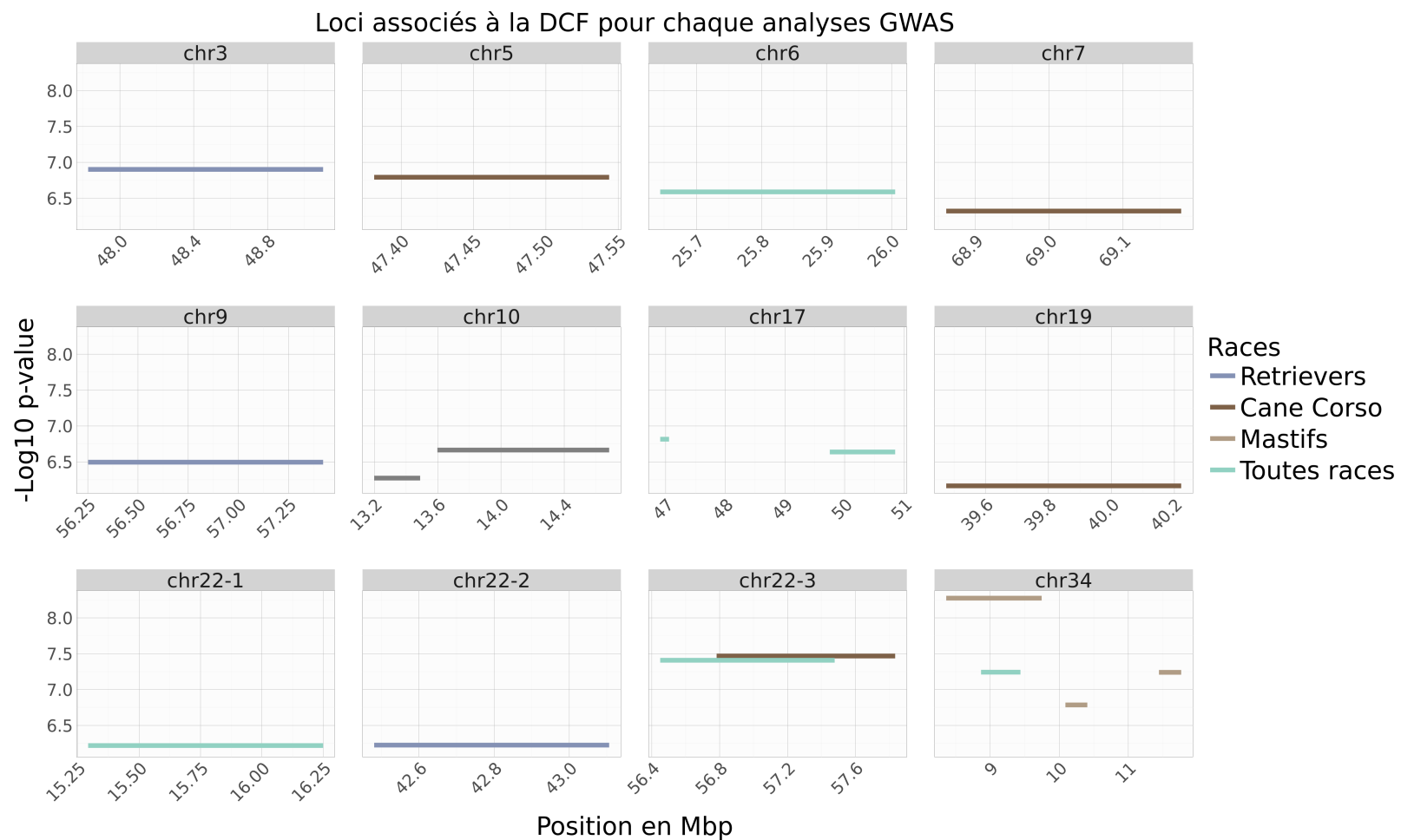


FIGURE 22 – Représentation des différents loci identifiés par race et combinaison de races par région chromosomique pour les GWAS retenus parmi tous les modèles statistiques.

RET= Labrador, Golden et mix Labrador x Golden, CC= Cane Corso, DDBX = Dogue de Bordeaux, MASTIF = DDBX et CC. En ordonnée est représenté la valeur de la p-value la plus basse pour le loci et l'abscisse représente la position du locus en millions de paires de bases sur le chromosome en question.

TABLE 4 – Loci génomiques associés à la Dysplasie Coxo-Fémorale chez le chien en fonction des races.

Inclut les races étudiées (RET= Labrador, Golden et mix Labrador x Golden, CC= Cane Corso, DDBX = Dogue de Bordeaux, MASTIF = DDBX et CC, All Breed = toutes races confondues), le nombre de cas et de contrôles, le nombre de SNP (dont indépendants), le modèle utilisé, les covariables utilisées, l'inflation génomique de l'analyse (λ_{GC}), les chromosomes et les positions des loci significatifs, le SNP le plus important pour chaque locus (Top SNP) et sa p-value associée et les gènes annotés situés dans les loci (ceux déjà associés dans la littérature chez le chien ou l'humain sont mis en gras avec l'article correspondant).

GWAS	Cases / Controls	nb SNP (ind)	Modèle & Covariable	λ_{GC}	Chr	Position en pb des loci	Top SNP	p-value max	Gènes dans la région
RET	112 / 203	6 516 758 (609 846)	GLM PC1-PC4	1.371	3	47827980-49101168	48690993:C:G	1.253E-7	FAM174B, ST8SIA2, SLCO3A1
					9	56251764-57422222	56587074:G:A	3.197E-7	RALGPS1 (Yan et al., 2019), ANGPTL2 (Cheng et al., 2021), ZBTB34, ZBTB43, LMX1B (Jacobsen et al., 2023), MVB12B, PBX3
					22	42480497-43106243	42789862:T:C	5.940E-7	GPC5 (Binversie et al., 2022a)
CC	63 / 97	6 488 265 (630 680)	Chi-deux	1.386	5	47381283-47543698	47507861:G:A	1.612E-7	
					7	68860634-69178731	68960838:T:A	4.774E-7	
					19	39475433-40221584	39548591:G:A	6.833E-7	R3HDM1 , MCM6 (Yan et al., 2019), ZRANB3, UBXN4, LCT, DARS1
DDBX	60 / 57	4 633 418 (230 200)	Chi-deux	1.231	22	56782321-57834954	57104724:T:G	3.406E-8	FAM155A (Cheng et al., 2021), LIG4, ABHD13, TNFSF13B, MYO16
					10	13198803-13489196	13254148:G:A	5.328E-7	LGR5 (Binversie et al., 2022a)
					10	13598191-14683685	14642775:G:A	2.166E-7	THAP2, TMEM19, RAB21 , TBC1D15 , TPH2 , TRHDE (Binversie et al., 2022a)
MASTIF	123 / 154	6 710 858 (663 032)	GLM Breed (CC/DDBX)	1.313	34	10089438-10406636	10323918:A:T	1.635E-7	IRX2
					34	11443842-11765806	11721111:T:C	5.747E-8	TERT, SLC6A18, SLC6A19, SLC12A7, NKD2, TRIP13, BRD9
					34	8364146-9745902	8989195:A:G	5.316E-9	ADAMTS16, IRX1
Toutes races	235 / 357	7 391 561 (900 954)	GLM PC1-PC6	1.456	6	25644070-26005377	25836338:C:G	2.578E-7	GPRC5B, IQCK, KNOP1, VPS35L, CCP110, GDE1, TMC5
					17	46908111-47056223	47043772:G:A	1.529E-7	LRRTM4 (Todhunter et al., 2019)
					17	49751794-50847566	50318444:A:G	2.303E-7	DUSP11, TPRKB, NAT8, ALMS1, EGR4, FBXO41, CCT7, PRADC1, SMYD5, NOTO, RAB11FIP5, SFXN5, EMX1, SPR, EXOC6B
					22	15292288-16249397	15306904:C:T	6.024E-7	DIAPH3 (Todhunter et al., 2019)
					22	56449805-57478341	56848318:G:C	3.903E-8	FAM155A (Cheng et al., 2021), LIG4, ABHD13, TNFSF13B, MYO16
					34	8868304-9438466	8989195:A:G	5.718E-8	

2 Seconde approche

Nous avons ainsi adopté une stratégie non paramétrique afin d'estimer directement la distribution nulle des statistiques de test. Pour chaque combinaison de races, nous avons réalisé un million de permutations Monte-Carlo du phénotype dans *plink* avec l'option `--mperm`. Cette procédure fournit pour chaque variant une p-value empirique : elle correspond à la proportion de permutations où la statistique observée est au moins aussi extrême que celle attendue sous l'hypothèse nulle.

L'avantage est double :

- le bruit de fond lié à la structure de population est fortement atténué ;
- le seuil de significativité s'adapte automatiquement à la complexité du jeu de données, sans choix arbitraire de λ_{GC} .

Après permutation, nous avons appliqué la même chaîne d'analyse : extraction automatique des régions associées à la maladie (DCF) avec *Manhattan Harvester* mais cette fois ci avec les options suivantes : `-dots 5 -inlimit 0.5 -peak-limit 1.3`. Les variants les plus significatifs ont ensuite été extraits de ces loci, puis la région a été annotée (*Ensembl v111*).

Les résultats présentés dans l'article dans les pages suivantes reposent exclusivement sur cette seconde approche, qui offre un compromis robuste entre contrôle du taux de faux positifs et puissance de détection.

A partir des données de bibliographie collectées pour la revue comparative de la dysplasie de la hanche entre l'humain et le chien, j'ai représenté les régions chromosomiques que nous avons identifiées sur le caryotype canin afin d'observer la colocalisation des loci identifiés dans notre analyse et ceux de la littérature chez l'humain et chez le chien. Tous ces loci sont représentés sur la figure 23 ainsi que la littérature associée en figure 24.

Quatre situations sont observées :

- confirmation de loci déjà identifiés chez l'humain et le chien : CFA9, CFA22, CFA30, CFA36 ;
- confirmation de loci uniquement identifiés chez l'humain : CFA6, CFA23, CFA34 ;
- confirmation de loci uniquement identifiés chez le chien : CFA3, CFA17 ;
- découverte de nouveaux loci chez le chien : CFA4, CFA10, CFA11, CFA16, CFA18, CFA31, CFA38.

Les loci ayant déjà été identifiés par des études précédentes sont très prometteurs et ceux nouvellement découverts nécessiteront une attention toute particulière afin de les confirmer. Il faudra déterminer leur implication réelle ou bien savoir s'ils sont artéfactuels.

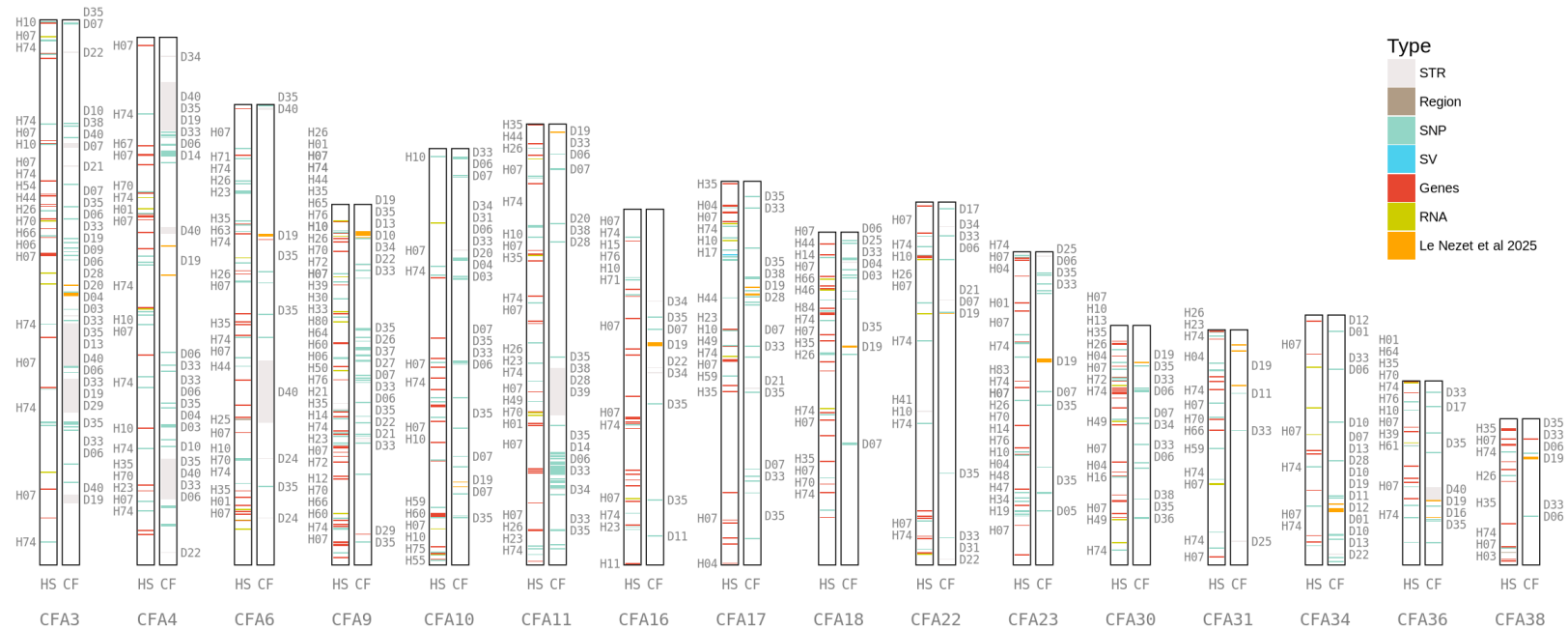


FIGURE 23 – Représentation schématique des loci impliqués dans la DCF sous forme de caryotype canin pour la littérature humaine (HS) et canine (CF).

Sont représenté les loci identifiés lors de la thèse (en orange) ainsi que l'ensemble des régions chromosomiques identifié dans la littérature chez le chien et chez l'humain (après lift-over sur le génome canin canFam4). La couleur des loci correspond au type de variants identifiés : STR = microsatellite, région chromosomique, SNP = Polymorphisme mono-nucléotidique, SV = variant de structure, gène, RNA = Acide RiboNucléique, D19 (orange) = fait référence à ce travail de thèse (Le Nézet et al 2025).

Bibliography

Dog

Human

D01	Akis et al 2020	D21	Liu et al 2007	H01	Aki et al 2018	H22	Harsanyi et al 2021	H43	Miyamoto et al 2007	H64	Wang et al 2008
D02	Alilovic et al 2023	D22	Marschall et Distl 2007	H02	Baghdadi et al 2019	H23	Hatzitokoulas et al 2018	H44	Mori et al 2023	H65	Wang et al 2010
D03	Ardicli et al 2024	D23	Martin 2018	H03	Basit et al 2017	H24	Igrek et al 2021	H45	Mototani et al 2005	H66	Wang et al 2020
D04	Ardicli et al 2025	D24	Mateescu et al 2005	H04	Basit et al 2018	H25	Ingvarsson et al 2001	H46	Ning et al 2019	H67	Wang et al 2022
D05	Atalay et Kok 2022	D25	Mateescu et al 2008	H05	Cengic et al 2015	H26	Jacobsen et al 2023	H47	Qiao et al 2017	H68	Xu et al 2016
D06	Bartolome et al 2015	D26	Mikkola a et al 2019	H06	Chen et al 2019	H27	Jawadi et al 2018	H48	Qiao et al 2017b	H69	Xu et al 2020
D07	Binversie et al 2022	D27	Mikkola b et al 2019	H07	Cheng et al 2021	H28	Ji et al 2020	H49	Rao et al 2025	H70	Xu et al 2021
D08	Chase et al 2004	D28	Mikkola et al 2021	H08	Da Silva et al 2009	H29	Jia et al 2012	H50	Rouault et al 2009	H71	Xu et al 2021b
D09	Chase et al 2005	D29	Padula et al 2023	H09	Dai et al 2008	H30	Jiang et al 2003	H51	Rouault et al 2010	H72	Xu et al 2022
D10	Distl et al 2009	D30	Pfahler et Distl 2012	H10	Dembic et al 2022	H31	Kapoor et al 2007	H52	Rubini et al 2008	H73	Yamanaka et al 2009
D11	Fealey et al 2017	D31	Phavaphutanon et al 2009	H11	Demirel et al 2023	H32	Kolundzic et al 2011	H53	Sadat-Ali et al 2018	H74	Yan et al 2019
D12	Fels et Distl 2014	D32	Quignon_ACGAO 2016	H12	Ding et al 2021	H33	Li et al 2010	H54	Sarikaya et al 2024	H75	Yan et al 2022
D13	Fels et al 2014	D33	Sanchez-Molano et al 2014	H13	Eytan et al 2014	H34	Li et al 2017	H55	Sekimoto et al 2013	H76	Yang et al 2022
D14	Friedenberg et al 2011	D34	Todhunter et al 2005	H14	Feldman et al 2013	H35	Li et al 2021	H56	Sekimoto et al 2017	H77	Yilmaz et al 2015
D15	Hayward et al 2016	D35	Todhunter et al 2019	H15	Feldman et al 2019	H36	Liu et al 2014	H57	Shi et al 2011	H78	Zhang et al 2018
D16	Huang et al 2017	D36	Topparmak et al 2025	H16	Ghosh et al 2009	H37	Lu et al 2023	H58	Shi et al 2014	H79	Zhang et al 2020
D17	Kieler et al 2024	D37	Wang et al 2021	H17	Gombosh et al 2024	H38	Luo et al 2021	H59	Sun et al 2015	H80	Zhao et al 2013
D18	Lavrijsen et al 2014	D38	Zhou et al 2010	H18	Granchi et al 2002	H39	Ma et al 2005	H60	Sun et al 2019	H81	Zhao et al 2013b
D19	Le Nezet et al 2025	D39	Zhu et al 2008	H19	Gumus et al 2021	H40	Ma et al 2017	H61	Tian et al 2012	H82	Zhao et al 2017
D20	Lee et al 2017	D40	Zhu et al 2012	H20	Gumus et al 2022	H41	Mabuchi et al 2006	H62	Topak et al 2021	H83	Zhu et al 2011
				H21	Hao et al 2014	H42	Meng et al 2024	H63	Torisu et al 1993	H84	Zhu et al 2019

FIGURE 24 – Ensemble des publications de la littérature utilisées pour représenter les régions chromosomiques (loci) impliquées dans la dysplasie de la hanche chez le chien et l’humain.

ARTICLE 5

“ Identification of loci predisposing to canine hip dysplasia in four dog breeds using imputed genotypes from low-pass sequencing ”

AUTEURS : Louis Le Nézet, Lauranne Machou, Benoit Hédan, Richard Guyon, Catherine André, Pascale Quignon

STATUT : En préparation et en relecture par les co-auteurs

JOURNAL : Publication prévu en tant qu'article dans *Plos Genet*

Identification of loci predisposing to canine hip dysplasia in four dog breeds using imputed genotypes from low-pass sequencing

Louis Le Nézet^{1,*}, Lauranne Machou¹, Benoit Hédan¹, Richard Guyon¹, Catherine André¹, Pascale Quignon^{1,*}

¹ UMR-6290, IGDR (Institute Genetics and Development of Rennes), University of Rennes, CNRS, Rennes, France

* Corresponding author. louislenezet@gmail.com pascale.quignon@univ-rennes.fr

1 Abstract

Canine hip dysplasia (CHD) is an abnormal development of the hip joint due to an excessive laxity between the acetabulum and the femoral head leading to osteoarthritis with age. CHD is homologous to the developmental dysplasia of the hip (DDH) in human, and its development is influenced by the environment as well as multiple genetic factors that still need to be elucidated. Dog breeds offer a rich and spontaneous model to study the genetic basis of such a complex disease. In this study we combined several genome wide association studies using more than 600 dog DNA samples from five dog breeds sequenced with the low-pass technique at 1× of coverage followed by genotype imputation using *nf-core/phaseimpute*. This methodology allowed us to validate genes already described in human DDH and in previous canine studies as well as identifying new promising genes.

2 Introduction

Canine Hip Dysplasia (CHD) is an abnormality in the development of the hip joint in dogs caused by excessive laxity of the joint, resulting in poor coaptation of the bones. It then leads to a deformation of the joint profiles and the development of osteoarthritis in later age (Fries & Remedios, 1995; King, 2017). It is considered as the most prevalent orthopaedic condition in dogs (Smith, 1998). Moreover, various dog breeds are severely affected with prevalence up to 50 % of a breed population in Cane Corso for example (Baldinger et al., 2020; Genevois et al., 2020). This

condition entails inflated costs for the organisations that train working dogs (guide dog schools for example) and impact the well-being of affected dogs. CHD is a complex, multifactorial disease with several genes and environmental factors (e.g., nutrition, weight, exercise) involved in its development (Fries & Remedios, 1995; King, 2017).

The disease is evaluated from a dorso-ventral X-ray and consider radiographic evidence (i.e., Norberg-Olsson angle, presence of osteophyte) as well as clinical evidence. The diagnostic is established at one-year-old for most of the breeds and evaluated through a scale from

A (i.e., unaffected) to E (i.e., strongly affected) by the Fédération Cynologique Internationale (FCI) in Europe.

CHD shares similarities with Developmental Dysplasia of the Hip (DDH) in human, both clinically and orthopedically, and share similar diagnosis and treatments (Willemsen et al., 2021). Furthermore, the dog species is a spontaneous and ethical model with characteristics that facilitate the discovery of the genetic basis of diseases. Dogs have shorter generation time; their information and samples are easier to obtain; they share the same environment as humans, and the diseases prevalence tends to be higher due to population structure and artificial selection (Pallotti et al., 2022). Dogs are particularly valuable for studying complex multigenic diseases due to intra-breed genetic homogeneity with high linkage disequilibrium and inter-breed genetic heterogeneity (Sutter et al., 2004). Indeed, each canine breed constitutes a genetic isolate with unique combinations of major and minor genes, which makes it possible to identify multiple genetic regions involved in disease (Hayward et al., 2016; Karlsson & Lindblad-Toh, 2008; Lequarré et al., 2011). In addition, we hypothesize that some of these genetic regions would be shared between CHD and DDH in human. In recent years, numerous studies in dog succeed to discover strong effect loci in complex diseases and phenotypes as compared to human (Binversie et al., 2022b; Meeson et al., 2019; Ostrander et al., 2008; Sutter & Ostrander, 2004).

Since the beginning of the 21st century, more than forty papers have been published on CHD genetics and some of these results were independently validated by Mikkola et al. (2021). Despite the numerous studies, no gene seem to make a consensus. This can be explained by many factors such as the difference in genetic data acquisition techniques, analysis methodology or the breeds studied and the precise phenotype analysed.

Due to the complexity of the disease a high number of individuals is necessary to discover the major genes involved in its development. A hundred affected and a hundred unaffected individuals have been set as the minimum requirement to detect major genes conferring a 3-to-5-fold increased risk in complex disease (Karlsson & Lindblad-Toh, 2008). However, this estimate was based on a Single Nucleotide Polymorphism (SNP) array density of 15 000 SNPs. Nowadays, modern technology allows to sequence, for a similar cost, the whole genome at around 1 × of coverage (i.e., low-pass sequencing) and obtain millions of genotypes. The SNP density gain coupled to an across-breed design as well as an increased number of affected and control individuals allows to highlight predisposing genes with moderate associated risk (Hayward et al., 2016).

However, the usage of low-pass sequencing techniques introduces a critical challenge: the accurate reconstruction of haplotypes and the imputation of missing genetic data generated by this

new sequencing strategy (Wragg et al., 2024). This challenge becomes particularly pronounced when dealing with species lacking dedicated phased panels. Accurate haplotype reconstruction is a fundamental step to build a reliable reference for subsequent imputation, thereby improving the comprehensiveness of genetic information (Buckley et al., 2021; Wragg et al., 2024).

The aim of this study is to contribute to the broader effort to understand the genetic and biological underpinnings of CHD, by both validating previously reported genes and identifying novel loci associated with the condition. Adopting a One Health perspective, we also aim to compare our findings with those from studies on developmental dysplasia of the hip (DDH) in humans, to deepen the cross-species understanding of this complex trait. In this work, the genetic basis of CHD was assessed by comparing affected (FCI grades D and E) and unaffected (FCI grade A) dogs across four breeds. Low-pass whole-genome sequencing followed by genotype imputation was used to generate the genetic dataset. Genome-wide association studies (GWAS) were then conducted within each breed and across grouped breeds to ensure a comprehensive analysis. These analyses not only confirmed loci previously reported in the canine and human literature but also uncovered novel loci not yet described.

3 Materials and Methods

3.1 Ethic statement

All dog blood samples were collected by DVM veterinarians, in the course of their medical care with the owner consent, and submitted to the Cani-DNA Biological Resource centre (BRC) at the University of Rennes (André et al., 2022), under the ethic statement and approval of the CREEA (Comité Rennais en Ethique et Expérimentation Animale): CE07-2020-11-CA, date of approval 23 November 2020.

3.2 Sample collection and sequencing

All blood samples were registered, stored and processed by the Cani-DNA BRC. In this study, only dogs from the following four strongly affected breed were selected: Labrador and Golden retriever as well as Cane Corso and Dogues de Bordeaux. Dogs with an FCI grade of A was considered as unaffected and dogs with an FCI grade of D or E as affected. Intermediary grades were not considered. The genomic DNA was isolated using the NucleoSpin Blood L Kit (Macherey-Nagel) according to the manufacturer's instructions. Each DNA sample was then quantified using the Quant-iT™ 1X ds-DNA BR Assay Standards kit (Invitrogen) and the DNA quantity was standardized to 200 ng before being dried in a vacuum centrifuge at 55 °C for 30 minutes. A total of 635 blood sample was sent for sequencing to *Psomagen* company (USA) in the following three batches (balanced

half affected, half unaffected per 96 well plate) on the NovaSeqX sequencer for a mean genome coverage of 1×.

3.3 Data processing

The sequence reads (i.e., fastq) were then aligned to the *canFam4* reference genome Wang et al., 2021 using the *nf-core/sarek* pipeline (v3.4.4) Garcia et al., 2020 and the *bwa-mem2* software (v2.2.1) Vasimuddin et al., 2019. Sequencing adaptors was trimmed using *fastp* (v0.23.4) Chen et al., 2018 and duplicated sequence was marked using the *markduplicates* software (GATK4 v4.5.0) McKenna et al., 2010. Sequence quality controls were performed with the pipeline using the *fastqc* software (v0.12.1) Andrews, 2010.

Genome coverage was performed using *mosdepth* tool Pedersen & Quinlan, 2018. A Linear Mixed Model (LMM) was fitted to assess whether genome coverage differs by CHD status, while accounting for sequencing batch effects and breed variation. The model included the CHD status as a fixed effect and both batches and breed as random effects.

For each individual i , a Normalized X Coverage Index (NXCI) was computed as x_{ij} is the mean genome coverage of the non-pseudo-autosomal region of the X chromosome of individual i for each j variants, M_i is the median coverage across individuals for each position of the genome and u_j is the mean genome coverage of the individual across all position.

$$E_i = \log \left(\frac{x_{ij}}{M_i \cdot u_j} \right)$$

The bam files were then imputed using the *nf-core/phaseimpute* pipeline (v1.0.0). To do so, the reference panel of the *Dog 10 000 genomes* consortium Meadows et al., 2023 was used. The variants for the panel were first normalized and only the biallelic SNPs were kept for imputation using *Glimpse2* software Rubinacci et al., 2021.

3.4 GWAS analyses

3.4.1 Variant and sample filtering

Before performing the GWAS, the following filters have been applied to the variants obtained by *nf-core/phaseimpute* using the *plink* software (v1.90b6.21). First every genotype probability less than 70 % was set to missing value (i.e., “./.”). Variant identifiers were normalized, and duplicate variants with inconsistent reference alleles were removed. Individuals and variants with more than 20 % missing genotype data were excluded (21 350 variants removed). Variants were then filtered based on Minor Allele Frequency (MAF): common variants (MAF > 5 %) were retained if their call rate exceeded 95 % (62 078 variants removed), while rare variants (1 % < MAF < 5 %) were included if their call rate exceeded 99 % (19 3129 variants removed). These two variant sets were then merged, and additional filtering was applied to retain only individuals with less than 3 % missing genotypes in the final dataset (14 dogs removed).

3.4.2 Breed assessment

To assess population structure and evaluate the clustering of samples by breed, multidimensional scaling (MDS) was performed. Prior to this, linkage disequilibrium (LD) pruning was applied using *plink* (v1.90b6.21) with a 50-SNP window, a 5-SNP step size, and an r^2 threshold of 0.5 (--indep-pairwise 50 5 0.5), resulting in a pruned dataset of 1 139 910 SNPs.

3.4.3 GWAS combination

Genome-wide association studies (GWAS) were conducted in a multi-level approach: first within each breed separately, then within breed groups - Retrievers (Labrador Retrievers and Golden Retrievers) and Mastiffs (Cane Corso and Dogues de Bordeaux) - and finally across all individuals.

To determine the most appropriate correction for population structure in each analysis, a series of increasingly complex models was tested. These models included unadjusted logistic regression, logistic regression with breed as a covariate, models incorporating the first 1 to 10 principal components (PCs) from a prior principal component analysis (PCA), and a mixed linear model using a genomic relationship matrix (GRM).

GWAS for each combination were made using a 1 df chi-square allelic test and false positive result was mitigated using one million Monte Carlo phenotype permutations with the *plink* software. The $\max(T)$ adjusted p-value was then used for the Manhattan plot as well as to define the loci using *Manhattan harvester* Haller

et al., 2019 with the parameters -dots 5 -inlimit 0.5 -peak-limit 1.3. Since we used permutation-based $\max(T)$ adjusted p-values, the threshold for noise control was already enforced, making these values more conservative than raw p-values. The -inlimit 0.5 parameter defines a preliminary filter that retains only SNPs with adjusted p-values below 0.5 to reduce computational load and is not used to determine significance. Peaks were only retained if they reached a height of at least 1.3 on the $-\log_{10}(p)$ scale (-peak-limit 1.3, corresponding to $p = 0.05$). Additionally, each peak had to consist of at least five SNPs (-dots 5). Only the top 10 peaks per GWAS were retained for downstream analyses.

3.4.4 Gene discovery and literature comparison

The variants from the above-mentioned loci for each GWAS were selected and annotated using the *snpEff* software (v4.3) Cingolani et al., 2012 and GO pathway enrichment was performed for each gene discovered. The comparison of regions and genes in both species was made through a database regrouping results from CHD and DDH genetic studies.

4 Results

4.1 Sequencing quality checks

Of the 635 samples initially submitted for sequencing, 12 failed the library preparation quality control and were excluded. Following the removal of individuals with excessive missing genotype data and

subsequent quality control filtering, a total of 609 dogs were retained for downstream analyses (see Table 1).

TABLE 1 - Sample distribution by breed and affection for Hip Dysplasia

FCI A are unaffected and FCI D/E are severely affected. FCI: Fédération Cynologique Internationale.

Group	Breed	FCI A	FCI D/E	Total
Retriever	Labrador	99	56	155
	Golden	92	46	138
	Lab×Gol	25	12	37
	Total	216	114	330
Mastiff	Cane Corso	97	64	161
	Dogue de Bor-	58	60	118
	deaux			
	Total	155	124	279
Total		371	238	609

4.2 Coverage by affection status, breed and batches

The mean genome coverage across all samples was $1.2 \times \pm 0.4 \times$ (mean $\pm$ SD), ranging from $0.2 \times$ to $2.1 \times$ (see Figure 1, top left) The distribution was normal (Shapiro-Wilk test: $W = 0.997$, p -value = 0.38). Diagnostic checks for the LMM confirmed that model assumptions were met Residuals were evenly scattered around zero (see Figure 1, top right), with no discernible pattern, indicating no systematic bias in model predictions. The Q-Q plot (see Figure 1, bottom left) showed that residuals closely followed a normal distribution, with minor deviations in the tails. The Scale-Location plot (see Figure 1, bottom right) suggested constant variance across fitted values,

ruling out heteroscedasticity. Given these results, no transformations were required. The fixed effect of CHD status was not statistically significant ($\beta = -0.025$, $SE = 0.026$, $p = 0.35$), indicating no measurable difference in genome coverage between CHD groups. The random effects for sequencing batch and breed contributed 23 % and 27 % of the variance, respectively, highlighting their influence on coverage variability (see Figure 2).

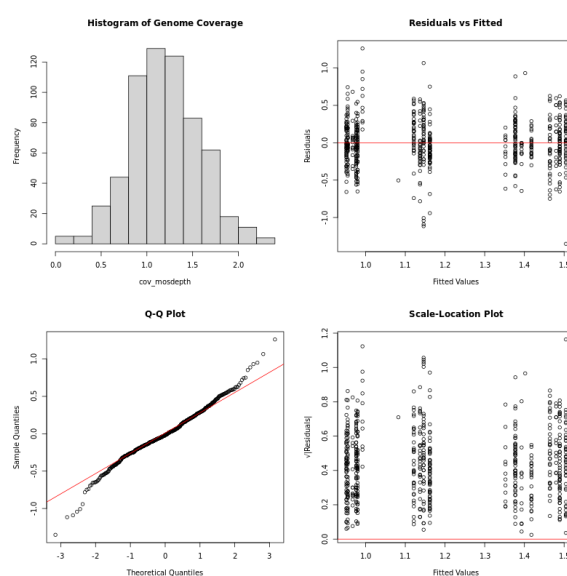


FIGURE 1 - Diagnostic plots evaluating the fit of the model predicting mean sequencing coverage across samples.

Top left: Histogram of residuals showing distribution of model errors. Top right: Residuals vs Fitted values, assessing linearity and homoscedasticity. Bottom left: Q-Q plot of standardized residuals, evaluating normality. Bottom right: Scale-Location plot, checking for homoscedasticity (constant variance).

4.3 Sex check

From the 623 individuals sequenced 389 were labelled as female, 225 as male, 4 as unknown and 5 as unsure. With thresholds set to between 0 and 0.5 for female and -0.5 and -1 for male the mean genomic sex ratio enables to

obtain the 9 unknown sex (5 females and 4 males). Three individuals labelled as female were male and three labelled as male were female. The threshold also allowed to identify one individual outside the accepted range due to low coverage (i.e., 0.15) (see Figure S1).

4.4 Variants and quality controls

After the different filtering steps, the final dataset contains 8 176 478 variants and 609 dogs with 238 cases and 371 controls.

Multidimensional scaling plots generated from the pruned dataset (1 139 910 SNPs), showed clear clustering by breed, with no evident separation between cases and controls (see Figure 3).

4.5 GWAS

For each breed group, a minor allele frequency (MAF) filter of 2 % was applied, resulting in SNP sets ranging from 7.4 million variants (all breeds combined) to 4.6 million SNPs (Dogue de Bordeaux). To estimate the number of independent tests, linkage disequilibrium (LD) pruning was performed and the number of pruned SNPs ranged from 294 000 (Golden Retrievers) to over 900 000 (all breeds combined).

Genome-wide association studies (GWAS) were conducted using a permutation-based approach to control false negative results, yielding 36 distinct loci across the different combinations: 10 in Labradors, 10 in the broader Retriever group, 2 in Dogue de Bordeaux, 4 in the Mastiff group, and 10 in the full multi-breed

cohort. These loci span a total of 57 genes, 11 of which have previously been associated with hip dysplasia in either canine or human studies. A summary of all GWAS results, including case/control numbers, SNP counts (with pruned SNPs in parentheses), significant loci, lead SNPs with p-values, and annotated genes is provided in Table 5.

A representative Manhattan plot is shown in Figure 4 for the Labrador GWAS, highlighting a notable locus on chromosome 9. This region overlaps with several genes implicated in developmental dysplasia of the hip (DDH) in human: *RALGPS1*, *ANGPTL2* and *LMX1B* Cheng et al., 2021; Jacobsen et al., 2023; Yan et al., 2019.

To further investigate the consistency of GWAS signals across breeds, we computed the minor allele frequency (MAF) difference between cases and controls for the top 10 SNPs from each GWAS. The results, presented as a heatmap (see Figure S2), reveal that several top SNPs show breed-specific effects, while others demonstrate more consistent allele frequency shifts across related breeds (e.g., between Labradors and Golden Retrievers within the Retriever group). Notably, both the Labrador and Retriever GWAS identified the SNP chr9:56598550:T:C as a top signal, with a consistent overrepresentation of the minor allele in affected individuals across all breed groups. This suggests a potentially shared genetic basis for hip dysplasia at this locus.

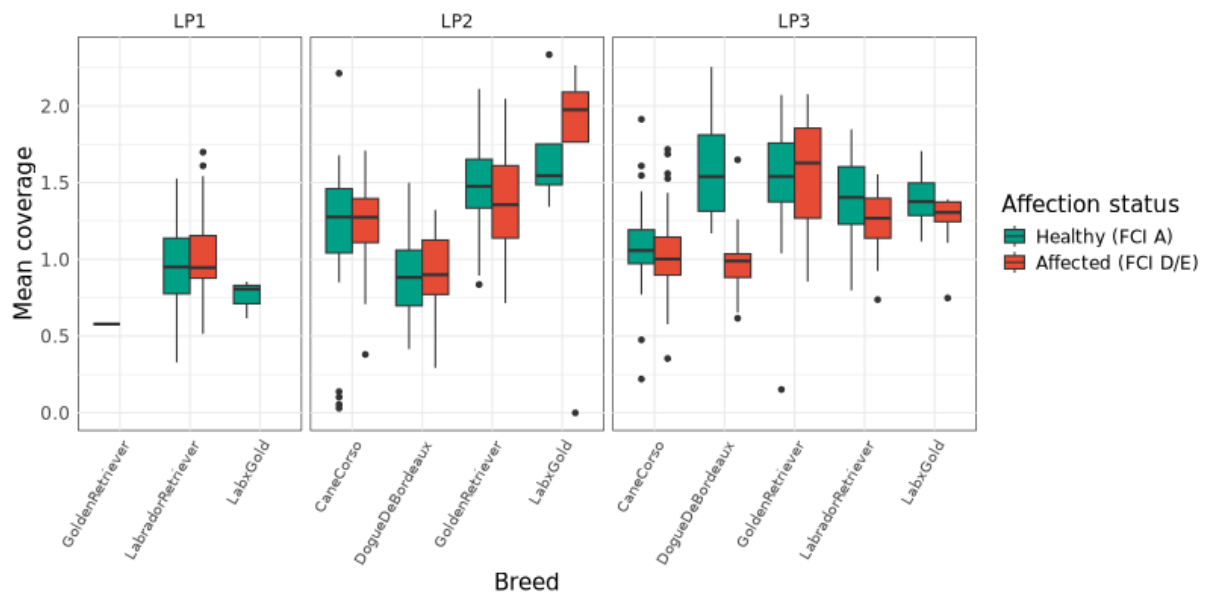


FIGURE 2 - Distribution of mean sequencing coverage across batches, dog breeds, and affection status.

Batch were split into three: LP1, LP2, and LP3. Affection is based on the *Fédération Cynologique Internationale (FCI)* scale (A are unaffected and D/E are strongly affected).

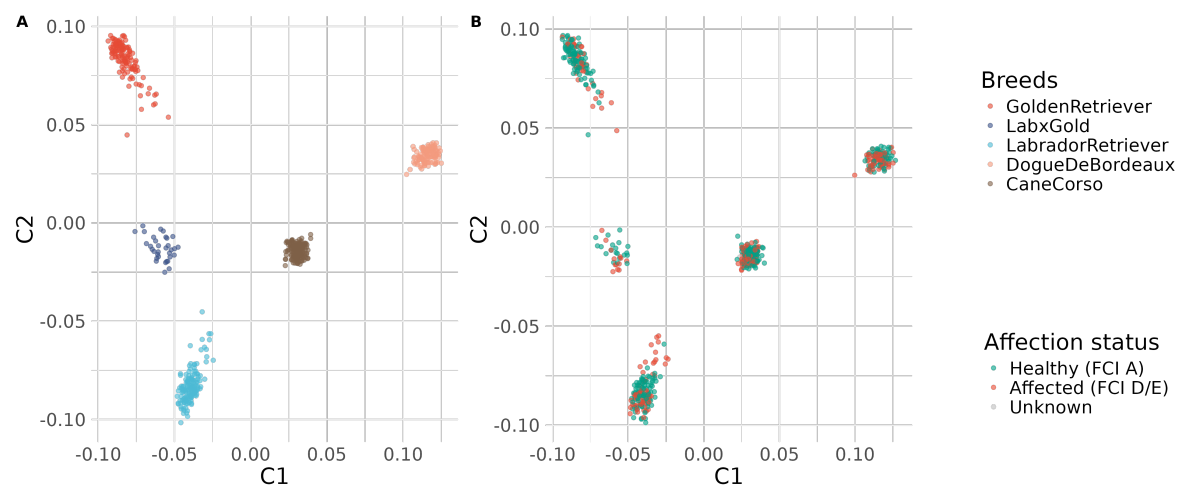


FIGURE 3 - Multi-Dimensional Scaling (MDS) plot of the samples by breed and affections status.

Affection is based on the *Fédération Cynologique Internationale (FCI)* scale (A are unaffected and D/E are strongly affected). Only dimension 1 and 2 are represented for a subsample set of independent variants using *plink v1.9*. The colours represent either the breed of each individual (A) or their affection status (B).

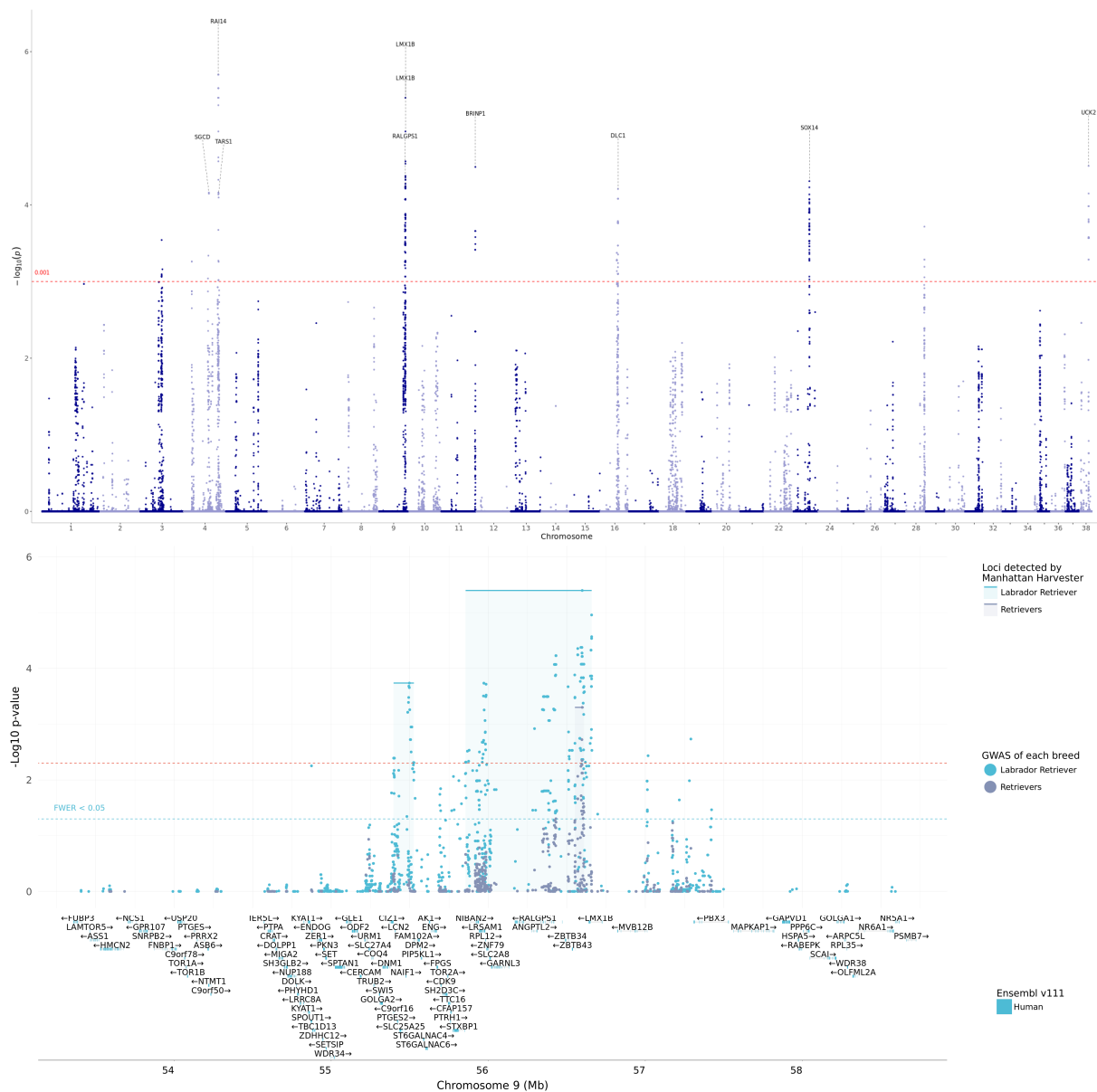


FIGURE 4 - Manhattan plot of the genome-wide association study (GWAS) for the Labrador population (56 cases, 96 controls).

Top panel: Whole genome GWAS based on 1 million phenotype permutations to control family wise error rates. Bottom panel: Detailed view of the significant loci on chromosome 9 (54-57 Mb) for the Labradors (blue dots) and for the Retrievers (purple dots) GWAS. The bottom part represents the genes' location from the Ensembl v111 genome annotation in dog. The highlighted box corresponds to the loci identified by Manhattan Harvester using -dots 5 -inlimit 0.5 -peak-limit 1.3.

TABLE 2 - Genome-wide significant loci identified in breed-specific GWAS for canine hip dysplasia

Include different breed groups studied (LAB = Labrador retriever, GOL = Golden retriever, RET= LAB, GOL and mix Labrador x Golden, CC= Cane Corso, DDBX = Dogue de Bordeaux, MAS = DDBX et CC (Mastifs), ALL = All Breed), sample size (cases / controls), number of SNPs (of which independent), chromosome and significant loci localisation, top SNP position and its associated p-values, and genes located in the locus. Genes already associated to hip dysplasia in literature are highlighted in bold and linked to the corresponding article: [1] Aki et al. (2018), [2] Todhunter et al. (2019), [3] Yan et al. (2019), [4] Cheng et al. (2021), [5] Jacobsen et al. (2023), [6] Padula et al. (2023), [7] Binversie et al. (2022a)

Breed(s)	Cases / Controls	nb SNP (ind)	Chr	Loci	Top SNP	p-value	Genes present in loci
LAB	56 / 96	5 911 045 (463 856)	4	54107384-54140940	54107704:G:A	0,000069	
			4	74549418-74626042	74560904:A:G	0,000002	
			4	75068320-75114606	75077444:C:T	0,000071	ADAMTS12
			9	55397150-55526694	55497812:C:T	0,000183	CIZ1, C9orf16, PTGES2, LCN2 [1,2] SLC25A25, NAIF1
			9	55855842-56658642	56598550:T:C	0,000004	NIBAN2, LRSAM1, RPL12, ZNF79, SLC2A8, GARNL3, RALGPS1 [3], ANGPTL2 [4], ZBTB34, ZBTB43, LMX1B [5]
			11	73265968-73580435	73549965:G:A	0,000032	
			16	37063435-37810972	37698727:T:C	0,000062	LONRF1, TRMT9B, DLC1
			23	34442797-34501186	34475352:G:A	0,000049	
			23	34512086-35013782	34524555:C:T	0,000154	SOX14, CLDN18, DZIP1L, A4GNT, RPS3A [3], DBR1, ARMC8
GOL	44 / 88	5 072 129 (294 290)	38	17904500-18371314	18242000:G:C	0,000031	FAM78B, UCK2, TMCO1, AP4S1, ALDH9A1, MGST3
RET	112 / 203	6 516 758 (609 846)	3	10999903-11216929	11196131:T:C	0,00211	
			3	26993868-27381654	27377912:C:T	0,000352	SERINC5, WBP11, THBS4 [6], MTX3, CMYA5
			3	45668890-46230199	46142459:C:A	0,00071	
			3	47411845-47443799	47423710:T:A	0,000389	
			9	56551929-56608705	56598550:T:C	0,000499	
			17	47017332-47125230	47049778:A:G	0,002885	LRRTM4 [2]
			22	42592532-42862787	42834405:T:A	0,001406	GPC5 [7]
			30	34304566-34387202	34386059:A:G	0,0004	TLE3 [3]
			31	36232164-36471324	36471324:G:C	0,000247	C2CD2, ZBTB21, UMODL1, ABCG1
RET	112 / 203	6 516 758 (609 846)	31	37262518-37342226	37334082:G:A	0,000047	

Breed(s)	Cases / Controls	nb SNP (ind)	Chr	Loci	Top SNP	p-value	Genes present in loci
CC	63 / 7	6 488 265 (630 680)					
DDBX	60 / 57	4 633 418 (230 200)	10	13261919-13275100	13264591:C:T	0,04791	
			10	14034917-14105302	14104157:G:A	0,03863	RAB21 [7]
			34	10311131-10325473	10317582:A:C	0,007726	
MAS	123 / 154	6 710 858 (663 032)	34	11558497-11728643	11565920:G:A	0,0212	
			34	8937903-9455401	8989195:A:G	0,002145	
			34	9455653-9655325	9455653:C:T	0,01894	
			3	46140578-46229694	46223756:A:C	0,02066	
			3	47424330-47453019	47441169:A:G	0,0079	
			4	49143382-49258307	49145736:C:G	0,00188	
			4	73364184-73385462	73371963:C:T	0,009248	UGT3A1
ALL	235 / 357	7 391 561 (900 954)	6	55777191-56142049	56140673:G:A	0,01025	BCAR3, FNBP1L
			17	45777604-46125087	46113023:C:T	0,01044	
			18	36920958-37240816	37131859:A:G	0,006333	MPPED2
			31	30411456-30418004	30418004:C:A	0,01339	
			36	10911541-10937059	10915458:A:G	0,01195	
			36	8072536-8094704	8090062:T:G	0,02216	GCA [3], KCNH7

5 Discussion

In this study, we conducted genome-wide association analyses (GWAS) of hip dysplasia across several dog breeds using both individual breed and grouped approaches. While our results uncovered several loci of interest, some of which are supported by the existing literature, some limitations and future directions merit discussion.

One of the primary limitations of our study is the small number of affected individuals, particularly in Retrievers. The power of GWAS depends heavily on the number of cases, especially for complex traits like hip dysplasia that are likely polygenic with moderate effect sizes. This limited sample size reduced our ability to detect associations reaching genome-wide significance.

We primarily relied on phenotype permutation-based GWAS using *plink*'s allelic model due to its robustness in small samples. More complex models such as mixed linear models incorporating genomic relatedness (e.g., *GCTA*) failed to identify genome-wide significant loci in most cases. This raises the possibility that the study may be underpowered for such approaches, or that overfitting and additional noise from covariates diluted genuine signals. Interestingly, despite these differences, both methods produced consistent peak signals, supporting the validity of our findings.

Strong variance in genome coverage between individuals and across sequencing batches introduced additional noise

in the data. Although normalization steps were applied, residual batch effects could have impacted the detection of subtle associations. Furthermore, the X chromosome was excluded from the analysis, as the pipeline does not yet allow haploid imputation. This may have led to missed sex-linked associations relevant to hip dysplasia. However, the canine literature does not show robust evidence of sexual dimorphism in prevalence in contrast to humans.

Out of the 36 loci identified, approximately 20 % overlap with genes previously implicated in hip dysplasia in either dogs or humans. This substantial concordance with prior studies provides strong validation of our analytical approach and supports the biological relevance of the loci highlighted here. One of the most compelling regions identified lies on chromosome 9, shared across the Labrador and Retriever GWAS. Within this locus, several genes with relevant biological functions were highlighted:

- *RALGPS1*: encodes a guanine nucleotide exchange factor involved in Ral signalling and cytoskeleton remodelling ;
- *ANGPTL2*: is part of the angiopoietin-like family of vascular growth factors and has known roles in inflammation and tissue remodelling, both of which are central to the progression of dysplasia ;
- *LMX1B*: a LIM-homeobox transcription factor, is involved in dorsal limb development and mutations

in this gene are associated with Nail-Patella Syndrome, which includes joint abnormalities, such as elbow dysplasia, as a core phenotype.

The convergence of these genes in a single locus suggests a potentially conserved pathway underlying hip development and its dysregulation in dysplasia. However, these candidate loci require further validation.

To confirm the relevance of the loci identified here, replication in an independent cohort is essential. Functional follow-up, such as transcriptomic profiling, would help elucidate potential mechanisms. The identification of robust and reproducible genetic markers could support the development of a genetic risk score, providing a valuable tool to guide breeding decisions and reduce the prevalence of hip dysplasia in predisposed dog breeds.

6 Acknowledgements

The authors would like to thank the Cani-DNA Biological Resource Centre (BRC), hosted at BIOSIT facility (<https://biosit.univ-rennes.fr/centre-de-ressources-biologiques-crb-cani-dna>), for managing and providing dog samples as part of several genetic research projects.

We would like to thank Anne-Sophie Guillory and Nadine Botherel for their dedicated contributions to this project.

We are also grateful to the GenOuest bioinformatics core facility (<https://www.genouest.org>) for providing the computing infrastructure used in this study.

We extend our thanks to the breed clubs that supported and promoted this research, including the Retriever Club, the Association Française des Bouviers Bernois, and the Société des Amateurs de Dogues de Bordeaux. We are especially thankful to the Association les Chiens Guides d'Aveugles de l'Ouest for their financial support and collaboration in the collection of samples within their facilities.

We also wish to acknowledge the valuable support of the veterinary experts who contributed to the phenotypic validation process: Jean-Pierre Genevois, Gilles Chaudieu, Philippe Haudiquet, Didier Fontaine, Didier Fau, and Francis Leg-eard.

6.1 Author contributions

LLN performed the experimental work, carried out the data analysis, and drafted the manuscript. LM coordinated collaborations and communications with partners. RG oversaw the management of the Cani-DNA BRC. BH contributed to data analysis and reviewed the manuscript. PQ supervised the project, reviewed and edited the manuscript. CA also reviewed and contributed to the final manuscript editing.

7 Funding

This work was performed in the frame of L.L.N.'s PhD research, as a part of a project on the comparative genetics of developmental genetic diseases in dogs and humans, funded by the "Fondation pour la Recherche Médicale" and the french guide dogs association: "Association les

Chiens Guides d'Aveugles de l'Ouest" (ACGAO) as well as the french foundation "Visio", which supports visually impaired children and adults.

8 Competing interests

No competing interest is declared.

9 Bibliography

- AKI, T. et al., 2018. A Whole-Genome Transcriptome Analysis of Articular Chondrocytes in Secondary Osteoarthritis of the Hip. *PLOS ONE* [en ligne]. Vol. 13, no. 6, e0199734 [visité le 2024-02-02]. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0199734.
- ANDRÉ, C. et al., 2022. Cani-DNA, un CRB qui a du chien! Réseau de collecte de prélèvements de chiens par les vétérinaires pour la recherche biomédicale et la diversité génétique. *NOV'AE - Ingénierie et savoir-faire innovants* [en ligne]. No. spécial 02 NOV'AE, p. 34–44 [visité le 2025-03-13]. ISSN 2823-3980. Disp. à l'adr. DOI: 10.17180/novae-2022-NS02-art05.
- ANDREWS, S., 2010. FastQC A Quality Control Tool for High Throughput Sequence Data [en ligne]. [visité le 2025-07-28].
- BALDINGER, A. et al., 2020. Prevalence of Canine Hip Dysplasia in 10 Breeds in France, a Retrospective Study of the 1997-2017 Radiographic Screening Period. *Plos one*. Vol. 15, no. 7, e0235847.
- BINVERSIE, E. E. et al., 2022a. Across-Breed Genetic Investigation of Canine Hip Dysplasia, Elbow Dysplasia, and Anterior Cruciate Ligament Rupture Using Whole-Genome Sequencing. *Frontiers in Genetics* [en ligne]. Vol. 13 [visité le 2022-12-06]. ISSN 1664-8021.
- BINVERSIE, E. E. et al., 2022b. Canine ACL Rupture: A Spontaneous Large Animal Model of Human ACL Rupture. *BMC Musculoskeletal Disorders* [en ligne]. Vol. 23, no. 1, p. 116 [visité le 2022-02-11]. ISSN 1471-2474. Disp. à l'adr. DOI: 10.1186/s12891-021-04986-z.
- BUCKLEY, R. M. et al., 2021. Best Practices for Analyzing Imputed Genotypes from Low-Pass Sequencing in Dogs. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*. ISSN 1432-1777. Disp. à l'adr. DOI: 10.1007/s00335-021-009914-z.
- CHEN, S. et al., 2018. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* [en ligne]. Vol. 34, no. 17, p. i884–i890 [visité le 2025-07-28]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/bty560.
- CHENG, B. et al., 2021. Integrative Analysis of MicroRNA and mRNA Sequencing Data Identifies Novel Candidate Genes and Pathways for Developmental Dysplasia of Hip. *CARTILAGE* [en ligne], p. 1947603521990859 [visité le 2021-09-02]. ISSN 1947-6035. Disp. à l'adr. DOI: 10.1177/1947603521990859.
- CINGOLANI, P. et al., 2012. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila Melanogaster* Strain W1118; Iso-2; Iso-3. *Fly*. Vol. 6, no. 2, p. 80–92. ISSN 1933-6942. Disp. à l'adr. DOI: 10.4161/fly.19695.
- FRIES, C. L. & REMEDIOS, A. M., 1995. The Pathogenesis and Diagnosis of Canine Hip Dysplasia: A Review. *The Canadian Veterinary Journal*. Vol. 36, no. 8, p. 494.
- GARCIA, M. et al., 2020. Sarek: A Portable Workflow for Whole-Genome Sequencing Analysis of Germline and Somatic Variants [en ligne]. No. 9:63 [visité le 2024-03-15]. Disp. à l'adr. DOI: 10.12688/f1000research.16665.2.
- GENEVOIS, J. -. et al., 2020. Prevalence of Canine Hip Dysplasia in 17 Breeds in France, a Retrospective Study of the 1993-2019 Radiographic Screening Period. *Revue Vétérinaire Clinique* [en ligne]. Vol. 55, no. 4, p. 123–146 [visité le 2021-04-14]. ISSN 2214-5672. Disp. à l'adr. DOI: 10.1016/j.anicom.2020.09.003.

- HALLER, T. et al., 2019. Manhattan Harvester and Cropper: A System for GWAS Peak Detection. *BMC Bioinformatics* [en ligne]. Vol. 20, no. 1, p. 22 [visité le 2025-06-02]. ISSN 1471-2105. Disp. à l'adr. DOI: 10.1186/s12859-019-2600-4.
- HAYWARD, J. J. et al., 2016. Complex Disease and Phenotype Mapping in the Domestic Dog. *Nature Communications* [en ligne]. Vol. 7, no. 1, p. 1–11 [visité le 2020-05-13]. ISSN 2041-1723. Disp. à l'adr. DOI: 10.1038/ncomm10460.
- JACOBSEN, K. K. et al., 2023. COL11A1 Is Associated with Developmental Dysplasia of the Hip and Secondary Osteoarthritis in the HUNT Study. *Osteoarthritis and Cartilage Open* [en ligne], p. 100424 [visité le 2024-01-08]. ISSN 2665-9131. Disp. à l'adr. DOI: 10.1016/j.ocarto.2023.100424.
- KARLSSON, E. K. & LINDBLAD-TOH, K., 2008. Leader of the Pack: Gene Mapping in Dogs and Other Model Organisms. *Nature Reviews Genetics* [en ligne]. Vol. 9, no. 9, p. 713–725 [visité le 2025-03-12]. ISSN 1471-0064. Disp. à l'adr. DOI: 10.1038/nrg2382.
- KING, M. D., 2017. Etiopathogenesis of Canine Hip Dysplasia, Prevalence, and Genetics. *Veterinary Clinics of North America: Small Animal Practice* [en ligne]. Vol. 47, no. 4, p. 753–767 [visité le 2022-11-04]. ISSN 0195-5616. Disp. à l'adr. DOI: 10.1016/j.cvs.2017.03.001.
- LEQUARRÉ, A.-S. et al., 2011. LUPA: A European Initiative Taking Advantage of the Canine Genome Architecture for Unravelling Complex Disorders in Both Human and Dogs. *Veterinary Journal* (London, England: 1997). Vol. 189, no. 2, p. 155–159. ISSN 1532-2971. Disp. à l'adr. DOI: 10.1016/j.tvjl.2011.06.013.
- MCKENNA, A. et al., 2010. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Research* [en ligne]. Vol. 20, no. 9, p. 1297–1303 [visité le 2022-11-10]. ISSN 1088-9051, ISSN 1549-5469. Disp. à l'adr. DOI: 10.1016/j.gr.107524.110.
- MEADOWS, J. R. S. et al., 2023. Genome Sequencing of 2000 Canids by the Dog10K Consortium Advances the Understanding of Demography, Genome Function and Architecture. *Genome Biology* [en ligne]. Vol. 24, no. 1, p. 187 [visité le 2025-05-10]. ISSN 1474-760X. Disp. à l'adr. DOI: 10.1186/s13059-023-03023-7.
- MEESON, R. L. et al., 2019. Spontaneous Dog Osteoarthritis—a One Medicine Vision. *Nature Reviews Rheumatology*. Vol. 15, no. 5, p. 273–287.
- MIKKOLA, L. et al., 2021. An Across-Breed Validation Study of 46 Genetic Markers in Canine Hip Dysplasia. *BMC Genomics* [en ligne]. Vol. 22, no. 1, p. 68 [visité le 2021-02-01]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/s12864-021-07375-x.
- OSTRANDER, E. A. et al., 2008. Canine Genetics Facilitates Understanding of Human Biology. *Genomics of Disease* [en ligne], p. 11–24 [visité le 2022-11-10]. Disp. à l'adr. DOI: 10.1007/978-0-387-76723-9_2.
- PADULA, G. et al., 2023. Preliminary Transcriptomic Analysis of Peripheral Blood from German Shepherd Dogs with Degenerative Joint Disease for the Identification of Diagnostic Biomarkers. *Gene* [en ligne]. Vol. 872, p. 147455 [visité le 2024-01-17]. ISSN 0378-1119. Disp. à l'adr. DOI: 10.1016/j.gene.2023.147455.
- PALLOTTI, S. et al., 2022. Dog-Human Translational Genomics: State of the Art and Genomic Resources. *Journal of Applied Genetics* [en ligne] [visité le 2022-09-13]. ISSN 2190-3883. Disp. à l'adr. DOI: 10.1007/s13353-022-00721-z.
- PEDERSEN, B. S. & QUINLAN, A. R., 2018. Mosdepth: Quick Coverage Calculation for Genomes and Exomes. *Bioinformatics* (Oxford, England). Vol. 34, no. 5, p. 867–868. ISSN

- 1367-4811. Disp. à l'adr. DOI: 10.1093/bioinformatics/btx699.
- RUBINACCI, S. et al., 2021. Efficient Phasing and Imputation of Low-Coverage Sequencing Data Using Large Reference Panels. *Nature Genetics* [en ligne]. Vol. 53, no. 1, p. 120–126 [visité le 2025-09-24]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-020-00756-0.
- SMITH, G., 1998. Canine Hip Dysplasia: Pathogenesis, Diagnosis, and Genetic Control. *Veterinary Quarterly* [en ligne]. Vol. 20, S22–S24 [visité le 2023-12-08]. ISSN 0165-2176, ISSN 1875-5941. Disp. à l'adr. DOI: 10.1080/01652176.1998.10807390.
- SUTTER, N. B. & OSTRANDER, E. A., 2004. Dog Star Rising: The Canine Genetic System. *Nature Reviews. Genetics*. Vol. 5, no. 12, p. 900–910. ISSN 1471-0056. Disp. à l'adr. DOI: 10.1038/nrg1492.
- SUTTER, N. B. et al., 2004. Extensive and Breed-Specific Linkage Disequilibrium in *Canis Familiaris*. *Genome Research* [en ligne]. Vol. 14, no. 12, p. 2388–2396 [visité le 2022-11-04]. ISSN 1088-9051, ISSN 1549-5469. Disp. à l'adr. DOI: 10.1101/gr.3147604.
- TODHUNTER, R. J. et al., 2019. Gene Expression in Hip Soft Tissues in Incipient Canine Hip Dysplasia and Osteoarthritis. *Journal of Orthopaedic Research®*. Vol. 37, no. 2, p. 313–324.
- VASIMUDDIN, M. et al., 2019. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) [en ligne], p. 314–324 [visité le 2025-07-28]. ISSN 1530-2075. Disp. à l'adr. DOI: 10.1109/IPDPS.2019.000041.
- WANG, C. et al., 2021. A Novel Canine Reference Genome Resolves Genomic Architecture and Uncovers Transcript Complexity. *Communications Biology* [en ligne]. Vol. 4, no. 1, p. 1–11 [visité le 2025-05-07]. ISSN 2399-3642. Disp. à l'adr. DOI: 10.1038/s42003-021-01698-x.
- WILLEMSSEN, K. et al., 2021. Comparing Hip Dysplasia in Dogs and Humans: A Review. *Frontiers in Veterinary Science* [en ligne]. Vol. 8 [visité le 2022-10-27]. ISSN 2297-1769.
- WRAGG, D. et al., 2024. A Cautionary Tale of Low-Pass Sequencing and Imputation with Respect to Haplotype Accuracy. *Genetics, Selection, Evolution : GSE* [en ligne]. Vol. 56, p. 6 [visité le 2024-01-17]. ISSN 0999-193X. Disp. à l'adr. DOI: 10.1186/s12711-024-00875-w.
- YAN, W. et al., 2019. A Genome-Wide Association Study Identifies New Genes Associated with Developmental Dysplasia of the Hip. *Clinical Genetics* [en ligne]. Vol. 95, no. 3, p. 345–355 [visité le 2024-02-01]. ISSN 1399-0004. Disp. à l'adr. DOI: 10.1111/cge.13483.

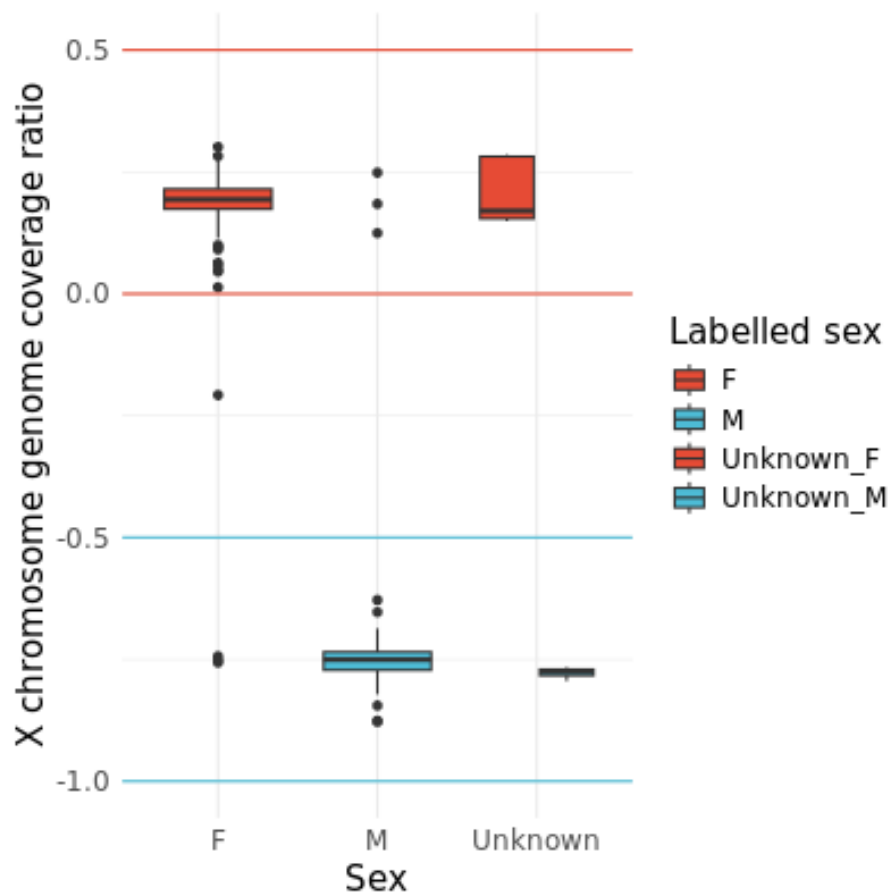


FIGURE S1 - Normalized X Coverage Index depending on affiliated sex

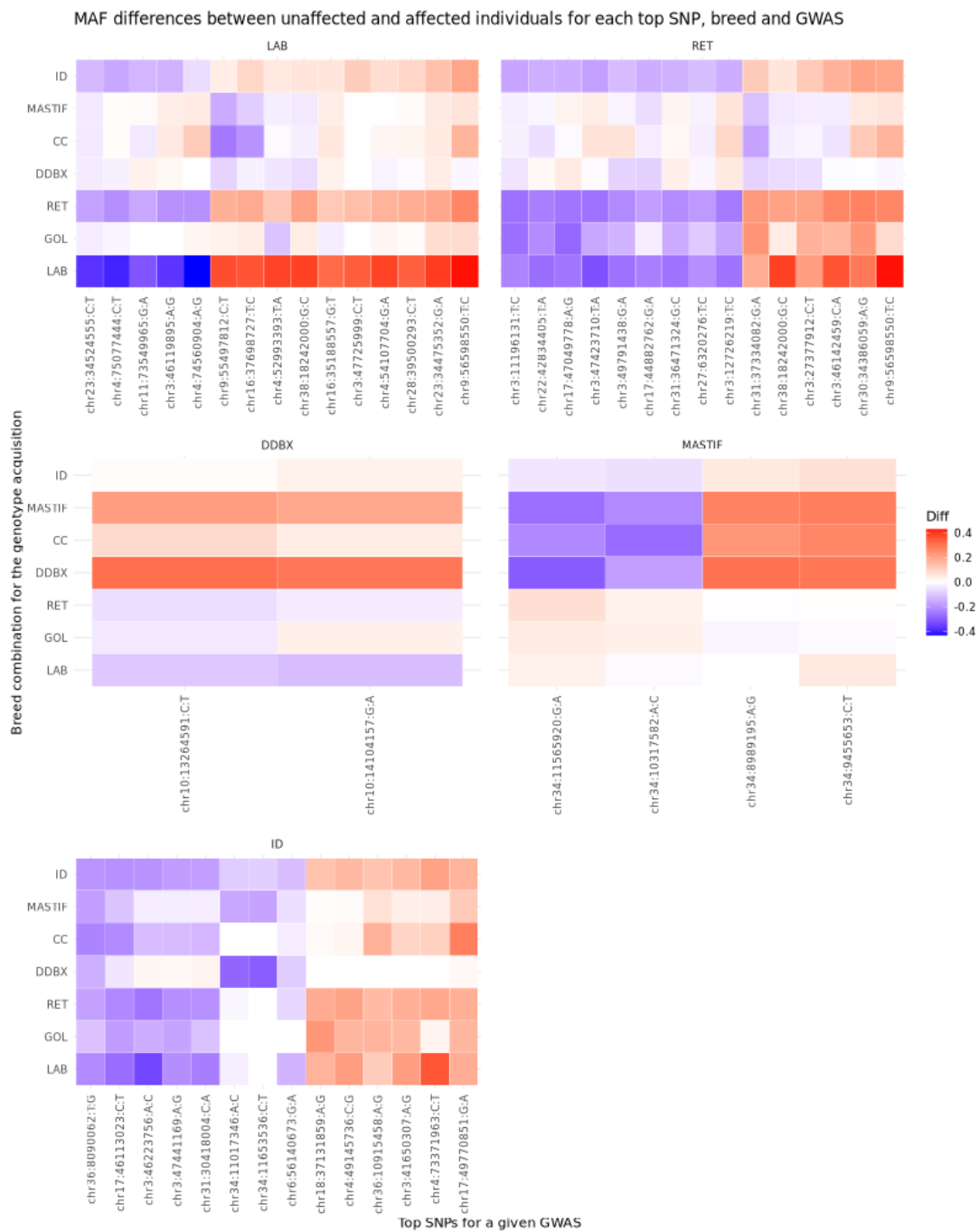


FIGURE S2 - Heatmap of case-control minor allele frequency (MAF) differences across populations for lead SNPs identified in each GWAS.

Quatrième partie

DISCUSSION & CONCLUSION



DISCUSSION

La présente thèse s'est intéressée aux bases génétiques de la dysplasie coxo-fémorale (DCF) chez le chien comme modèle de la dysplasie de la hanche (DDH) chez l'humain. Ce projet, initié de longue date en collaboration avec l'Association des Chiens Guides de l'Ouest (ACGAO), conserve toute sa pertinence scientifique. En effet, malgré l'identification de nombreux loci associés à cette affection tant chez l'humain que chez le chien, seuls quelques gènes candidats semblent émerger, sans qu'aucune preuve fonctionnelle solide n'ait confirmé leur rôle. Par ailleurs, aucun test génétique de risque fiable n'est aujourd'hui disponible ni pour l'humain ni pour le chien. C'est dans ce contexte, encore largement ouvert et riche en perspectives, que j'ai rejoint le laboratoire en 2020 pour contribuer à ce projet.

Afin d'étudier la DCF chez le chien, nous avons sélectionné cinq races présentant une prévalence variable mais élevée de la DCF et séquencé par technologie *low-pass* un échantillonnage de plus de 700 individus, regroupant la moitié de cas et de contrôles. L'analyse génétique a combiné des études d'association génome entier (GWAS) avec une comparaison des loci identifiés dans la littérature humaine et canine existante, permettant d'évaluer la conservation et la spécificité des loci identifiés inter-races canines et interspèces humain - chien.

Cette étude a de même permis de mettre en lumière les forces, mais aussi les limites, de la comparaison entre les dysplasies de la hanche chez l'humain et le chien, notamment sur les aspects anatomiques, mécaniques et physiopathologiques. Les similitudes et différences identifiées permettront d'affiner l'interprétation des résultats obtenus en génétique et la transposition des connaissances d'une espèce à l'autre.

Enfin le développement et l'application d'outils de génétique computationnelle et de science ouverte, tels que *Bioconductor* *Pedixplorer* et les workflows *nf-core/phaseimpute*, ont favorisé une approche reproductible et automatisée et serviront, je l'espère, la communauté scientifique du domaine.

Ces éléments seront détaillés et discutés successivement dans les sections suivantes, qui abordent la collecte des échantillons, le choix du phénotype et des populations étudiées, les analyses génétiques et enfin le développement en science ouverte d'outils de génétique computationnels.

A Contraintes méthodologiques et logistiques

1 Recrutement des individus

Si la dysplasie coxo-fémorale (DCF) est une affection couramment dépistée chez le chien, le contexte juridique, social et émotionnel qui l'entoure freine fortement les capacités de collecte et la participation volontaire de certains éleveurs et propriétaires.

D'un point de vue légal, bien que le chien soit désormais reconnu comme un « être vivant doué de sensibilité », il reste, selon la loi, assimilé à un bien meuble. La DCF figure parmi les six vices rédhibitoires reconnus par le Code rural (*Article R213-2*, 2003), ce qui signifie que tout chien vendu doit être garanti indemne de cette affection. Chez les chiens de race inscrits au LOF, les individus diagnostiqués dysplasiques sont systématiquement écartés de la reproduction. De ce fait, les éleveurs et les clubs de race sont souvent réticents à communiquer sur la présence de cas dans leurs lignées. Cette réticence est renforcée par des croyances persistantes et une compréhension seulement partielle du caractère polygénique et multifactoriel de la DCF. En effet, la reproduction de deux individus sains ne garantit en aucun cas l'absence d'individus atteints dans la progéniture. Cela fragilise la confiance dans les recherches sur le sujet mais aussi sur le diagnostic. Faute d'outils fiables, la sélection repose souvent sur des choix empiriques sans garantie réelle et déroute les éleveurs et acquéreurs de chiens.

Du côté des propriétaires, les réactions à l'annonce d'un diagnostic sont également contrastées. Certains expriment de l'incompréhension ou de la colère, se sentant lésés par l'éleveur ou par le lecteur des radiographies. Pour d'autres, c'est l'inquiétude qui domine face à la perspective de complications de santé futures ou la dévalorisation de leur animal. Un grand nombre de propriétaires n'ont ainsi pas donné suite à leur volonté initiale de participer au projet. Certains, ne croyant pas au diagnostic, ont alors fait relire la radio à l'étranger, là où la notation est plus permissive, bien que cela ne change en rien l'état clinique du chien, voire aggrave la situation si ce chien reproduit.

Face à ces difficultés, un travail de communication intensif a été nécessaire : échanges réguliers avec les clubs des races étudiées, éleveurs, vétérinaires, séances d'explication et de vulgarisation et surtout, accompagnement individualisé. À ce titre, l'implication du Dr vétérinaire Lauranne Machou a été déterminante pour établir un dialogue de confiance avec les propriétaires, leur expliquer les enjeux et apporter des conseils concrets pour la prise en charge de leur chien.

Ce travail de terrain a porté ses fruits : malgré les difficultés, plusieurs centaines de chiens atteints ont pu être inclus ces quatre dernières années, permettant d'atteindre une taille d'échantillons suffisante pour réaliser les premières analyses génomiques. La collecte, amorcée dès 2008, s'est poursuivie tout au long de la thèse et les échantillons

les plus récents permettront à l'équipe et à la future thèse de valider de manière indépendante les résultats obtenus. L'obtention des données de séquençage s'est elle aussi étalée dans le temps, le premier lot a été préparé et envoyé à la société de séquençage aux Etats-Unis courant 2022 tandis que les données du 3^{ème} et dernier lot ont été reçues pendant l'été 2024, soit à la fin de la 3^{ème} année de thèse.

Compte tenu de ces délais et à la suite de la soumission de trois demandes, une prolongation d'un an a été obtenue par la Fondation pour la Recherche Médicale, afin de finaliser les analyses sur l'ensemble du jeu de données et de valoriser les outils développés au cours de la thèse.

2 Qualité et standardisation du phénotype

Le diagnostic de la DCF, bien qu'encadré par des normes strictes en Europe, reste une étape complexe et sensible dans la chaîne d'exams cliniques vétérinaires. Il ne peut être considéré comme officiel par la FCI qu'à partir de l'âge d'un an pour la plupart des races, une fois le développement articulaire terminé. Plusieurs sources de variabilité technique et interprétative peuvent influencer la fiabilité du diagnostic.

D'une part, la radiographie standardisée utilisée pour le score FCI est sensible à des paramètres opératoires. La sédation doit garantir un relâchement musculaire optimal afin de pouvoir bien observer l'espace interarticulaire. De même le bon positionnement de l'individu sur la table et la bonne orientation de son bassin et de ses membres inférieurs sont primordiaux afin d'obtenir une radiographie fiable et précise. Une rotation asymétrique du bassin peut en effet masquer un défaut de coaptation articulaire sur une des hanches, induisant un faux négatif.

D'autre part, la lecture de la radiographie elle-même présente une part de subjectivité. Le score FCI combine une mesure quantitative, l'angle de Norberg-Olsson, mais aussi des critères qualitatifs plus difficiles à standardiser, comme la forme de l'acétabulum ou la congruence articulaire, ce qui induit une variabilité interlecteurs, selon l'expérience et la rigueur du lecteur et certainement, une variabilité inter-races, peu connue et non prise en compte à ce jour, par exemple pour le groupe des Setter (communication personnelle Pr F. Legeard).

Face à ces limites, plusieurs méthodes alternatives ont été proposées dans la littérature. Il est ainsi désormais recommandé de réaliser une deuxième radiographie en subluxation. L'indice de distraction permet de quantifier la laxité articulaire mais est encore peu pratiqué. Enfin des approches de lecture automatisée par intelligence artificielle (IA) émergent chez l'humain (Chen et al., 2023 ; Ahn et al., 2025) comme chez le chien (McEvoy et al., 2021 ; Hennessey et al., 2022 ; Boufenar et al., 2023). Ces approches alternatives peinent néanmoins à se généraliser. La lecture radiographique officielle

reste aujourd'hui un système bien établi et rentable, structuré autour de quelques lecteurs agréés en France (Genevois, 2018), ce qui limite l'incitation au changement.

Afin de réduire le bruit phénotypique et de fiabiliser notre cohorte, nous avons opté pour une approche fondée sur des phénotypes extrêmes (i.e., indemnes vs fortement atteints), moins sensibles aux imprécisions de notation. Nous avons également vérifié que les lectures d'individus indemnes avaient bien été réalisées après l'âge d'un an et priorisé les évaluations effectuées par le Pr Genevois, reconnu pour son exigence diagnostique et son expertise, afin d'assurer une cohérence stricte des phénotypes retenus.

Cette stratégie nous permet d'espérer une bonne robustesse statistique, en minimisant les erreurs de classification.

B Validation du choix des races canines étudiées

La forte structuration génétique des races canines constitue à la fois un atout et une limite pour l'étude des maladies complexes. D'un côté, elle facilite la détection de loci associés en réduisant la variabilité génétique de fond au sein de chaque race. De l'autre, elle s'accompagne d'un déséquilibre de liaison marqué, formant de longs blocs d'haplotypes homozygotes dans lesquels plusieurs variants sont cotransmis. Cette configuration complique la résolution fine des analyses d'association, puisqu'il devient difficile d'isoler le variant causal parmi les nombreux variants corrélés identifiés dans un même bloc haplotypique.

Dans ce travail, nous nous sommes concentrés sur deux grands groupes : les retrievers, largement utilisés comme chiens guides et les molossoïdes (e.g., Cane Corso, Dogue de Bordeaux), réputés particulièrement affectés par la DCF. Ces deux groupes, déjà étudiés au laboratoire pour d'autres affections génétiques, disposaient par ailleurs de prélèvements et de données cliniques préexistantes. Phylogénétiquement, ils sont assez éloignés (voir figure 25), les retrievers appartenant au clade rouge (en bas à gauche de la figure 25) et les molossoïdes aux clades bleu et violet (en bas à droite de la figure 25). L'intégration de races apparentées dans des analyses GWAS présente un intérêt majeur : les différences dans la structure du LD entre races proches permettent de raffiner les régions associées, en exploitant les points de rupture distincts. C'est précisément cette approche comparative, combinant retrievers et molossoïdes, que nous avons mise en œuvre afin d'améliorer la résolution des loci candidats.

Cependant, pour enrichir la compréhension globale de la dysplasie coxo-fémorale (DCF), il serait pertinent d'inclure des races issues d'autres branches phylogénétiques, telles que le Berger Allemand, des races anciennes comme l'Akita Inu ou l'Alaskan Malamute ou bien encore les Setter, proches cousins des retrievers et intéressant sur le plan de l'anatomie de la hanche, apparemment différente entre les Setter (i.e., Irlandais,

Gordon et Anglais). Ces populations pourraient mettre en évidence de nouveaux loci associés à la maladie, en raison de leur histoire évolutive distincte et de l'hypothèse de combinaisons géniques spécifiques selon les races.

Néanmoins, plusieurs contraintes logistiques ont limité cette diversification dans la collecte des prélèvements. Malgré nos sollicitations auprès du club, il n'a pas été possible d'obtenir des prélèvements suivis de chiens Berger Allemand. Par ailleurs, les races anciennes présentent souvent en France des effectifs très réduits, ce qui rend leur recrutement particulièrement complexe.

Ces éléments soulignent l'importance de la diversité raciale dans les études génétiques canines et appellent à renforcer les collaborations avec les éleveurs et clubs de race pour optimiser le recrutement futur.

C Limites de l'étude et interprétation des résultats

1 Comparabilité inter espèces

Malgré la transition de la quadrupédie à la bipédie, l'anatomie humaine reste très proche de celle du chien (voir figure 26). Ainsi, la DCF chez le chien et la DDH chez l'humain partagent de nombreuses similitudes, qu'elles soient anatomiques, physiopathologiques ou diagnostiques. Néanmoins, des différences notables existent et doivent être soigneusement considérées dans une approche comparative.

L'un des éléments les plus différenciants est le rôle de la laxité articulaire. Chez le chien, la laxité est un facteur déterminant dans le développement de la DCF et elle serait mesurable de façon prédictive dès l'âge de quatre mois (Adams et al., 1998). Chez l'humain, la compréhension de son rôle a considérablement évolué au fil des décennies. Historiquement, la laxité (souvent testée via le signe de Barlow / Ortolani chez le nourrisson) était considérée comme le principal déterminant de la DDH. Avec les progrès de l'imagerie (i.e., échographie puis IRM), le consensus s'est nuancé : la laxité précoce joue bien un rôle, mais son lien direct avec la progression vers une dysplasie pathologique est plus complexe. Certains nourrissons présentent une laxité transitoire qui se résout spontanément, tandis que dans d'autres cas, elle persiste et contribue à des anomalies structurales (Shipman et al., 2006 ; Harsanyi et al., 2020a).

Par ailleurs, le passage de la quadrupédie à la bipédie a considérablement modifié la configuration tridimensionnelle de l'articulation coxo-fémorale humaine. La lordose lombaire, associée à une rétroversion du bassin et à un valgus fémoral accru, réduit la couverture antérieure de la tête fémorale par l'acétabulum. Ces adaptations influencent les contraintes et charges mécaniques sur l'articulation et pourraient donc moduler différemment l'usure et les mécanismes dégénératifs selon l'espèce. Cependant, il est important de noter que la marche bipède ne s'installe que vers 12 à 18 mois chez l'enfant, soit après le développement initial de la DDH. Cela relativise l'influence directe

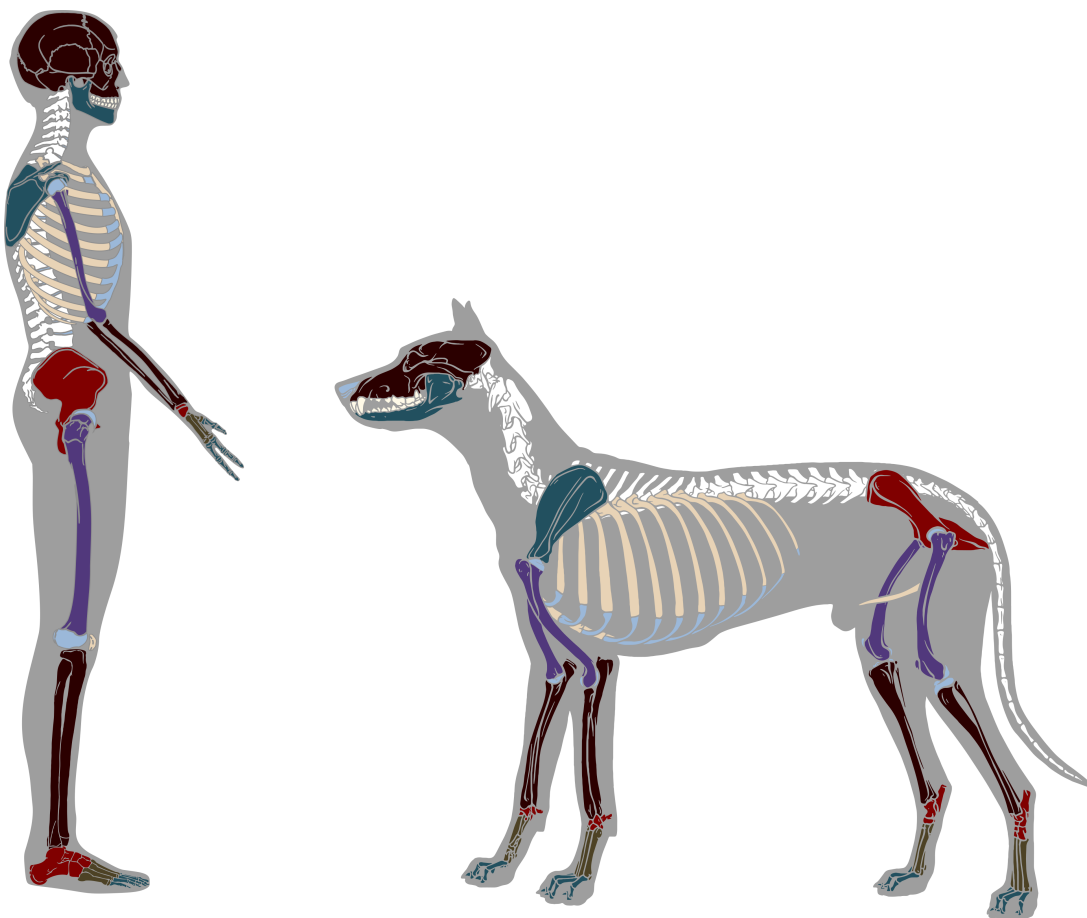


FIGURE 26 – Comparaison des différents os du squelette chez l'humain et le chien

de la bipédie dans la genèse de ces anomalies articulaires, qui semblent s'installer en amont.

Il est également important de souligner la différence de vitesse de maturation articulaire : elle est d'environ un an chez le chien (Riser, 1975) contre huit ans chez l'humain (Lee & Eberson, 2006). Cette divergence temporelle pourrait avoir un impact sur la période critique de développement des dysplasies.

Malgré ces différences, les deux espèces partagent une structure articulaire de base similaire et la progression de la dysplasie suit des trajectoires physiopathologiques largement comparables (Meeson et al., 2019). Ce parallèle est renforcé par l'identification de loci génétiques communs associés à cette affection dans les deux espèces, mis à jour dans la revue rédigée pour cette thèse. Si tous les gènes impliqués chez l'une ne sont pas nécessairement retrouvés chez l'autre, un noyau commun de gènes candidats orthologues dans certains loci en synténie semble émerger appuyant la pertinence du modèle canin de DCF dans l'étude de la DDH humaine.

2 Analyses génétiques

À partir d'une cohorte de plus de 900 chiens, pour lesquels nous avons vérifié les statuts dysplasiques et obtenu de l'ADN de qualité et en quantité suffisante, nous avons réalisé un séquençage à faible profondeur du génome entier (*Low-Pass Whole Genome Sequencing*, LP-WGS) pour 707 d'entre eux. En complément, nous avons également à notre disposition des données de génotypage issues d'autres projets menés au sein de l'équipe. Les analyses génétiques se sont appuyées sur des études d'association à l'échelle du génome (GWAS) et des analyses de liaison sont prévues afin d'explorer plus finement l'hérédité et la transmission des loci identifiés. Je passe ici en revue ces différentes approches, en discutant leurs atouts respectifs ainsi que leurs limites.

a Données de génotypage

Les analyses d'association présentées dans cette thèse se fondent uniquement sur les données nouvellement acquises par séquençage low-pass. Cependant, l'équipe « Génétique du Chien » dispose également d'un volume important de données de génotypage sur puces *Illumina* (170 000 SNPs) et *Affymetrix* (600 000 SNPs) sur des individus indemnes et atteints de DCF, appartenant aux mêmes races que celles étudiées ici.

À ce stade, ces données n'ont pas encore été intégrées, notamment en raison des limitations actuelles du pipeline *nf-core/phaseimpute*, qui n'inclut pas encore les outils d'imputation de données de génotypage tels que *Beagle5*, *Minimac4* ou *Impute5* (Browning et al., 2018 ; Rubinacci et al., 2020 ; Browning et al., 2021). Un stagiaire en Master 1 de statistique s'y est cependant consacré avec mon aide ces derniers mois. Par ailleurs, l'ajout de données plus anciennes, générées à partir de technologies différentes, soulève un risque non négligeable de biais dans les résultats dus aux effets

batch. Nous envisageons toutefois d'intégrer ces jeux de données hétérogènes dans les futures analyses, en utilisant un modèle tenant compte de la technologie de génotypage comme covariable. Une telle stratégie permettra également de structurer l'analyse selon une approche d'apprentissage supervisé, avec une répartition des individus en un jeu d'entraînement (80-90 %) et un jeu de validation (10-20 %), destiné à la réplication des analyses, pour vérifier la reproductibilité des signaux d'association.

Ce travail sera réalisé dans le cadre du développement continu du pipeline *nf-core/phaseimpute*, par l'intégration des modules d'imputation nécessaires, que je poursuivrai après cette thèse.

b Analyses de liaison génétique

Pour une partie des individus inclus dans nos cohortes, nous disposons de données généalogiques, obtenues à partir des pedigrees fournis par les éleveurs ou retrouvés sur le site *Loft Select* de la SCC (Société Centrale Canine : <https://www.centrale-canine.fr/lofselect>). Ces informations nous ont permis d'identifier plusieurs familles de retrievers suffisamment nombreuses et structurées pour être exploitables dans le cadre d'analyses de liaison génétique.

Ces familles présentent une distribution informative du phénotype au sein de leur arbre généalogique, avec des cas atteints sur plusieurs générations. Elles offrent donc une opportunité d'analyses génétiques complémentaires aux analyses d'association, en permettant de rechercher des co-ségrégations de régions génomiques avec le phénotype dans un contexte familial.

Nous sommes toutefois en attente de données généalogiques supplémentaires provenant du CESECAH (Centre d'Élevage et de Sélection des Chiens d'Accompagnement aux Handicapés), qui devraient compléter certains des pedigrees actuellement partiels. Ces données permettront de confirmer ou d'infirmer les signaux identifiés par GWAS, en recourant à une approche statistique indépendante, fondée sur les liens de parenté entre individu et donc moins sensible à la structuration de population propre aux races canines.

Avec le développement de l'outil *Pedixplorer* et la disponibilité des données de séquençage, produit lors de cette thèse, lorsque le CESECAH nous confiera les données de généalogie, ce travail sera réalisable dans la suite de mon travail de thèse.

c Analyses d'Association génétique : GWAS

i Phénotype utilisé

Pour les analyses statistiques réalisées dans cette thèse, nous avons retenu un phénotype binaire, en opposant les chiens classés A (indemnes) aux chiens classés D ou E (atteints), selon les stades de la FCI (Fédération Cynologique Internationale). Ce choix visait à maximiser la fiabilité du phénotype, en évitant les zones d'incertitude représentées par les catégories intermédiaires B et C.

En effet, la classification des individus B et C est sujette à variabilité interlecteurs et à de nombreux critères d'interprétation, pouvant significativement varier entre lecteurs officiels. De plus, dans certaines études, les individus B sont inclus avec les contrôles, mais les C peuvent être inclus avec les cas ou avec les contrôles. Ces choix permettent certes d'augmenter la taille des échantillons, mais au prix d'un risque accru de bruit phénotypique, susceptible d'affaiblir la puissance des analyses et de donner de fausses pistes de résultats, voire de faux gènes candidats.

Une autre stratégie aurait été de modéliser un trait quantitatif, tel que l'angle de Norberg-Olsson ou l'indice de distraction, qui reflètent plus précisément la progression et le caractère développemental de la maladie. Toutefois, ces mesures ne sont pas systématiquement disponibles et leur intégration aurait conduit à un taux élevé de données manquantes. Nous avons donc choisi de ne pas les intégrer, afin d'éviter une perte trop importante de données et de préserver la robustesse statistique de nos analyses.

Enfin, il est également envisageable d'exploiter le score FCI comme variable semi-quantitative, approche déjà utilisée dans la littérature (Wang et al., 2021b). Cela permettrait à la fois d'intégrer un plus grand nombre d'individus et de modéliser plus finement les phénotypes intermédiaires, ce qui serait particulièrement utile pour le développement d'outils de génétique prédictive, notamment des scores de risque polygénique (Bartolomé et al., 2015).

ii Sélection des modèles statistiques

Le choix du modèle statistique approprié pour les analyses GWAS a représenté l'un des aspects les plus complexes de cette thèse. La littérature regorge de modèles, chacun se présentant comme une amélioration des précédents, rendant la sélection difficile pour des données spécifiques et fortement structurées comme celles des races canines.

Nous avons appliqué plusieurs approches statistiques aux données, incluant des corrections par analyse en composantes principales (ACP) et des modèles mixtes utilisant en covariable la matrice de relation génomique (GRM). Cette diversité d'approches a généré un ensemble de résultats parfois complémentaires, mais complexes à interpréter et à comparer.

Dans un contexte de forte structuration génétique, comme c'est le cas ici, les corrections trop strictes peuvent conduire à des modèles excessivement conservateurs, augmentant le risque de faux négatifs. L'inclusion d'un trop grand nombre de composantes principales ou de la GRM peuvent ainsi masquer des signaux réels d'association.

C'est pourquoi il est essentiel de contrôler rigoureusement l'inflation génomique (e.g., via le λ_{GC} ou les QQ plots) et de sélectionner les modèles offrant le meilleur compromis entre correction de la structure et sensibilité statistique, en fonction de la population analysée. Dans ce travail, j'ai réalisé une approche exploratoire, accompagnée d'une évaluation critique des performances de chaque modèle.

Ce travail exploratoire nous a permis d'identifier les configurations de modèles les plus robustes en fonction de nos données, en particulier ceux intégrant une correction modérée de la structure populationnelle. La variabilité des résultats souligne qu'aucun modèle unique ne peut être universellement et seulement optimal et que le choix du cadre statistique doit être adapté à la structure génétique propre à chaque population étudiée. À l'avenir, l'intégration de méthodes bayésiennes ou multi-modèles pourrait constituer une piste prometteuse pour tirer le meilleur parti de la complémentarité des approches tout en limitant le risque de faux négatifs.

iii Concordance des résultats avec la littérature

Malgré la variété des différentes analyses réalisables, les différentes méthodologies appliquées aux données ont donné des résultats similaires. Comme précisé dans les résultats, les analyses statistiques de cette cohorte ont permis d'identifier une vingtaine de loci. L'annotation de ceux-ci par la littérature existante ainsi que par les gènes présents les plus pertinents, a permis de mettre en évidence des gènes candidats prometteurs pour la suite de ce projet.

Nous pouvons notamment citer le gène *LMX1B* sur le chromosome 9 déjà identifié par Jacobsen et al. (2023) chez l'humain au sein de la cohorte HUNT (étude Trøndelag Health de 69 500 personnes) (Krokstad et al., 2013). Ce gène n'est mentionné dans le tableau 3 des appendices de l'article, cependant nous l'avons retenu compte tenu de son implication dans le développement des membres. En effet, il code un facteur de transcription de la famille des LIM-homeodomain, caractérisé par deux domaines LIM N-terminaux, un homéodomaine et une région riche en glutamine. Il joue un rôle essentiel dans plusieurs processus du développement, notamment la différenciation des structures dorsales des membres, la formation de la membrane basale glomérulaire, le développement du segment antérieur de l'œil, ainsi que la différenciation des neurones dopaminergiques et sérotoninergiques. Dans le contexte de la dysplasie coxo-fémorale, l'implication de *LMX1B* apparaît particulièrement intéressante du fait de son rôle crucial dans la spécification du destin dorsal des membres et dans l'organisation

des structures articulaires. Ainsi, son identification dans notre cohorte renforce l'hypothèse d'une convergence entre les mécanismes génétiques impliqués chez le chien et certains modèles pathologiques décrits chez l'humain.

GPC5 est aussi un gène prometteur déjà identifié par Binversie et al. (2022a) dans une étude comprenant plus de 20 races différentes. Ce gène code une protéine de surface de la famille des glypicans, des protéoglycanes membranaires ancrés via un groupement glycosyl-phosphatidylinositol (GPI) et porteurs de chaînes d'héparane sulfate. Des altérations de *GPC5* ont été associées à plusieurs affections humaines, notamment certains syndromes dysmorphiques osseux (e.g., l'omodysplasie), reflétant son rôle dans l'organisation tissulaire et la signalisation cellulaire. Dans le cadre de la dysplasie coxo-fémorale, l'intérêt de *GPC5* réside dans sa fonction de protéoglycane de surface impliqué dans les interactions cellule-matrice et dans le contrôle du développement squelettique.

Cependant, notre étude n'a pas permis de retrouver certains autres gènes majeurs décrits dans la littérature du domaine, tels que *CX3CR1*, *ASPN*, *COL1A1*, *HOXB9* et *GDF5* chez l'humain (Gkiatas et al., 2019 ; Wen et al., 2022), *FBN2* chez le chien (Friedenberg et al., 2011) ou encore *CTBP2* (CFA 28 / HSA 10, sur le même locus en LD avec le gène *ZRANB1*, impliqué dans la DDH (Feldman et al., 2014 ; Hayward et al., 2016)) ni l'un des 21 loci, pourtant répliqués par (Mikkola et al., 2021).

Le fait de ne pas retrouver les mêmes loci et régions chromosomiques que dans les précédentes études peut s'expliquer par de nombreux facteurs, notamment les différences d'espèces et de populations ou races étudiées, les variations dans la définition et la mesure des phénotypes, la définition et le positionnement des loci et des gènes candidats pointés (en LD ou pas), ainsi que de l'hétérogénéité des méthodes d'acquisition et d'analyse des données biologiques et génétiques. De plus, bien que de nombreuses études aient identifié des centaines de loci associés à la dysplasie coxo-fémorale, la majorité d'entre eux n'ont pas été confirmés dans des cohortes indépendantes de réplification (Meeson et al., 2019 ; Mikkola, 2020).

Dans cette perspective, un objectif essentiel de la suite de mes travaux sera de répliquer ces analyses à partir d'un jeu de données plus large intégrant des données de génotypage, afin de renforcer la robustesse et la validité des associations génétiques identifiées.

D Science ouverte et développement d'outils de génétique computationnelle

1 Science ouverte

Au cours de cette thèse, j'ai dressé un état des lieux des connaissances sur les bases génétiques de la dysplasie coxo-fémorale (DCF), tant chez l'humain que chez le chien. Cette exploration bibliographique a révélé une difficulté majeure pour bien comparer et intégrer les études : le manque d'accessibilité et de standardisation des données publiées.

Dans de nombreuses études, seules des tables résumées sont disponibles, sans accès aux jeux de données complets. Plus problématique encore, les métadonnées essentielles, telles que la version du génome de référence, sont souvent absentes : parmi les 150 articles consultés, plus de la moitié ne la mentionnaient pas. J'ai dû, à plusieurs reprises, déduire cette information à partir de la date de publication ou d'indices indirects. Cette absence de précision, tant sur les référentiels et matériels que sur les méthodes employées, a entravé toute comparaison rigoureuse avec plusieurs travaux, en particulier ceux des années 2000 basés sur des microsatellites non documentés ou aujourd'hui inaccessibles.

Dans certains cas, les bases de données référencées dans les articles ne sont plus disponibles en ligne. Ainsi, grâce à des outils comme WebArchive (Kahle, 1997), j'ai pu en récupérer une partie, mais plusieurs ressources semblent définitivement perdues, avec un impact direct sur la pérennité des connaissances scientifiques.

Concernant les différents outils que j'ai pu utiliser tout au long de cette thèse, j'ai rencontré plusieurs difficultés techniques liées au fait de travailler sur une espèce non humaine. Certains outils très utilisés, comme regenie (Mbatchou et al., 2021) servant à la modélisation de GWAS complexes, sont explicitement conçus pour le génome humain : la longueur des chromosomes y est codée en dur, rendant l'outil inutilisable en l'état pour l'utiliser pour le chien. J'ai donc dû modifier le code source, le recompiler (souvent sans documentation claire sur les dépendances), puis tester son bon fonctionnement sur mes données. D'autres outils, notamment pour l'analyse de pedigrees, sont aujourd'hui obsolètes, non maintenus ou présentent des bogues non corrigés. J'ai essayé, dans la mesure du possible, de contacter les développeurs ou ouvrir des issues *GitHub*. Ce système de suivi collaboratif permet de déclarer des problèmes ainsi que de suggérer des améliorations pour un logiciel. Pour une partie de ces outils, je n'ai malheureusement jamais eu de réponses et ai donc dû réorienter mes analyses avec d'autres logiciels.

2 Démarche et communautés

Face à ce constat, j'ai choisi de m'appuyer sur des communautés bio-informatiques engagées dans la science ouverte, en particulier *Bioconductor* et *nf-core*, pour développer respectivement deux outils dont j'ai eu besoin pour avancer dans le projet : *Pedixplorer* et *nf-core/phaseimpute*. Ces deux projets de développements bio-informatiques ont été conçus selon des principes exigeants : code ouvert, versionné, documenté, reproductible et testé sur des jeux de données représentatifs. Ce cadre garantit la fiabilité à long terme des analyses, au prix d'un investissement initial important.

Le développement de *Pedixplorer* m'a demandé plus de quatre mois et celui de *nf-core/phaseimpute* plus de six mois, sans compter les semaines de maintenance (e.g., résolution de bogues, mises à jour de dépendances, ajouts de fonctionnalités, etc.).

J'aurais pu, comme cela se fait souvent, me contenter d'utiliser les scripts internes ou d'écrire des commandes ponctuelles écrite en *Bash*, adaptées à mes données et à mes besoins immédiats. Cette approche aurait certainement été plus rapide à mettre en œuvre et aurait permis de produire des résultats plus rapidement. Cependant, ce type de solution « maison », bien qu'efficace sur le moment, est souvent peu reproductible, difficilement partageable et fragile face à l'évolution des outils ou des formats de données. Une erreur de configuration, un mauvais paramétrage ou simplement une absence de documentation peut compromettre la validité des résultats et rendre leur traçabilité impossible.

En comparaison, les outils développés dans un cadre standardisé et communautaire, bien que plus longs et exigeants à concevoir, garantissent une meilleure fiabilité, une meilleure reproductibilité et une pérennité du code dans le temps.

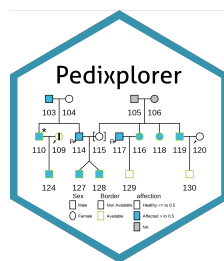
Avec mon expérience initialement limitée en génomique computationnelle, j'ai très vite compris l'importance de pouvoir m'appuyer sur une communauté structurée : des développeurs, des bio-informaticiens et des experts du domaine, qui contribuent à tester, valider et faire évoluer ces outils bio-informatiques.

Ce choix, entre rapidité et robustesse, a été structurant dans ma démarche. Il m'a conduit à investir davantage de temps dans le développement d'outils comme *Pedixplorer* ou *nf-core/phaseimpute*, mais avec la conviction que cela profite à la fois à mon projet et à ceux qui suivront, voire à la communauté scientifique de ce domaine.

Cette approche a également renforcé la fiabilité de mes propres analyses. Par exemple, pour l'alignement des séquences sur le génome canin, j'ai utilisé le pipeline *nf-core/sarek* (Hanssen et al., 2024). Celui-ci a fonctionné immédiatement, a réalisé l'ensemble des contrôles qualité et généré un rapport clair et interprétable. Cette étape est aussi aisément reproductible, simplement en fournissant le fichier de configuration *ad hoc* (i.e., format *yaml*) et la version exacte de *nf-core/sarek* utilisée.



(a) nf-core/phaseimpute



(b) Pedixplorer



(c) files2db

FIGURE 27 – Logos des trois outils développés et publiés lors de cette thèse

3 Développement et essaimage des outils

Au cours de cette thèse, j'ai développé quatre outils bio-informatiques, dont trois pour lesquels le code est public (voir figure 27), dans des langages variés, afin de répondre à des besoins spécifiques non couverts par des outils existants :

- *nf-core/phaseimpute* (*Nextflow/Bash*), un pipeline d'imputation génomique adapté aux données de séquençage *low-pass* ;
- *Pedixplorer* (*R*), un package pour la construction et l'exploration interactive de pedigrees complexes ;
- *files2db* (*Python*), un outil permettant l'agrégation automatisé de données hétérogènes ;
- un ensemble de scripts d'exploration (*R*) de la littérature scientifique permettant de comparer des loci entre espèces.

N'ayant initialement qu'une expérience modeste en développement bio-informatique, j'ai choisi de développer et de diffuser ces outils au sein de communautés bio-informatiques actives afin de bénéficier d'un retour critique sur mon code et d'en garantir la qualité. Ce choix s'est révélé essentiel pour l'apprentissage, mais aussi pour assurer la reproductibilité et la qualité de mes contributions, ainsi que leur pérennité et diffusion à la communauté en retour.

Au sein de *nf-core*, chaque composant (module, sous-workflow) est développé dans un dépôt dédié, avant d'être intégré dans différents pipelines via une *pull request* (PR), c'est-à-dire une proposition formelle de modification permettant la revue collaborative du code. Chaque PR fait l'objet de revues de code détaillées, de tests automatiques et de discussions techniques approfondies sur les bonnes pratiques (e.g., nommage, cohérence, logique d'implémentation, documentation).

Cette démarche collaborative m'a permis d'acquérir une maîtrise pratique de *Git* (i.e., système de gestion de versions distribué) ainsi que des principes d'intégration continue (CI), consistant à tester et valider automatiquement chaque modification de code afin d'assurer la stabilité du projet. J'ai également pu automatiser ces tests via *GitHub Actions*, un outil facilitant la création de *workflows* de CI/CD (intégration et déploiement

continus). Ces compétences ont non seulement enrichi le développement des outils étudiés, mais elles m'ont également permis de gérer efficacement mes autres projets de développement informatique, en garantissant rigueur, reproductibilité et qualité du code.

La bibliothèque *Pedixplorer* s'inscrit dans l'écosystème *R/Bioconductor*. Elle a été conçue de manière plus isolée car l'interaction communautaire au sein de *Bioconductor* est plus limitée que pour *Nextflow*. Cela est cependant compensé par une documentation exhaustive. J'ai pu échanger avec plusieurs membres du projet via le réseau professionnel *Slack*. Ces discussions ont porté principalement sur l'implémentation de la classe *S4* *Pedigree*, élément central du package. Elles m'ont amené à revoir en profondeur certaines parties de mon code pour en améliorer la robustesse, l'ergonomie et la conformité aux standards *Bioconductor*. J'y ai également intégré une batterie de tests unitaires et un système d'évaluation automatisée pour chaque nouvelle mise à jour. Le comportement des modules *shiny* reste cependant plus difficile à tester : malgré l'usage de *shinytest2*, il est complexe d'assurer leur bon fonctionnement sur toutes les architectures et navigateurs utilisés par la communauté. Les tests sont actuellement réalisés sous l'environnement *Windows* et *Linux* avec le navigateur *Chrome* et sont ainsi recommandés.

Enfin, mes deux dernières contributions (i.e., *files2db* et les scripts de comparaison interspèces) sont pour l'instant uniquement disponibles sur mon dépôt *GitHub* personnel. Un travail de généralisation reste à faire pour les adapter à d'autres types de projets. Je prévois de les publier à moyen terme, après la soutenance de thèse, une fois stabilisés et correctement documentés.

CONCLUSION

Les premières bases de ce projet ont été lancées dès 2008 par la collecte de prélèvements de chiens atteints et indemnes dans plusieurs races, projet monté en collaboration et avec le soutien financier de l'ACGAO, qui après une interruption de quelques années a repris avec mon projet de thèse, grâce à un financement de la Fondation Visio.

Les objectifs de ce projet étaient :

- le développement d'un test génétique de risque à partir des résultats de génétique obtenus dans les races en question ;
- la comparaison de la physiopathologie et de la génétique des dysplasies de la hanche chez le chien (i.e., Dysplasie Coxo-Fémorale : DCF) et l'humain (i.e., Dysplasie Développementale de la Hanche : DDH), grâce à une collaboration de longue date avec l'équipe GGB, INSERM-CHU de Brest du Pr Claude Ferec, puis du Dr Emmanuelle Génin.

Comme décrit dans les résultats, les différents sous-objectifs pour mener à bien les analyses génétiques ont pu être réalisés pendant ma thèse dans les 5 races (i.e., Golden Retriever, Labrador Retriever, Bouvier Bernois, Cane Corso et Dogue de Bordeaux). Certes, la collecte de cas et de contrôles dans ces races avait débuté dès les années 2010, mais les données étaient éparses et parcellaires. Néanmoins tous les protocoles étaient rédigés et la collecte des prélèvements et des informations avec l'ACGAO a pu reprendre ou elle avait été arrêtée en 2015.

La première année de thèse fut consacrée à la structuration des données en un fichier unique et à la formalisation de la collecte des informations phénotypiques et généalogiques. Les deux années suivantes ont été dédiées au développement d'outils bio-informatiques destinés au traitement des données de génétique. Dans un premier temps, *Pedixplorer* a été créé pour gérer les données de pedigree, suivi de *nf-core/phaseimpute* afin d'imputer les données de séquençage à faible couverture sur le génome de référence *canFam4*.

Le développement de ces outils, mené au cours de la deuxième et de la troisième année, a représenté un véritable défi : il a fallu les concevoir, coder, tester, corriger en interaction avec les communautés scientifiques concernées, puis les appliquer sur des jeux de données tests avant de les utiliser pour nos propres analyses. Cette démarche s'est inscrite dans une logique de science ouverte et dans le respect des principes FAIR

(i.e., *Findable, Accessible, Interoperable, Reusable*). Les outils ont été conçus pour être documentés, standardisés et accessibles publiquement afin de favoriser leur réutilisation par d'autres équipes. Leur intégration dans des écosystèmes collaboratifs (i.e., communautés *Bioconductor* et *nf-core*) garantit leur interopérabilité avec d'autres pipelines existants et leur amélioration continue grâce aux retours de la communauté. Ainsi, au-delà de leur utilité immédiate pour ce projet, ces outils constituent des ressources durables, contribuant à la transparence, à la reproductibilité et à la mutualisation des efforts en bio-informatique.

Une autre composante essentielle de cette thèse a été la collecte de prélèvements biologiques, menée tout au long du projet. Grâce à l'implication du Dr vétérinaire Lauranne Machou, cette collecte a pu être structurée et suivie en collaboration avec de nombreux partenaires. Nous avons pu alors sélectionner plus de 900 chiens avec données phénotypiques dont une solide cohorte de chiens guides de l'ACGAO (Golden et Labrador Retriever) formés depuis 2003, qui ont été prélevés et suivi depuis 2008. Cette collection d'ADN associée à des données phénotypiques de ces chiens est une ressource importante en soi, stockée au CRB Cani-DNA.

Un total de 707 ADN, issus des cinq races étudiées, a été dosé et préparé avant d'être séquencé en LP-WGS (1× de couverture) par la société *Psomagen* (États-Unis). Les données ainsi obtenues ont ensuite été exploitées pour réaliser les analyses génétiques.

Les loci identifiés au cours des différentes analyses sont particulièrement prometteurs. Certains confirment des associations déjà décrites dans la littérature canine (6 loci en CFA 9, 10, 17 et 22) que humaine (4 loci en CFA 9, 19 et 22) et 9 représentent de nouveaux loci (CFA 3, 5, 6, 7, 17 et 34). Toutefois, une réplication indépendante de ces résultats à partir d'une deuxième cohorte sera nécessaire pour confirmer ou bien infirmer certains locus. Cela permettra d'estimer de manière fiable le pouvoir prédictif des variants qui seront retenus pour développer un test génétique de risque.

En effet, cette recherche a pour objectif de développer un test génétique de risque, afin d'aider les éleveurs des races étudiées à mieux gérer l'élevage, par le dépistage des chiens à risque pour orienter les choix de reproduction en excluant les reproducteurs à fort risque et d'accompagner l'ACGAO dans la sélection des chiots les moins à risque, destinés à devenir chiens guides.

Ce projet de recherche se poursuivra par la réplication des résultats sur une cohorte indépendante, puis par la sélection d'une vingtaine de marqueurs parmi les loci les plus prometteurs, afin de développer un test génétique de prédiction du risque. Les variants génétiques qui seront sélectionnés pour le test ne seront probablement pas les variants impliqués eux-mêmes, en raison du fort déséquilibre de liaison propre à chaque race, qui étend les associations à des régions plus larges. Il sera donc essentiel de séquencer

finement les loci d'intérêt, puis d'annoter l'ensemble des variants présents, dans l'espoir d'identifier les mutations des différents loci impliquées dans cette affection. En parallèle, une approche fonctionnelle sera donc envisagée, pour contribuer à sélectionner les meilleurs gènes candidats et loci et comprendre le lien entre altérations génétiques et physiopathologie.

Cette étape consistera en une analyse transcriptomique des tissus articulaires issus de chiens atteints et sains, collectés au cours de ma thèse. Ce travail permettra de mettre en évidence les voies biologiques perturbées au cours du développement de la dysplasie et de les relier, lorsque possible, aux variants identifiés. À terme, ces résultats pourront contribuer à une meilleure compréhension des liens entre génotype et phénotype et ouvrir la voie à l'identification de nouveaux biomarqueurs ou de cibles thérapeutiques pour cette affection locomotrice débilante chez le chien et l'humain.

Enfin, ce projet de thèse se place dans le contexte *One Health* (i.e., une seule santé) entre chiens et humains, par une collaboration de longue date avec l'équipe du Pr Claude Ferec et maintenant du Dr Emmanuelle Génin, qui dirige le laboratoire INSERM GGB à Brest. En effet, cette équipe a collecté depuis de nombreuses années des cas de DDH chez des Bigoudens (Hôpital de Pont Labbé) et cette cohorte est en cours de séquençage. Comme je poursuis mon parcours scientifique par un post-doctorat dans leur équipe, je serais au fait de ces analyses chez l'humain et pourrai participer aux comparaisons entre les analyses GWAS faites à Brest chez les Bigoudens et à Rennes chez les chiens guides et autres races canines.

Ce projet se prolongera dans le cadre d'une seconde thèse menée au sein de l'équipe « Génétique du Chien », grâce au soutien renouvelé de la Fondation Visio et de l'AC-GAO. Cela permet d'assurer la continuité des analyses, le développement d'un test génétique pour les races canines considérées et la valorisation des outils de génétique computationnelle développés au cours de ma thèse.

BIBLIOGRAPHY



BIBLIOGRAPHIE

- ACADÉMIE NATIONALE DE MÉDECINE, 2025a. Dictionnaire de l'Académie Nationale de Médecine: Dysplasie [<http://91.209.229.113/search/results?titre=dysplasie>]. [visité le 2025-04-15].
- ACADÉMIE NATIONALE DE MÉDECINE, 2025b. Dictionnaire de l'Académie Nationale de Médecine: Dysplasie Coxo-Fémorale [<http://91.209.229.113/search/results?titre=dysplasie%20coxo%C3%A9morale>]. [visité le 2025-04-15].
- ADAMS, W. et al., 1998. Early Detection of Canine Hip Dysplasia: Comparison of Two Palpation and Five Radiographic Methods. *Journal of the American Animal Hospital Association* [en ligne]. Vol. 34, no. 4, p. 339–347 [visité le 2020-08-05]. ISSN 0587-2871. Disp. à l'adr. DOI: 10.5326/15473317-34-4-339.
- AHN, K.-S. et al., 2025. Deep Learning-Based Automated Guide for Defining a Standard Imaging Plane for Developmental Dysplasia of the Hip Screening Using Ultrasonography: A Retrospective Imaging Analysis. *BMC Medical Informatics and Decision Making* [en ligne]. Vol. 25, no. 1, p. 91 [visité le 2025-02-21]. ISSN 1472-6947. Disp. à l'adr. DOI: 10.1186/s12911-025-02926-8.
- ALEXANDER, J. W., 1992. The Pathogenesis of Canine Hip Dysplasia. *Veterinary Clinics of North America: Small Animal Practice* [en ligne]. Vol. 22, no. 3, p. 503–511 [visité le 2020-08-05]. ISSN 0195-5616. Disp. à l'adr. DOI: 10.1016/S0195-5616(92)50051-1.
- AMERICAN KENNEL CLUB, 1997. *The Complete Dog Book* [en ligne]. New York : Howell Book House [visité le 2025-05-04]. ISBN 978-0-87605-047-7.
- ANDRÉ, C. et al., 2014. Methods for Diagnosing Skin Diseases [en ligne]. [visité le 2025-05-18]. EP2501825B1.
- ANDRÉ, C. et al., 2022. Cani-DNA, un CRB qui a du chien! Réseau de collecte de prélèvements de chiens par les vétérinaires pour la recherche biomédicale et la diversité génétique. *NOV'AE - Ingénierie et savoir-faire innovants* [en ligne]. No. spécial 02 NOV'AE, p. 34–44 [visité le 2025-03-13]. ISSN 2823-3980. Disp. à l'adr. DOI: 10.17180/novae-2022-NS02-art05.
- ARENDT, M. et al., 2016. Diet Adaptation in Dog Reflects Spread of Prehistoric Agriculture. *Heredity* [en ligne]. Vol. 117, no. 5, p. 301–306 [visité le 2025-05-03]. ISSN 1365-2540. Disp. à l'adr. DOI: 10.1038/hdy.2016.48.
- Article R213-2, 2003 [en ligne]. [visité le 2025-09-09]. No. 213-2.
- AUTON, A. et al., 2013. Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *PLOS Genetics* [en ligne]. Vol. 9, no. 12, e1003984 [visité le 2025-06-30]. ISSN 1553-7404. Disp. à l'adr. DOI: 10.1371/journal.pgen.1003984.
- BADANO, J. L. & KATSANIS, N., 2002. Beyond Mendel: An Evolving View of Human Genetic Disease Transmission. *Nature Reviews Genetics* [en ligne]. Vol. 3, no. 10, p. 779–789 [visité le 2025-05-17]. ISSN 1471-0064. Disp. à l'adr. DOI: 10.1038/nrg910.
- BAI, B. et al., 2015. DoGSD: The Dog and Wolf Genome SNP Database. *Nucleic Acids Research*. Vol. 43, no. Database issue, p. D777–783. ISSN 1362-4962. Disp. à l'adr. DOI: 10.1093/nar/gku1174.

-
- BAILEY-WILSON, J. E. & WILSON, A. F., 2011. Linkage Analysis in the Next-Generation Sequencing Era. *Human Heredity* [en ligne]. Vol. 72, no. 4, p. 228–236 [visit  le 2025-05-22]. ISSN 0001-5652. Disp.   l'adr. DOI: 10.1159/000334381.
- BAKER, M., 2016. 1,500 Scientists Lift the Lid on Reproducibility. *Nature* [en ligne]. Vol. 533, no. 7604, p. 452–454 [visit  le 2025-05-23]. ISSN 1476-4687. Disp.   l'adr. DOI: 10.1038/533452a.
- BARTOLOM , N. et al., 2015. A Genetic Predictive Model for Canine Hip Dysplasia: Integration of Genome Wide Association Study (GWAS) and Candidate Gene Approaches. *PLoS One*. Vol. 10, no. 4, e0122558.
- BENECKE, N., 1987. Studies on Early Dog Remains from Northern Europe. *Journal of Archaeological Science* [en ligne]. Vol. 14, no. 1, p. 31–49 [visit  le 2025-05-01]. ISSN 0305-4403. Disp.   l'adr. DOI: 10.1016/S0305-4403(87)80004-3.
- BERGSTR M, A. et al., 2020. Origins and Genetic Legacy of Prehistoric Dogs. *Science* (New York, N.Y.) [en ligne]. Vol. 370, no. 6516, p. 557–564 [visit  le 2025-05-04]. ISSN 0036-8075. Disp.   l'adr. DOI: 10.1126/science.aba9572.
- BERGSTR M, A. et al., 2022. Grey Wolf Genomic History Reveals a Dual Ancestry of Dogs. *Nature* [en ligne]. Vol. 607, no. 7918, p. 313–320 [visit  le 2025-05-02]. ISSN 1476-4687. Disp.   l'adr. DOI: 10.1038/s41586-022-04824-9.
- BINVERSIE, E. E. et al., 2022a. Across-Breed Genetic Investigation of Canine Hip Dysplasia, Elbow Dysplasia, and Anterior Cruciate Ligament Rupture Using Whole-Genome Sequencing. *Frontiers in Genetics* [en ligne]. Vol. 13 [visit  le 2022-12-06]. ISSN 1664-8021.
- BINVERSIE, E. E. et al., 2022b. Canine ACL Rupture: A Spontaneous Large Animal Model of Human ACL Rupture. *BMC Musculoskeletal Disorders* [en ligne]. Vol. 23, no. 1, p. 116 [visit  le 2022-02-11]. ISSN 1471-2474. Disp.   l'adr. DOI: 10.1186/s12891-021-04986-z.
- BIOCONDUCTOR PACKAGE, 2025. liftOver [en ligne]. Bioconductor. [visit  le 2025-07-28]. Disp.   l'adr. DOI: 10.18129/B9.BIOC.LIFTOVER.
- BJ RNERFELDT, S. et al., 2006. Relaxation of Selective Constraint on Dog Mitochondrial DNA Following Domestication. *Genome Research* [en ligne]. Vol. 16, no. 8, p. 990–994 [visit  le 2025-05-04]. ISSN 1088-9051. Disp.   l'adr. DOI: 10.1101/gr.5117706.
- BLACK, M. et al., 2022. <p>EDAM: The Bioscientific Data Analysis Ontology (Update 2021)</P>. F1000Research [en ligne]. Vol. 11 [visit  le 2025-07-25]. Disp.   l'adr. DOI: 10.7490/f1000research.1118900.1.
- BORGE, K. S. et al., 2011. Litter Size at Birth in Purebred Dogs—a Retrospective Study of 224 Breeds. *Theriogenology*. Vol. 75, no. 5, p. 911–919. ISSN 1879-3231. Disp.   l'adr. DOI: 10.1016/j.theriogenology.2010.10.034.
- BOTIGU , L. R. et al., 2017. Ancient European Dog Genomes Reveal Continuity since the Early Neolithic. *Nature Communications* [en ligne]. Vol. 8, p. 16082 [visit  le 2025-05-01]. ISSN 2041-1723. Disp.   l'adr. DOI: 10.1038/ncomms16082.
- BOUFENAR, C. et al., 2023. Computer-Aided Diagnosis of Canine Hip Dysplasia Using Deep Learning Approach in a Novel X-ray Image Dataset. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* [en ligne]. Vol. 0, no. 0, p. 1–11 [visit  le 2023-11-09]. ISSN 2168-1163. Disp.   l'adr. DOI: 10.1080/21681163.2023.2274947.

-
- BOYKO, A. R., 2011. The Domestic Dog: Man's Best Friend in the Genomic Era. *Genome Biology* [en ligne]. Vol. 12, no. 2, p. 216 [visit  le 2025-05-04]. Disp.   l'adr. DOI: 10.1186/gb-2011-12-2-216.
- BOYKO, A. R. et al., 2009. Complex Population Structure in African Village Dogs and Its Implications for Inferring Dog Domestication History. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 106, no. 33, p. 13903–13908. ISSN 1091-6490. Disp.   l'adr. DOI: 10.1073/pnas.0902129106.
- BRAND, R. A., 2008. Etiology of Congenital Dislocation of the Hip: Carl E. Badgley MD (1893-1973) The 11th President of the AAOS 1942. *Clinical Orthopaedics & Related Research* [en ligne]. Vol. 466, no. 1, p. 90–103 [visit  le 2022-11-04]. ISSN 0009-921X. Disp.   l'adr. DOI: 10.1007/s11999-007-0020-7.
- BREEN, M. et al., 2001. Chromosome-Specific Single-Locus FISH Probes Allow Anchorage of an 1800-Marker Integrated Radiation-Hybrid/Linkage Map of the Domestic Dog Genome to All Chromosomes. *Genome Research* [en ligne]. Vol. 11, no. 10, p. 1784–1795 [visit  le 2020-09-04]. ISSN 1088-9051, ISSN 1549-5469. Disp.   l'adr. DOI: 10.1101/gr.189401.
- BREEN, M. et al., 2004. An Integrated 4249 Marker FISH/RH Map of the Canine Genome. *BMC Genomics* [en ligne]. Vol. 5, no. 1, p. 65 [visit  le 2022-11-10]. ISSN 1471-2164. Disp.   l'adr. DOI: 10.1186/1471-2164-5-65.
- BROWN, S. K. et al., 2011. Phylogenetic Distinctiveness of Middle Eastern and Southeast Asian Village Dog Y Chromosomes Illuminates Dog Origins. *PloS One*. Vol. 6, no. 12, e28496. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0028496.
- BROWNING, B. L. et al., 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics*. Vol. 103, no. 3, p. 338–348. ISSN 1537-6605. Disp.   l'adr. DOI: 10.1016/j.ajhg.2018.07.015.
- BROWNING, B. L. et al., 2021. Fast Two-Stage Phasing of Large-Scale Sequence Data. *The American Journal of Human Genetics* [en ligne]. Vol. 108, no. 10, p. 1880–1890 [visit  le 2025-08-20]. ISSN 0002-9297. Disp.   l'adr. DOI: 10.1016/j.ajhg.2021.08.005.
- BUCKLEY, R. M. et al., 2021. Best Practices for Analyzing Imputed Genotypes from Low-Pass Sequencing in Dogs. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*. ISSN 1432-1777. Disp.   l'adr. DOI: 10.1007/s00335-021-09914-z.
- BULIK-SULLIVAN, B. K. et al., 2015. LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. *Nature Genetics*. Vol. 47, no. 3, p. 291–295. ISSN 1546-1718. Disp.   l'adr. DOI: 10.1038/ng.3211.
- BUNEL, M. et al., 2019. Natural Models for Retinitis Pigmentosa: Progressive Retinal Atrophy in Dog Breeds. *Human Genetics* [en ligne]. Vol. 138, no. 5, p. 441–453 [visit  le 2020-07-24]. ISSN 1432-1203. Disp.   l'adr. DOI: 10.1007/s00439-019-01999-6.
- CAMACHO, P. et al., 2016. Large Mammalian Animal Models of Heart Disease. *Journal of Cardiovascular Development and Disease* [en ligne]. Vol. 3, no. 4, p. 30 [visit  le 2025-05-18]. ISSN 2308-3425. Disp.   l'adr. DOI: 10.3390/jcdd3040030.
- CAMAR S, E. et al., 2016. The Evolution of Paleolithic Hominin–Carnivore Interaction Written in Teeth: Stories from the Swabian Jura (Germany). *Journal of Archaeological Science: Reports* [en ligne]. Vol. 6, p. 798–809 [visit  le 2025-05-01]. ISSN 2352-409X. Disp.   l'adr. DOI: 10.1016/j.jasrep.2015.11.010.

-
- CANO-GAMEZ, E. & TRYNKA, G., 2020. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics* [en ligne]. Vol. 11, p. 424 [visité le 2025-05-23]. ISSN 1664-8021. Disp. à l'adr. DOI: 10.3389/fgene.2020.00424.
- CARDON, L. R. & BELL, J. I., 2001. Association Study Designs for Complex Diseases. *Nature Reviews Genetics* [en ligne]. Vol. 2, no. 2, p. 91–99 [visité le 2025-05-21]. ISSN 1471-0064. Disp. à l'adr. DOI: 10.1038/35052543.
- CHANG, W. et al., 2024. Shiny: Web Application Framework for R. Version 1.10.0. Aussi disponible à l'adresse: <https://shiny.posit.co/>.
- CHASE, K. et al., 2004. Bilaterally Asymmetric Effects of Quantitative Trait Loci (QTLs): QTLs That Affect Laxity in the Right versus Left Coxofemoral (Hip) Joints of the Dog (*Canis Familiaris*). *American Journal of Medical Genetics Part A*. Vol. 124, no. 3, p. 239–247.
- CHEN, Y.-P. et al., 2023. Automatic and Human Level Graf's Type Identification for Detecting Developmental Dysplasia of the Hip. *Biomedical Journal* [en ligne], p. 100614 [visité le 2023-06-13]. ISSN 2319-4170. Disp. à l'adr. DOI: 10.1016/j.bj.2023.100614.
- CHEN, Z. et al., 2021. Revisiting the Genome-Wide Significance Threshold for Common Variant GWAS. *G3: Genes|Genomes|Genetics* [en ligne]. Vol. 11, no. 2, jkaa056 [visité le 2025-07-01]. ISSN 2160-1836. Disp. à l'adr. DOI: 10.1093/g3journal/jkaa056.
- CHENG, B. et al., 2021. Integrative Analysis of MicroRNA and mRNA Sequencing Data Identifies Novel Candidate Genes and Pathways for Developmental Dysplasia of Hip. *CARTILAGE* [en ligne], p. 1947603521990859 [visité le 2021-09-02]. ISSN 1947-6035. Disp. à l'adr. DOI: 10.1177/1947603521990859.
- COPPINGER, R. & COPPINGER, L., 2001. *Dogs-A Startling New Understanding of Canine Origin, Behavior & Evolution*. Bibliovault OAI Repository, the University of Chicago Press.
- CRUZ, F. et al., 2008. The Legacy of Domestication: Accumulation of Deleterious Mutations in the Dog Genome. *Molecular Biology and Evolution* [en ligne]. Vol. 25, no. 11, p. 2331–2336 [visité le 2025-05-04]. ISSN 0737-4038. Disp. à l'adr. DOI: 10.1093/molbev/msn177.
- CYRIL CHAPMAN et al., 1999. Cyrillic. Cherwell Scientific Limited. Version 3.0.400. Aussi disponible à l'adresse: <https://www.apbenson.com/about-cyrillic-3>.
- D'AQUIN, M. et al., 2023. FAIREST: A Framework for Assessing Research Repositories. *Data Intelligence* [en ligne]. Vol. 5, no. 1, p. 202–241 [visité le 2025-05-23]. ISSN 2641-435X. Disp. à l'adr. DOI: 10.1162/dint_a_00159.
- DAVIES, R. W. et al., 2016. Rapid Genotype Imputation from Sequence without Reference Panels. *Nature Genetics* [en ligne]. Vol. 48, no. 8, p. 965–969 [visité le 2024-03-15]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/ng.3594.
- DAVIES, R. W. et al., 2021. Rapid Genotype Imputation from Sequence with Reference Panels. *Nature Genetics* [en ligne]. Vol. 53, no. 7, p. 1104–1111 [visité le 2024-03-15]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-021-00877-0.
- DE RISIO, L. et al., 2011. Prevalence, Heritability and Genetic Correlations of Congenital Sensorineural Deafness and Pigmentation Phenotypes in the Border Collie. *Veterinary Journal (London, England: 1997)*. Vol. 188, no. 3, p. 286–290. ISSN 1532-2971. Disp. à l'adr. DOI: 10.1016/j.tvjl.2010.05.012.

-
- DERRIEN, T., 2007. L'analyse Comparee Des Genomes : Applications à l'identification de Nouveaux Gènes Canins [en ligne]. [visit  le 2022-11-10]. Theses. Universit  Rennes 1.
- DERRIEN, T. et al., 2009. Revisiting the Missing Protein-Coding Gene Catalog of the Domestic Dog. BMC Genomics [en ligne]. Vol. 10, no. 1, p. 62 [visit  le 2025-07-22]. ISSN 1471-2164. Disp.   l'adr. DOI: 10.1186/1471-2164-10-62.
- DEVLIN, B. & ROEDER, K., 1999. Genomic Control for Association Studies. Biometrics. Vol. 55, no. 4, p. 997–1004. ISSN 0006-341X. Disp.   l'adr. DOI: 10.1111/j.0006-341x.1999.00997.x.
- DING, Z.-L. et al., 2012. Origins of Domestic Dog in Southern East Asia Is Supported by Analysis of Y-chromosome DNA. Heredity [en ligne]. Vol. 108, no. 5, p. 507–514 [visit  le 2025-05-01]. ISSN 1365-2540. Disp.   l'adr. DOI: 10.1038/hdy.2011.114.
- DOBNEY, K. & LARSON, G., 2006. Genetics and Animal Domestication: New Windows on an Elusive Process. Journal of Zoology [en ligne]. Vol. 269, no. 2, p. 261–271 [visit  le 2025-05-03]. ISSN 1469-7998. Disp.   l'adr. DOI: 10.1111/j.1469-7998.2006.00042.x.
- DOWNS, J. F., 1960. Domestication : An Examination of the Changing Social Relationships between Man and Animals. In: [en ligne] [visit  le 2025-05-03].
- DYER, S. C. et al., 2025. Ensembl 2025. Nucleic Acids Research [en ligne]. Vol. 53, no. D1, p. D948–D957 [visit  le 2025-05-17]. ISSN 1362-4962. Disp.   l'adr. DOI: 10.1093/nar/gkae1071.
- EIZENGA, J. M. et al., 2020. Pangenome Graphs. Annual Review of Genomics and Human Genetics. Vol. 21, p. 139–162. ISSN 1545-293X. Disp.   l'adr. DOI: 10.1146/annurev-genom-120219-080406.
- EWELS, P. A. et al., 2020. The Nf-Core Framework for Community-Curated Bioinformatics Pipelines. Nature Biotechnology [en ligne]. Vol. 38, no. 3, p. 276–278 [visit  le 2023-03-16]. ISSN 1546-1696. Disp.   l'adr. DOI: 10.1038/s41587-020-0439-x.
- FCI Breeds Nomenclature, 2025 [<https://fci.be/en/Nomenclature/>]. [visit  le 2025-05-04].
- FELDMAN, G. J. et al., 2014. Linkage Mapping and Whole Exome Sequencing Identify a Shared Variant in CX3CR1 in a Large Multi-Generation Family. The Journal of Arthroplasty [en ligne]. Vol. 29, no. 9, Supplement, p. 238–241 [visit  le 2022-11-04]. ISSN 0883-5403. Disp.   l'adr. DOI: 10.1016/j.arth.2014.05.014.
- FERNANDES, M. et al., 2016. CanisOme–The Protein Signatures of Canis Lupus Familiaris Diseases. Journal of Proteomics. Vol. 136, p. 193–201. ISSN 1876-7737. Disp.   l'adr. DOI: 10.1016/j.jprot.2016.01.005.
- FRANTZ, L. A. F. et al., 2016. Genomic and Archaeological Evidence Suggest a Dual Origin of Domestic Dogs. Science (New York, N.Y.) Vol. 352, no. 6290, p. 1228–1231. ISSN 1095-9203. Disp.   l'adr. DOI: 10.1126/science.aaf3161.
- FREEDMAN, A. H. et al., 2014. Genome Sequencing Highlights the Dynamic Early History of Dogs. PLoS Genetics [en ligne]. Vol. 10, no. 1, e1004016 [visit  le 2025-05-01]. ISSN 1553-7390. Disp.   l'adr. DOI: 10.1371/journal.pgen.1004016.
- FRIEDENBERG, S. G. et al., 2011. Evaluation of a Fibrillin 2 Gene Haplotype Associated with Hip Dysplasia and Incipient Osteoarthritis in Dogs. American journal of veterinary research. Vol. 72, no. 4, p. 530–540.

-
- GALIBERT, F. et al., 2011. Toward Understanding Dog Evolutionary and Domestication History. *Comptes Rendus Biologies*. Vol. 334, no. 3, p. 190–196. ISSN 1768-3238. Disp. à l'adr. DOI: 10.1016/j.crvi.2010.12.011.
- GALTON, F., 1865. The First Steps towards the Domestication of Animals. *Transactions of the Ethnological Society of London* [en ligne]. Vol. 3, p. 122–138 [visité le 2025-05-03]. ISSN 1368-0366. Disp. à l'adr. DOI: 10.2307/3014161.
- GARDNER, H. L. et al., 2016. Dogs as a Model for Cancer. *Annual review of animal biosciences* [en ligne]. Vol. 4, p. 199–222 [visité le 2022-11-10]. ISSN 2165-8102. Disp. à l'adr. DOI: 10.1146/annurev-animal-022114-110911.
- GENEVOIS, J. .-, 2018. Dépistage Précoce de La Dysplasie Coxo-Fémorale: Indication, Procédés Résultats. *La Revue de l'Association Française du Berger Blanc*. No. 10, p. 10–12.
- GENEVOIS, J.-P. et al., 2008. Prevalence of Hip Dysplasia According to Official Radiographic Screening, among 31 Breeds of Dogs in France. *Veterinary and Comparative Orthopaedics and Traumatology: V.C.O.T.* Vol. 21, no. 1, p. 21–24. ISSN 0932-0814. Disp. à l'adr. DOI: 10.3415/VCOT-07-02-0014.
- GENEVOIS, J.-P., 2020. Correspondance Avec Le Professeur Genevois.
- GERMONPRÉ, M. et al., 2009. Fossil Dogs and Wolves from Palaeolithic Sites in Belgium, the Ukraine and Russia: Osteometry, Ancient DNA and Stable Isotopes. *Journal of Archaeological Science* [en ligne]. Vol. 36, no. 2, p. 473–490 [visité le 2025-05-01]. ISSN 0305-4403. Disp. à l'adr. DOI: 10.1016/j.jas.2008.09.033.
- GERMONPRÉ, M. et al., 2012. Palaeolithic Dog Skulls at the Gravettian Předmostí Site, the Czech Republic. *Journal of Archaeological Science* [en ligne]. Vol. 39, no. 1, p. 184–202 [visité le 2025-05-01]. ISSN 0305-4403. Disp. à l'adr. DOI: 10.1016/j.jas.2011.09.022.
- GHIRLANDA, S. et al., 2013. Fashion vs. Function in Cultural Evolution: The Case of Dog Breed Popularity. *PLOS ONE* [en ligne]. Vol. 8, no. 9, e74770 [visité le 2025-05-04]. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0074770.
- GIAMBARTOLOMEI, C. et al., 2014. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics* [en ligne]. Vol. 10, no. 5, e1004383 [visité le 2025-07-02]. ISSN 1553-7404. Disp. à l'adr. DOI: 10.1371/journal.pgen.1004383.
- GIGER, U. et al., 2006. Breed-Specific Hereditary Diseases and Genetic Screening. *The Dog and Its Genome*, p. 249–289.
- GKIATAS, I. et al., 2019. Developmental Dysplasia of the Hip: A Systematic Literature Review of the Genes Related with Its Occurrence. *EFORT open reviews*. Vol. 4, no. 10, p. 595–601. ISSN 2058-5241. Disp. à l'adr. DOI: 10.1302/2058-5241.4.190006.
- GOU, X. et al., 2014. Whole-Genome Sequencing of Six Dog Breeds from Continuous Altitudes Reveals Adaptation to High-Altitude Hypoxia. *Genome Research*. Vol. 24, no. 8, p. 1308–1315. ISSN 1549-5469. Disp. à l'adr. DOI: 10.1101/gr.171876.113.
- GRALL, A. et al., 2012. PNPLA1 Mutations Cause Autosomal Recessive Congenital Ichthyosis in Golden Retriever Dogs and Humans. *Nature Genetics*. Vol. 44, no. 2, p. 140–147. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/ng.1056.

-
- GUAGUERE, E. et al., 2009. Clinical, Histopathological and Genetic Data of Ichthyosis in the Golden Retriever: A Prospective Study. *The Journal of Small Animal Practice*. Vol. 50, no. 5, p. 227–235. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.2009.00730.x.
- GUYON, R. et al., 2003. A 1-Mb Resolution Radiation Hybrid Map of the Canine Genome. *Proceedings of the National Academy of Sciences of the United States of America* [en ligne]. Vol. 100, no. 9, p. 5296–5301 [visité le 2022-01-04]. ISSN 0027-8424. Disp. à l'adr. DOI: 10.1073/pnas.0831002100.
- GYÓRÖDI, C. et al., 2015. A Comparative Study: MongoDB vs. MySQL. In: 2015 13th International Conference on Engineering of Modern Electric Systems (EMES) [en ligne], p. 1–6 [visité le 2025-08-18]. Disp. à l'adr. DOI: 10.1109/EMES.2015.7158433.
- HALO, J. V. et al., 2021. Long-Read Assembly of a Great Dane Genome Highlights the Contribution of GC-rich Sequence and Mobile Elements to Canine Genomes. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 118, no. 11, e2016274118. ISSN 1091-6490. Disp. à l'adr. DOI: 10.1073/pnas.2016274118.
- HANSEN, K. & KHANNA, C., 2004. Spontaneous and Genetically Engineered Animal Models; Use in Preclinical Cancer Drug Development. *European Journal of Cancer* (Oxford, England: 1990). Vol. 40, no. 6, p. 858–880. ISSN 0959-8049. Disp. à l'adr. DOI: 10.1016/j.ejca.2003.11.031.
- HANSSEN, F. et al., 2024. Scalable and Efficient DNA Sequencing Analysis on Different Compute Infrastructures Aiding Variant Discovery [en ligne]. *bioRxiv* [visité le 2025-08-19]. Disp. à l'adr. DOI: 10.1101/2023.07.19.549462.
- HARSANYI, S. et al., 2020a. Developmental Dysplasia of the Hip: A Review of Etiopathogenesis, Risk Factors, and Genetic Aspects. *Medicina*. Vol. 56, no. 4, p. 153.
- HARSANYI, S. et al., 2020b. Genetics of Developmental Dysplasia of the Hip. *European Journal of Medical Genetics* [en ligne]. Vol. 63, no. 9, p. 103990 [visité le 2021-02-02]. ISSN 1769-7212. Disp. à l'adr. DOI: 10.1016/j.ejmg.2020.103990.
- HAYWARD, J. J. et al., 2016. Complex Disease and Phenotype Mapping in the Domestic Dog. *Nature Communications* [en ligne]. Vol. 7, no. 1, p. 1–11 [visité le 2020-05-13]. ISSN 2041-1723. Disp. à l'adr. DOI: 10.1038/ncomms10460.
- HÉDAN, B. et al., 2006. Coat Colour in Dogs: Identification of the Merle Locus in the Australian Shepherd Breed. *BMC Veterinary Research* [en ligne]. Vol. 2, p. 9 [visité le 2025-05-18]. ISSN 1746-6148. Disp. à l'adr. DOI: 10.1186/1746-6148-2-9.
- HENNESSEY, E. et al., 2022. Artificial Intelligence in Veterinary Diagnostic Imaging: A Literature Review. *Veterinary Radiology & Ultrasound* [en ligne]. Vol. 63, no. S1, p. 851–870 [visité le 2022-12-20]. ISSN 1740-8261. Disp. à l'adr. DOI: 10.1111/vru.13163.
- HERVELLA, M. et al., 2022. The Domestic Dog That Lived ~17,000 Years Ago in the Lower Magdalenian of Erralla Site (Basque Country): A Radiometric and Genetic Analysis. *Journal of Archaeological Science: Reports* [en ligne]. Vol. 46, p. 103706 [visité le 2025-07-24]. ISSN 2352-409X. Disp. à l'adr. DOI: 10.1016/j.jasrep.2022.103706.
- HITTE, C. et al., 2005. Facilitating Genome Navigation: Survey Sequencing and Dense Radiation-Hybrid Gene Mapping. *Nature reviews Genetics* [en ligne]. Vol. 6, no. 8, p. 643–648 [visité le 2022-11-10]. ISSN 1471-0064. Disp. à l'adr. DOI: 10.1038/nrg1658.

-
- HOEPPNER, M. P. et al., 2014. An Improved Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding Transcripts. *PLOS ONE* [en ligne]. Vol. 9, no. 3, e91172 [visit  le 2025-05-09]. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0091172.
- HORMOZDIARI, F. et al., 2016. Colocalization of GWAS and eQTL Signals Detects Target Genes. *American Journal of Human Genetics*. Vol. 99, no. 6, p. 1245–1260. ISSN 1537-6605. Disp.   l'adr. DOI: 10.1016/j.ajhg.2016.10.003.
- HUANG, C. et al., 2017. FGWAS: Functional Genome Wide Association Analysis. *NeuroImage*. Vol. 159, p. 107–121. ISSN 1095-9572. Disp.   l'adr. DOI: 10.1016/j.neuroimage.2017.07.030.
- HUBER, W. et al., 2015. Orchestrating High-Throughput Genomic Analysis with Bioconductor. *Nature Methods* [en ligne]. Vol. 12, no. 2, p. 115–121 [visit  le 2025-05-23]. ISSN 1548-7105. Disp.   l'adr. DOI: 10.1038/nmeth.3252.
- H LSMEYER, V. et al., 2010. Epilepsy in Border Collies: Clinical Manifestation, Outcome, and Mode of Inheritance. *Journal of Veterinary Internal Medicine*. Vol. 24, no. 1, p. 171–178. ISSN 0891-6640. Disp.   l'adr. DOI: 10.1111/j.1939-1676.2009.0438.x.
- IOTCHKOVA, V. et al., 2019. GARFIELD Classifies Disease-Relevant Genomic Features through Integration of Functional Annotations with Association Signals. *Nature Genetics* [en ligne]. Vol. 51, no. 2, p. 343–353 [visit  le 2025-07-02]. ISSN 1546-1718. Disp.   l'adr. DOI: 10.1038/s41588-018-0322-6.
- JACOBSEN, K. K. et al., 2023. COL11A1 Is Associated with Developmental Dysplasia of the Hip and Secondary Osteoarthritis in the HUNT Study. *Osteoarthritis and Cartilage Open* [en ligne], p. 100424 [visit  le 2024-01-08]. ISSN 2665-9131. Disp.   l'adr. DOI: 10.1016/j.ocarto.2023.100424.
- JAGANNATHAN, V. et al., 2021. Dog10K_Boxer_Tasha_1.0: A Long-Read Assembly of the Dog Reference Genome. *Genes* [en ligne]. Vol. 12, no. 6, p. 847 [visit  le 2025-05-09]. ISSN 2073-4425. Disp.   l'adr. DOI: 10.3390/genes12060847.
- JAYASINGHE, D. et al., 2024. Advancements and Limitations in Polygenic Risk Score Methods for Genomic Prediction: A Scoping Review. *Human Genetics* [en ligne]. Vol. 143, no. 12, p. 1401–1431 [visit  le 2025-05-23]. ISSN 1432-1203. Disp.   l'adr. DOI: 10.1007/s00439-024-02716-8.
- JUNG, C. & P RTL, D., 2018. Scavenging Hypothesis: Lack of Evidence for Dog Domestication on the Waste Dump. *Dog behavior* [en ligne]. Vol. 4, no. 2, p. 41–56 [visit  le 2025-05-03]. ISSN 2421-5678. Disp.   l'adr. DOI: 10.4454/db.v4i2.73.
- KAHLE, B., 1997. Preserving the Internet. *Scientific American* [en ligne]. Vol. 276, no. 3, p. 82–83 [visit  le 2025-08-19]. ISSN 0036-8733. Disp.   l'adr. JSTOR: 24993660.
- KALLA, S. E. et al., 2020. Polymorphic SINEC_Cf Retrotransposons in the Genome of the Dog (*Canis Familiaris*) [en ligne]. *bioRxiv* [visit  le 2025-07-24]. Disp.   l'adr. DOI: 10.1101/2020.10.27.358119.
- KARLSSON, E. K. & LINDBLAD-TOH, K., 2008. Leader of the Pack: Gene Mapping in Dogs and Other Model Organisms. *Nature Reviews Genetics* [en ligne]. Vol. 9, no. 9, p. 713–725 [visit  le 2025-03-12]. ISSN 1471-0064. Disp.   l'adr. DOI: 10.1038/nrg2382.
- KARLSSON, E. K. et al., 2007. Efficient Mapping of Mendelian Traits in Dogs through Genome-Wide Association. *Nature Genetics* [en ligne]. Vol. 39, no. 11, p. 1321–1328 [visit  le 2022-11-10]. ISSN 1546-1718. Disp.   l'adr. DOI: 10.1038/ng.2007.10.
- KHANNA, C. et al., 2006. The Dog as a Cancer Model. *Nature Biotechnology*. Vol. 24, no. 9, p. 1065–1066. ISSN 1087-0156. Disp.   l'adr. DOI: 10.1038/nbt0906-1065b.

-
- KING, M. D., 2017. Etiopathogenesis of Canine Hip Dysplasia, Prevalence, and Genetics. *Veterinary Clinics of North America: Small Animal Practice* [en ligne]. Vol. 47, no. 4, p. 753–767 [visit  le 2022-11-04]. ISSN 0195-5616. Disp.   l'adr. DOI: 10.1016/j.cvsm.2017.03.001.
- KIRKNESS, E. F. et al., 2003. The Dog Genome: Survey Sequencing and Comparative Analysis. *Science* [en ligne]. Vol. 301, no. 5641, p. 1898–1903 [visit  le 2020-09-01]. ISSN 0036-8075, ISSN 1095-9203. Disp.   l'adr. DOI: 10.1126/science.1086432.
- K NIG, I. R., 2011. Validation in Genetic Association Studies. *Briefings in Bioinformatics* [en ligne]. Vol. 12, no. 3, p. 253–258 [visit  le 2025-05-23]. ISSN 1467-5463. Disp.   l'adr. DOI: 10.1093/bib/bbq074.
- KROKSTAD, S. et al., 2013. Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology* [en ligne]. Vol. 42, no. 4, p. 968–977 [visit  le 2025-08-20]. ISSN 0300-5771. Disp.   l'adr. DOI: 10.1093/ije/dys095.
- LAFRAMBOISE, T., 2009. Single Nucleotide Polymorphism Arrays: A Decade of Biological, Computational and Technological Advances. *Nucleic Acids Research* [en ligne]. Vol. 37, no. 13, p. 4181–4193 [visit  le 2025-05-21]. ISSN 0305-1048. Disp.   l'adr. DOI: 10.1093/nar/gkp552.
- LAHTINEN, M. et al., 2021. Excess Protein Enabled Dog Domestication during Severe Ice Age Winters. *Scientific Reports*. Vol. 11, no. 1, p. 7. ISSN 2045-2322. Disp.   l'adr. DOI: 10.1038/s41598-020-78214-4.
- LARSON, G. & FULLER, D., 2014. The Evolution of Animal Domestication. *Annual Review of Ecology, Evolution, and Systematics*. Vol. 66, p. 115–36. Disp.   l'adr. DOI: 10.1146/annurev-ecolsys-120213-091620.
- LARSON, G. et al., 2012. Rethinking Dog Domestication by Integrating Genetics, Archeology, and Biogeography. *Proceedings of the National Academy of Sciences* [en ligne]. Vol. 109, no. 23, p. 8878–8883 [visit  le 2025-05-01]. Disp.   l'adr. DOI: 10.1073/pnas.1203005109.
- LE N ZET, L. et al., 2025. Pedexplorer: A Bioconductor Package to Streamline Pedigree Design and Visualization. *Bioinformatics* [en ligne]. Vol. 41, no. 6, btaf329 [visit  le 2025-07-02]. ISSN 1367-4811. Disp.   l'adr. DOI: 10.1093/bioinformatics/btaf329.
- LEE, M. C. & EBERSON, C. P., 2006. Growth and Development of the Child's Hip. *Orthopedic Clinics of North America* [en ligne]. Vol. 37, no. 2, p. 119–132 [visit  le 2025-07-04]. ISSN 0030-5898. Disp.   l'adr. DOI: 10.1016/j.ocl.2005.12.001.
- LEQUARR , A.-S. et al., 2011. LUPA: A European Initiative Taking Advantage of the Canine Genome Architecture for Unravelling Complex Disorders in Both Human and Dogs. *Veterinary Journal* (London, England: 1997). Vol. 189, no. 2, p. 155–159. ISSN 1532-2971. Disp.   l'adr. DOI: 10.1016/j.tvjl.2011.06.013.
- LI, J. H. et al., 2021. Low-Pass Sequencing Increases the Power of GWAS and Decreases Measurement Error of Polygenic Risk Scores Compared to Genotyping Arrays. *Genome Research* [en ligne]. Vol. 31, no. 4, p. 529–537 [visit  le 2024-03-15]. ISSN 1088-9051. Disp.   l'adr. DOI: 10.1101/gr.266486.120.
- LINDBLAD-TOH, K. et al., 2005. Genome Sequence, Comparative Analysis and Haplotype Structure of the Domestic Dog. *Nature* [en ligne]. Vol. 438, no. 7069, p. 803–819 [visit  le 2020-09-01]. ISSN 1476-4687. Disp.   l'adr. DOI: 10.1038/nature04338.

-
- LOF SELECT, S., 2025. LOF Select: Aide à l'élevage | Société Centrale Canine [<https://www.centrale-canine.fr/lofselect>]. [visité le 2025-05-17].
- LORD, K. et al., 2013. Variation in Reproductive Traits of Members of the Genus *Canis* with Special Attention to the Domestic Dog (*Canis Familiaris*). *Behavioural Processes*. Vol. 92, p. 131–142. ISSN 1872-8308. Disp. à l'adr. DOI: 10.1016/j.beproc.2012.10.009.
- LORD, K. et al., 2016. Evolution of Working Dogs. In: SERPELL, J. (ed.). *The Domestic Dog: Its Evolution, Behavior and Interactions with People* [en ligne]. 2nd ed. Cambridge: Cambridge University Press, p. 42–66 [visité le 2025-05-04]. ISBN 978-1-107-02414-4. Disp. à l'adr. DOI: 10.1017/9781139161800.004.
- LÖSCHER, W., 2022. Dogs as a Natural Animal Model of Epilepsy. *Frontiers in Veterinary Science* [en ligne]. Vol. 9, p. 928009 [visité le 2025-05-18]. ISSN 2297-1769. Disp. à l'adr. DOI: 10.3389/fvets.2022.928009.
- MACHOU, L. et al., 2022. Faire avancer la recherche sur la génétique de la dysplasie coxo-fémorale du chien. *La semaine vétérinaire*. No. 1928, p. 14–15.
- MARSCHALL, Y. & DISTL, O., 2007. Mapping Quantitative Trait Loci for Canine Hip Dysplasia in German Shepherd Dogs. *Mammalian Genome* [en ligne]. Vol. 18, no. 12, p. 861–870 [visité le 2020-08-31]. ISSN 1432-1777. Disp. à l'adr. DOI: 10.1007/s00335-007-9071-z.
- MARSDEN, C. D. et al., 2016. Bottlenecks and Selective Sweeps during Domestication Have Increased Deleterious Genetic Variation in Dogs. *Proceedings of the National Academy of Sciences* [en ligne]. Vol. 113, no. 1, p. 152–157 [visité le 2025-05-03]. Disp. à l'adr. DOI: 10.1073/pnas.1512501113.
- MARTÍN DEL PICO, E. et al., 2024. FAIRsoft—a Practical Implementation of FAIR Principles for Research Software. *Bioinformatics* [en ligne]. Vol. 40, no. 8, btae464 [visité le 2025-05-23]. ISSN 1367-4811. Disp. à l'adr. DOI: 10.1093/bioinformatics/btae464.
- MAZZATENTA, A. et al., 2017. The Companion Dog as a Unique Translational Model for Aging. *Seminars in Cell & Developmental Biology*. Vol. 70, p. 141–153. ISSN 1096-3634. Disp. à l'adr. DOI: 10.1016/j.semcdb.2017.08.024.
- MBATCHOU, J. et al., 2021. Computationally Efficient Whole-Genome Regression for Quantitative and Binary Traits. *Nature Genetics* [en ligne]. Vol. 53, no. 7, p. 1097–1103 [visité le 2025-08-19]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-021-00870-7.
- MCEVOY, F. J. et al., 2021. Deep Transfer Learning Can Be Used for the Detection of Hip Joints in Pelvis Radiographs and the Classification of Their Hip Dysplasia Status. *Veterinary Radiology & Ultrasound* [en ligne]. Vol. 62, no. 4, p. 387–393 [visité le 2025-07-15]. ISSN 1740-8261. Disp. à l'adr. DOI: 10.1111/vru.12968.
- MCGREEVY, P. D. & NICHOLAS, F. W., 1999. Some Practical Solutions to Welfare Problems in Dog Breeding. *Animal Welfare* [en ligne]. Vol. 8, no. 4, p. 329–341 [visité le 2025-05-04]. ISSN 0962-7286, ISSN 2054-1538. Disp. à l'adr. DOI: 10.1017/S0962728600021965.
- MCLAREN, W. et al., 2016. The Ensembl Variant Effect Predictor. *Genome Biology* [en ligne]. Vol. 17, no. 1, p. 122 [visité le 2025-07-02]. ISSN 1474-760X. Disp. à l'adr. DOI: 10.1186/s13059-016-0974-4.

-
- MEADOWS, J. R. S. et al., 2023. Genome Sequencing of 2000 Canids by the Dog10K Consortium Advances the Understanding of Demography, Genome Function and Architecture. *Genome Biology* [en ligne]. Vol. 24, no. 1, p. 187 [visité le 2025-05-10]. ISSN 1474-760X. Disp. à l'adr. DOI: 10.1186/s13059-023-03023-7.
- MEESON, R. L. et al., 2019. Spontaneous Dog Osteoarthritis—a One Medicine Vision. *Nature Reviews Rheumatology*. Vol. 15, no. 5, p. 273–287.
- MELLERSH, C. S. & OSTRANDER, E. A., 1997. The Canine Genome. *Advances in Veterinary Medicine*. Vol. 40, p. 191–216. ISSN 1093-975X. Disp. à l'adr. DOI: 10.1016/s0065-3519(97)80009-2.
- MELLERSH, C. S. et al., 2000. An Integrated Linkage-Radiation Hybrid Map of the Canine Genome. *Mammalian Genome* [en ligne]. Vol. 11, no. 2, p. 120–130 [visité le 2020-09-02]. ISSN 1432-1777. Disp. à l'adr. DOI: 10.1007/s003350010024.
- MENOR-CAMPOS, D. J., 2024. Ethical Concerns about Fashionable Dog Breeding. *Animals : an Open Access Journal from MDPI* [en ligne]. Vol. 14, no. 5, p. 756 [visité le 2025-05-04]. ISSN 2076-2615. Disp. à l'adr. DOI: 10.3390/ani14050756.
- MIGNOT, E. et al., 1994. An Immunoglobulin Switchlike Sequence Is Linked with Canine Narcolepsy. *Sleep*. Vol. 17, no. 8 Suppl, S68–76. ISSN 0161-8105. Disp. à l'adr. DOI: 10.1093/sleep/17.suppl_8.s68.
- MIKKOLA, L., 2020. Dissertation: It's Complex: Studies on the Genetics of Canine Hip Dysplasia.
- MIKKOLA, L. et al., 2021. An Across-Breed Validation Study of 46 Genetic Markers in Canine Hip Dysplasia. *BMC Genomics* [en ligne]. Vol. 22, no. 1, p. 68 [visité le 2021-02-01]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/s12864-021-07375-x.
- MOBASHERI, A., 2015. Comparative Medicine in the Twenty-First Century: Where Are We Now and Where Do We Go from Here? *Frontiers in Veterinary Science* [en ligne]. Vol. 2, p. 2 [visité le 2025-05-17]. ISSN 2297-1769. Disp. à l'adr. DOI: 10.3389/fvets.2015.00002.
- MOREY, D. F., 2014. In Search of Paleolithic Dogs: A Quest with Mixed Results. *Journal of Archaeological Science* [en ligne]. Vol. 52, p. 300–307 [visité le 2025-05-01]. ISSN 0305-4403. Disp. à l'adr. DOI: 10.1016/j.jas.2014.08.015.
- MORICEAU, J.-M., 2007. Histoire du méchant loup: 3000 attaques sur l'homme en France (XVe-XXe siècle). Fayard. ISBN 978-2-213-62880-6.
- MORRILL, K. et al., 2022. Ancestry-Inclusive Dog Genomics Challenges Popular Breed Stereotypes. *Science (New York, N.Y.)* Vol. 376, no. 6592, eabk0639. ISSN 1095-9203. Disp. à l'adr. DOI: 10.1126/science.abk0639.
- MURPHY, S. C. et al., 2018. Length Variations within the Merle Retrotransposon of Canine PMEL: Correlating Genotype with Phenotype. *Mobile DNA* [en ligne]. Vol. 9, p. 26 [visité le 2025-07-25]. ISSN 1759-8753. Disp. à l'adr. DOI: 10.1186/s13100-018-0131-6.
- NEFF, M. W. et al., 1999. A Second-Generation Genetic Linkage Map of the Domestic Dog, *Canis Familiaris*. *Genetics*. Vol. 151, no. 2, p. 803–820. ISSN 0016-6731.
- NG, P. C. & HENIKOFF, S., 2001. Predicting Deleterious Amino Acid Substitutions. *Genome Research*. Vol. 11, no. 5, p. 863–874. ISSN 1088-9051. Disp. à l'adr. DOI: 10.1101/gr.176601.
- NICHOLAS, F. & TAMMEN, I., 2025. Online Mendelian Inheritance in Animals (OMIA) [en ligne]. University of Sydney [visité le 2025-07-18]. Disp. à l'adr. DOI: 10.25910/2AMR-PV70.

-
- OLLIVIER, M. et al., 2016. Amy2B Copy Number Variation Reveals Starch Diet Adaptations in Ancient European Dogs. *Royal Society Open Science*. Vol. 3, no. 11, p. 160449. ISSN 2054-5703. Disp. à l'adr. DOI: 10.1098/rsos.160449.
- OLLIVIER, M. et al., 2018. Dogs Accompanied Humans during the Neolithic Expansion into Europe. *Biology Letters*. Vol. 14, no. 10, p. 20180286. ISSN 1744-957X. Disp. à l'adr. DOI: 10.1098/rsbl.2018.0286.
- OMIA, 2025. OMIA - Online Mendelian Inheritance in Animals [<https://www.omia.org/home/>]. [visité le 2025-05-18].
- OSTRANDER, E. A. et al., 2000. Canine Genetics Comes of Age. *Trends in Genetics* [en ligne]. Vol. 16, no. 3, p. 117–124 [visité le 2022-11-04]. ISSN 0168-9525. Disp. à l'adr. DOI: 10.1016/S0168-9525(99)01958-7.
- OSTRANDER, E. A. & WAYNE, R. K., 2005. The Canine Genome. *Genome Research* [en ligne]. Vol. 15, no. 12, p. 1706–1716 [visité le 2025-05-07]. ISSN 1088-9051, ISSN 1549-5469. Disp. à l'adr. DOI: 10.1101/gr.3736605.
- OSTRANDER, E. A. et al., 2008. Canine Genetics Facilitates Understanding of Human Biology. *Genomics of Disease* [en ligne], p. 11–24 [visité le 2022-11-10]. Disp. à l'adr. DOI: 10.1007/978-0-387-76723-9_2.
- OSTRANDER, E. A. et al., 2017. Demographic History, Selection and Functional Diversity of the Canine Genome. *Nature Reviews. Genetics*. Vol. 18, no. 12, p. 705–720. ISSN 1471-0064. Disp. à l'adr. DOI: 10.1038/nrg.2017.67.
- OTT, J. et al., 2015. Genetic Linkage Analysis in the Age of Whole-Genome Sequencing. *Nature reviews. Genetics* [en ligne]. Vol. 16, no. 5, p. 275–284 [visité le 2024-12-04]. ISSN 1471-0056. Disp. à l'adr. DOI: 10.1038/nrg3908.
- OVERALL, K. L., 2000. Natural Animal Models of Human Psychiatric Conditions: Assessment of Mechanism and Validity. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* [en ligne]. Vol. 24, no. 5, p. 727–776 [visité le 2025-05-18]. ISSN 0278-5846. Disp. à l'adr. DOI: 10.1016/S0278-5846(00)00104-4.
- OVODOV, N. D. et al., 2011. A 33,000-Year-Old Incipient Dog from the Altai Mountains of Siberia: Evidence of the Earliest Domestication Disrupted by the Last Glacial Maximum. *PloS One*. Vol. 6, no. 7, e22821. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0022821.
- PANG, J.-F. et al., 2009. mtDNA Data Indicate a Single Origin for Dogs South of Yangtze River, Less than 16,300 Years Ago, from Numerous Wolves. *Molecular Biology and Evolution*. Vol. 26, no. 12, p. 2849–2864. ISSN 1537-1719. Disp. à l'adr. DOI: 10.1093/molbev/msp195.
- PARKER, H. G. et al., 2004. Genetic Structure of the Purebred Domestic Dog. *Science (New York, N.Y.)*. Vol. 304, no. 5674, p. 1160–1164. ISSN 1095-9203. Disp. à l'adr. DOI: 10.1126/science.1097406.
- PARKER, H. G. et al., 2017. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*. Vol. 19, no. 4, p. 697–708. ISSN 2211-1247. Disp. à l'adr. DOI: 10.1016/j.celrep.2017.03.079.
- PASCUAL-GARRIDO, C. et al., 2018. Canine Hip Dysplasia: A Natural Animal Model for Human Developmental Dysplasia of the Hip. *Journal of Orthopaedic Research* [en ligne]. Vol. 36, no. 7, p. 1807–1817 [visité le 2022-10-05]. ISSN 1554-527X. Disp. à l'adr. DOI: 10.1002/jor.23828.

-
- PASTOR, M. et al., 2009. Genetic and Environmental Risk Indicators in Canine Non-Hodgkin's Lymphomas: Breed Associations and Geographic Distribution of 608 Cases Diagnosed throughout France over 1 Year. *Journal of Veterinary Internal Medicine* [en ligne]. Vol. 23, no. 2, p. 301–310 [visit  le 2025-09-04]. ISSN 1939-1676. Disp.   l'adr. DOI: 10.1111/j.1939-1676.2008.0255.x.
- PATTERSON, D. F., 2000. Companion Animal Medicine in the Age of Medical Genetics. *Journal of Veterinary Internal Medicine* [en ligne]. Vol. 14, no. 1, p. 1–9 [visit  le 2022-11-10]. ISSN 1939-1676. Disp.   l'adr. DOI: 10.1111/j.1939-1676.2000.tb01492.x.
- PERRI, A., 2016. A Wolf in Dog's Clothing: Initial Dog Domestication and Pleistocene Wolf Variation. *Journal of Archaeological Science* [en ligne]. Vol. 68, p. 1–4 [visit  le 2025-07-24]. ISSN 0305-4403. Disp.   l'adr. DOI: 10.1016/j.jas.2016.02.003.
- PERRI, A. R. et al., 2021. Dog Domestication and the Dual Dispersal of People and Dogs into the Americas. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 118, no. 6, e2010083118. ISSN 1091-6490. Disp.   l'adr. DOI: 10.1073/pnas.2010083118.
- PILALIS, E. et al., 2025. Genome-Wide Functional Annotation of Variants: A Systematic Review of State-of-the-Art Tools, Techniques and Resources. *Frontiers in Pharmacology* [en ligne]. Vol. 16, p. 1474026 [visit  le 2025-05-23]. ISSN 1663-9812. Disp.   l'adr. DOI: 10.3389/fphar.2025.1474026.
- PIONNIER-CAPITAN, M. et al., 2011. New Evidence for Upper Palaeolithic Small Domestic Dogs in South-Western Europe. *Journal of Archaeological Science* [en ligne]. Vol. 38, no. 9, p. 2123–2140 [visit  le 2025-05-01]. ISSN 0305-4403. Disp.   l'adr. DOI: 10.1016/j.jas.2011.02.028.
- PLASSAIS, J. et al., 2015. A Spontaneous KRT16 Mutation in a Dog Breed: A Model for Human Focal Non-Epidermolytic Palmoplantar Keratoderma (FNEPPK). *The Journal of Investigative Dermatology*. Vol. 135, no. 4, p. 1187–1190. ISSN 1523-1747. Disp.   l'adr. DOI: 10.1038/jid.2014.526.
- PLASSAIS, J. et al., 2022. Natural and Human-Driven Selection of a Single Non-Coding Body Size Variant in Ancient and Modern Canids. *Current biology: CB*. Vol. 32, no. 4, 889–897.e9. ISSN 1879-0445. Disp.   l'adr. DOI: 10.1016/j.cub.2021.12.036.
- PURUGGANAN, M. D., 2022. What Is Domestication? *Trends in Ecology & Evolution*. Vol. 37, no. 8, p. 663–671. ISSN 1872-8383. Disp.   l'adr. DOI: 10.1016/j.tree.2022.04.006.
- QUIGNON, P. et al., 2007. Canine Population Structure: Assessment and Impact of Intra-Breed Stratification on SNP-based Association Studies. *PloS One*. Vol. 2, no. 12, e1324. ISSN 1932-6203. Disp.   l'adr. DOI: 10.1371/journal.pone.0001324.
- RANGE, F. & MARSHALL-PESCINI, S., 2022. Comparing Wolves and Dogs: Current Status and Implications for Human 'Self-Domestication'. *Trends in Cognitive Sciences* [en ligne]. Vol. 26, no. 4, p. 337–349 [visit  le 2025-05-19]. ISSN 1364-6613. Disp.   l'adr. DOI: 10.1016/j.tics.2022.01.003.
- REO, N. J. & WHYTE, K. P., 2012. Hunting and Morality as Elements of Traditional Ecological Knowledge. *Human Ecology* [en ligne]. Vol. 40, no. 1, p. 15–27 [visit  le 2025-05-03]. ISSN 1572-9915. Disp.   l'adr. DOI: 10.1007/s10745-011-9448-1.
- RIPOLLES-GARCIA, A. et al., 2023. Natural Disease History of a Canine Model of Oligogenic RPGRIP1-cone-rod Dystrophy Establishes Variable Effects of Previously and Newly Mapped Modifier Loci. *Human Molecular Genetics* [en ligne]. Vol. 32, no. 13, p. 2139–2151 [visit  le 2025-05-17]. ISSN 0964-6906. Disp.   l'adr. DOI: 10.1093/hmg/ddad046.

-
- RISER, W. H., 1975. The Dog as a Model for the Study of Hip Dysplasia. Growth, Form, and Development of the Normal and Dysplastic Hip Joint. *Veterinary Pathology*. Vol. 12, no. 4, p. 234–334. ISSN 0300-9858. Disp. à l'adr. DOI: 10.1177/030098587501200401.
- RISER, W. H. & SHIRER, J. F., 1966. Hip Dysplasia: Coxafemoral Abnormalities in Neonatal German Shepherd Dogs. *Journal of Small Animal Practice* [en ligne]. Vol. 7, no. 1, p. 7–12 [visité le 2023-12-07]. ISSN 0022-4510, ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.1966.tb04372.x.
- RUBINACCI, S. et al., 2020. Genotype Imputation Using the Positional Burrows Wheeler Transform. *PLOS Genetics* [en ligne]. Vol. 16, no. 11, e1009049 [visité le 2024-03-15]. ISSN 1553-7404. Disp. à l'adr. DOI: 10.1371/journal.pgen.1009049.
- RUBINACCI, S. et al., 2023. Imputation of Low-Coverage Sequencing Data from 150,119 UK Biobank Genomes. *Nature Genetics* [en ligne]. Vol. 55, no. 7, p. 1088–1090 [visité le 2025-09-26]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/s41588-023-01438-3.
- RUNGE, J. J. et al., 2010. Distraction Index as a Risk Factor for Osteoarthritis Associated with Hip Dysplasia in Four Large Dog Breeds*. *Journal of Small Animal Practice* [en ligne]. Vol. 51, no. 5, p. 264–269 [visité le 2022-11-04]. ISSN 1748-5827. Disp. à l'adr. DOI: 10.1111/j.1748-5827.2010.00937.x.
- SACKS, B. N. et al., 2013. Y Chromosome Analysis of Dingoes and Southeast Asian Village Dogs Suggests a Neolithic Continental Expansion from Southeast Asia Followed by Multiple Austronesian Dispersals. *Molecular Biology and Evolution* [en ligne]. Vol. 30, no. 5, p. 1103–1118 [visité le 2025-05-01]. ISSN 0737-4038. Disp. à l'adr. DOI: 10.1093/molbev/mst027.
- SAJJANAR, B. et al., 2025. Stress Responses to Changing Environmental Factors in the Domestic Animals: An Epigenetic Perspective. *Journal of Animal Physiology and Animal Nutrition* [en ligne]. Vol. n/a, no. n/a [visité le 2025-05-05]. ISSN 1439-0396. Disp. à l'adr. DOI: 10.1111/jpn.14115.
- SAVOLAINEN, P. et al., 2002. Genetic Evidence for an East Asian Origin of Domestic Dogs. *Science* (New York, N.Y.) Vol. 298, no. 5598, p. 1610–1613. ISSN 1095-9203. Disp. à l'adr. DOI: 10.1126/science.1073906.
- SCHMIDT, R. & TIMM, R., 2007. Bad Dogs: Why Do Coyotes and Other Canids Become Unruly ?
- SEABY, E. G. et al., 2016. Exome Sequencing Explained: A Practical Guide to Its Clinical Application. *Briefings in Functional Genomics* [en ligne]. Vol. 15, no. 5, p. 374–384 [visité le 2025-05-22]. ISSN 2041-2649. Disp. à l'adr. DOI: 10.1093/bfpg/elv054.
- SEGURA, V. et al., 2012. An Efficient Multi-Locus Mixed-Model Approach for Genome-Wide Association Studies in Structured Populations. *Nature Genetics* [en ligne]. Vol. 44, no. 7, p. 825–830 [visité le 2025-08-18]. ISSN 1546-1718. Disp. à l'adr. DOI: 10.1038/ng.2314.
- SERPELL, J. (ed.), 1995. The Domestic Dog: Its Evolution, Behaviour, and Interactions with People. 1. publ. Cambridge: Cambridge Univ. Press. ISBN 978-0-521-42537-7 978-0-521-41529-3.
- SERPELL, J. A., 2021. Commensalism or Cross-Species Adoption? A Critical Review of Theories of Wolf Domestication. *Frontiers in Veterinary Science* [en ligne]. Vol. 8, p. 662370 [visité le 2025-05-03]. ISSN 2297-1769. Disp. à l'adr. DOI: 10.3389/fvets.2021.662370.
- SERPELL, J. A. & PAUL, E. S., 2011. Pets in the Family: An Evolutionary Perspective. In: SHACKELFORD, T. K. & SALMON, C. (eds.). *The Oxford Handbook of Evolutionary Family Psychology* [en ligne]. Oxford University Press [visité le 2025-05-03]. ISBN 978-0-19-539669-0. Disp. à l'adr. DOI: 10.1093/oxfordhb/9780195396690.013.0017.

-
- SHANNON, L. M. et al., 2015. Genetic Structure in Village Dogs Reveals a Central Asian Domestication Origin. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 112, no. 44, p. 13639–13644. ISSN 1091-6490. Disp. à l'adr. DOI: 10.1073/pnas.1516215112.
- SHIPMAN, P., 2010. The Animal Connection and Human Evolution. *Current Anthropology*. Vol. 51, p. 519–538. Disp. à l'adr. DOI: 10.1086/653816.
- SHIPMAN, P., 2015. How Do You Kill 86 Mammoths? Taphonomic Investigations of Mammoth Megasites. *Quaternary International* [en ligne]. Vol. 359–360, p. 38–46 [visité le 2025-05-01]. ISSN 1040-6182. Disp. à l'adr. DOI: 10.1016/j.quaint.2014.04.048.
- SHIPMAN, S. A. et al., 2006. Screening for Developmental Dysplasia of the Hip: A Systematic Literature Review for the US Preventive Services Task Force. *Pediatrics* [en ligne]. Vol. 117, no. 3, e557–e576 [visité le 2025-08-19]. ISSN 0031-4005. Disp. à l'adr. DOI: 10.1542/peds.2005-1597.
- SHORT, A. D. et al., 2007. Hardy–Weinberg Expectations in Canine Breeds: Implications for Genetic Studies. *Journal of Heredity* [en ligne]. Vol. 98, no. 5, p. 445–451 [visité le 2025-05-14]. ISSN 0022-1503. Disp. à l'adr. DOI: 10.1093/jhered/esm020.
- SIMOONS, F. J. & BALDWIN, J. A., 1982. Breast-Feeding of Animals by Women: Its Socio-Cultural Context and Geographic Occurrence. *Anthropos* [en ligne]. Vol. 77, no. 3/4, p. 421–448 [visité le 2025-05-03]. ISSN 0257-9774. Disp. à l'adr. JSTOR: 40460478.
- SINNELL, J. P. et al., 2014. The Kinship2 R Package for Pedigree Data. *Human heredity* [en ligne]. Vol. 78, no. 2, p. 91–93 [visité le 2024-08-26]. ISSN 0001-5652. Disp. à l'adr. DOI: 10.1159/000363105.
- SKOGLUND, P. et al., 2015. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Current biology: CB*. Vol. 25, no. 11, p. 1515–1519. ISSN 1879-0445. Disp. à l'adr. DOI: 10.1016/j.cub.2015.04.019.
- SLOWIKOWSKI, K. et al., 2014. SNPsea: An Algorithm to Identify Cell Types, Tissues and Pathways Affected by Risk Loci. *Bioinformatics (Oxford, England)*. Vol. 30, no. 17, p. 2496–2497. ISSN 1367-4811. Disp. à l'adr. DOI: 10.1093/bioinformatics/btu326.
- STRAIN, G. et al., 2009. Prevalence of Deafness in Dogs Heterozygous or Homozygous for the Merle Allele. *Journal of Veterinary Internal Medicine* [en ligne]. Vol. 23, no. 2, p. 282–286 [visité le 2025-07-02]. ISSN 1939-1676. Disp. à l'adr. DOI: 10.1111/j.1939-1676.2008.0257.x.
- STREET, M. et al., 2014. New Analyses of the Late Palaeolithic Dog from Bonn-Oberkassel (German Rhineland). In: *Hugo Obermaier Annual Meeting*. Brunswick (Lower Saxony), Germany.
- SUTTER, N. B. et al., 2004. Extensive and Breed-Specific Linkage Disequilibrium in *Canis Familiaris*. *Genome Research* [en ligne]. Vol. 14, no. 12, p. 2388–2396 [visité le 2022-11-04]. ISSN 1088-9051, ISSN 1549-5469. Disp. à l'adr. DOI: 10.1101/gr.3147604.
- SYKES, J. E., 2024. Lessons Learned: Shaping the Evolution of Veterinary Specialty Education. *Journal of Veterinary Internal Medicine* [en ligne]. Vol. 38, no. 1, p. 375–380 [visité le 2025-05-17]. ISSN 1939-1676. Disp. à l'adr. DOI: 10.1111/jvim.16936.
- TANCREDI, D. & CARDINALI, I., 2023. Being a Dog: A Review of the Domestication Process. *Genes* [en ligne]. Vol. 14, no. 5, p. 992 [visité le 2025-05-03]. ISSN 2073-4425. Disp. à l'adr. DOI: 10.3390/genes14050992.

-
- TEAM, R. C., 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- THALMANN, O. et al., 2013. Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science* (New York, N.Y.) Vol. 342, no. 6160, p. 871–874. ISSN 1095-9203. Disp. à l'adr. DOI: 10.1126/science.1243650.
- THE PANDAS DEVELOPMENT TEAM, 2025. Pandas-Dev/Pandas: Pandas [Zenodo]. [visité le 2025-08-18]. Disp. à l'adr. DOI: 10.5281/zenodo.15831829.
- TODHUNTER, R. J. et al., 2005. Quantitative Trait Loci for Hip Dysplasia in a Cross-Breed Canine Pedigree. *Mammalian genome : official journal of the International Mammalian Genome Society*. Vol. 16, no. 9, p. 720–730. ISSN 0938-8990. Disp. à l'adr. DOI: 10.1007/s00335-005-0004-4.
- TODHUNTER, R. J. et al., 2019. Gene Expression in Hip Soft Tissues in Incipient Canine Hip Dysplasia and Osteoarthritis. *Journal of Orthopaedic Research®*. Vol. 37, no. 2, p. 313–324.
- TORRES DE LA RIVA, G. et al., 2013. Neutering Dogs: Effects on Joint Disorders and Cancers in Golden Retrievers. *PloS One*. Vol. 8, no. 2, e55937. ISSN 1932-6203. Disp. à l'adr. DOI: 10.1371/journal.pone.0055937.
- TRAGER, E. H. et al., 2007. Madeline 2.0 PDE: A New Program for Local and Web-Based Pedigree Drawing. *Bioinformatics* [en ligne]. Vol. 23, no. 14, p. 1854–1856 [visité le 2024-08-26]. ISSN 1367-4803. Disp. à l'adr. DOI: 10.1093/bioinformatics/btm242.
- TRUT, L. et al., 2009. Animal Evolution during Domestication: The Domesticated Fox as a Model. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*. Vol. 31, no. 3, p. 349–360. ISSN 1521-1878. Disp. à l'adr. DOI: 10.1002/bies.200800070.
- TRUT, L. N., 1999. Early Canid Domestication: The Farm-Fox Experiment: Foxes Bred for Tamability in a 40-Year Experiment Exhibit Remarkable Transformations That Suggest an Interplay between Behavioral Genetics and Development. *American Scientist* [en ligne]. Vol. 87, no. 2, p. 160–169 [visité le 2025-05-03]. ISSN 0003-0996. Disp. à l'adr. JSTOR: 27857815.
- TrynkaLab/CHEERS, 2024 [Trynka group at Wellcome Sanger Institute]. [visité le 2025-07-02].
- TSAI, K. L. et al., 2007. Understanding Hereditary Diseases Using the Dog and Human as Companion Model Systems. *Mammalian Genome* [en ligne]. Vol. 18, no. 6, p. 444–451 [visité le 2025-05-18]. ISSN 1432-1777. Disp. à l'adr. DOI: 10.1007/s00335-007-9037-1.
- TURNER, S. et al., 2011. Quality Control Procedures for Genome Wide Association Studies. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* [en ligne]. Vol. CHAPTER, Unit1.19 [visité le 2025-05-22]. ISSN 1934-8266. Disp. à l'adr. DOI: 10.1002/0471142905.hg0119s68.
- ucscGenomeBrowser/Kent, 2025 [UCSC Genome Browser Group]. [visité le 2025-07-02].
- VAUGHAN-SCOTT, T. & TAYLOR, J. H., 1997. The Pathophysiology and Medical Management of Canine Osteoarthritis. *Journal of the South African Veterinary Association*. Vol. 68, no. 1, p. 21–25. ISSN 1019-9128. Disp. à l'adr. DOI: 10.4102/jsava.v68i1.861.
- VAYSSE, A. et al., 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds Using Selection Mapping. *PLoS* [en ligne] [visité le 2022-11-10]. ISSN 1553-7390.

-
- VIGNE, J.-D., 2005. L'humérus de Chien Magdalénien de Erralla (Gipuzkoa, Espagne) et La Domestication Tardiglaciaire Du Loup En Europe. *Munibe*. Vol. 57, p. 279–287.
- VILÀ, C. et al., 1999. Phylogenetic Relationships, Evolution, and Genetic Diversity of the Domestic Dog. *Journal of Heredity* [en ligne]. Vol. 90, no. 1, p. 71–77 [visité le 2025-05-02]. ISSN 0022-1503. Disp. à l'adr. DOI: 10.1093/jhered/90.1.71.
- VILÀ, C. et al., 1997. Multiple and Ancient Origins of the Domestic Dog. *Science* [en ligne]. Vol. 276, no. 5319, p. 1687–1689 [visité le 2022-11-10]. Disp. à l'adr. DOI: 10.1126/science.276.5319.1687.
- VONHOLDT, B. M. et al., 2010. Genome-Wide SNP and Haplotype Analyses Reveal a Rich History Underlying Dog Domestication. *Nature*. Vol. 464, no. 7290, p. 898–902. ISSN 1476-4687. Disp. à l'adr. DOI: 10.1038/nature08837.
- WANG, C. et al., 2021a. A Novel Canine Reference Genome Resolves Genomic Architecture and Uncovers Transcript Complexity. *Communications Biology* [en ligne]. Vol. 4, no. 1, p. 1–11 [visité le 2025-05-07]. ISSN 2399-3642. Disp. à l'adr. DOI: 10.1038/s42003-021-01698-x.
- WANG, G.-D. et al., 2016. Out of Southern East Asia: The Natural History of Domestic Dogs across the World. *Cell Research* [en ligne]. Vol. 26, no. 1, p. 21–33 [visité le 2025-05-01]. ISSN 1001-0602. Disp. à l'adr. DOI: 10.1038/cr.2015.147.
- WANG, G.-d. et al., 2013. The Genomics of Selection in Dogs and the Parallel Evolution between Dogs and Humans. *Nature Communications* [en ligne]. Vol. 4, no. 1, p. 1860 [visité le 2025-05-01]. ISSN 2041-1723. Disp. à l'adr. DOI: 10.1038/ncomms2814.
- WANG, K. et al., 2010. ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Research* [en ligne]. Vol. 38, no. 16, e164 [visité le 2025-07-02]. ISSN 0305-1048. Disp. à l'adr. DOI: 10.1093/nar/gkq603.
- WANG, S. et al., 2021b. Genome-Wide Association Studies for Canine Hip Dysplasia in Single and Multiple Populations – Implications and Potential Novel Risk Loci. *BMC Genomics* [en ligne]. Vol. 22, no. 1, p. 636 [visité le 2021-09-06]. ISSN 1471-2164. Disp. à l'adr. DOI: 10.1186/s12864-021-07945-z.
- WANG, W. & KIRKNESS, E. F., 2005. Short Interspersed Elements (SINEs) Are a Major Source of Canine Genomic Diversity. *Genome Research* [en ligne]. Vol. 15, no. 12, p. 1798–1808 [visité le 2025-05-12]. ISSN 1088-9051, ISSN 1549-5469. Disp. à l'adr. DOI: 10.1101/gr.3765505.
- WEN, J. et al., 2022. Developmental Dysplasia of the Hip: A Systematic Review of Susceptibility Genes and Epigenetics. *Gene* [en ligne], p. 147067 [visité le 2022-11-28]. ISSN 0378-1119. Disp. à l'adr. DOI: 10.1016/j.gene.2022.147067.
- WERNER, P. et al., 1999. Anchoring of Canine Linkage Groups with Chromosome-Specific Markers. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*. Vol. 10, no. 8, p. 814–823. ISSN 0938-8990. Disp. à l'adr. DOI: 10.1007/s003359901096.
- WILKINS, A. S. et al., 2014. The “Domestication Syndrome” in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics* [en ligne]. Vol. 197, no. 3, p. 795–808 [visité le 2025-05-03]. ISSN 0016-6731. Disp. à l'adr. DOI: 10.1534/genetics.114.165423.
- WILKINSON, M. D. et al., 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* [en ligne]. Vol. 3, no. 1, p. 160018 [visité le 2025-05-23]. ISSN 2052-4463. Disp. à l'adr. DOI: 10.1038/sdata.2016.18.

-
- WILLEMSSEN, K. et al., 2021. Comparing Hip Dysplasia in Dogs and Humans: A Review. *Frontiers in Veterinary Science* [en ligne]. Vol. 8 [visité le 2022-10-27]. ISSN 2297-1769.
- WONG, A. K. et al., 2010. A Comprehensive Linkage Map of the Dog Genome. *Genetics* [en ligne]. Vol. 184, no. 2, p. 595–605 [visité le 2022-11-09]. ISSN 0016-6731. Disp. à l'adr. DOI: 10.1534/genetics.109.106831.
- WOOLPY, J. & GINSBURG, B. E., 1967. Wolf Socialization: A Study of Temperament in a Wild Social Species. *American Zoologist* [en ligne]. Vol. 7, no. 2, p. 357–363 [visité le 2025-05-03]. ISSN 0003-1569. Disp. à l'adr. DOI: 10.1093/icb/7.2.357.
- WORBOYS, M. et al., 2018. *The Invention of the Modern Dog: Breed and Blood in Victorian Britain* [en ligne]. Johns Hopkins University Press [visité le 2025-05-04]. ISBN 978-1-4214-2658-7 978-1-4214-2659-4.
- WRAGG, D. et al., 2024. A Cautionary Tale of Low-Pass Sequencing and Imputation with Respect to Haplotype Accuracy. *Genetics, Selection, Evolution : GSE* [en ligne]. Vol. 56, p. 6 [visité le 2024-01-17]. ISSN 0999-193X. Disp. à l'adr. DOI: 10.1186/s12711-024-00875-w.
- WU, J. et al., 2016. Whole Genome Sequencing and Its Applications in Medical Genetics. *Quantitative Biology* [en ligne]. Vol. 4, no. 2, p. 115–128 [visité le 2025-05-21]. ISSN 2095-4697. Disp. à l'adr. DOI: 10.1007/s40484-016-0067-0.
- WUCHER, V. et al., 2017. FEELnc: A Tool for Long Non-Coding RNA Annotation and Its Application to the Dog Transcriptome. *Nucleic Acids Research*. Vol. 45, no. 8, e57. ISSN 1362-4962. Disp. à l'adr. DOI: 10.1093/nar/gkw1306.
- WURSTER-HILL, D. H. & CENTERWALL, W. R., 1982. The Interrelationships of Chromosome Banding Patterns in Canids, Mustelids, Hyena, and Felids. *Cytogenetics and Cell Genetics*. Vol. 34, no. 1-2, p. 178–192. ISSN 0301-0171. Disp. à l'adr. DOI: 10.1159/000131806.
- YAN, W. et al., 2019. A Genome-Wide Association Study Identifies New Genes Associated with Developmental Dysplasia of the Hip. *Clinical Genetics* [en ligne]. Vol. 95, no. 3, p. 345–355 [visité le 2024-02-01]. ISSN 1399-0004. Disp. à l'adr. DOI: 10.1111/cge.13483.
- YANG, J. et al., 2011. GCTA: A Tool for Genome-Wide Complex Trait Analysis. *American Journal of Human Genetics*. Vol. 88, no. 1, p. 76–82. ISSN 1537-6605. Disp. à l'adr. DOI: 10.1016/j.ajhg.2010.11.011.
- ZENG, F. et al., 2022. PNPLA1-Mediated Acylceramide Biosynthesis and Autosomal Recessive Congenital Ichthyosis. *Metabolites* [en ligne]. Vol. 12, no. 8, p. 685 [visité le 2025-05-18]. ISSN 2218-1989. Disp. à l'adr. DOI: 10.3390/metabo12080685.
- ZHANG, Z. et al., 2020. Deciphering the Puzzles of Dog Domestication. *Zoological Research* [en ligne]. Vol. 41, no. 2, p. 97–104 [visité le 2025-05-03]. ISSN 2095-8137. Disp. à l'adr. DOI: 10.24272/j.issn.2095-8137.2020.002.
- ZHAO, H. et al., 2018. A Practical Approach to Adjusting for Population Stratification in Genome-Wide Association Studies: Principal Components and Propensity Scores (PCAPS). *Statistical Applications in Genetics and Molecular Biology*. Vol. 17, no. 6, /j/sagmb.2018.17.issue-6/sagmb-2017-0054/sagmb-2017-0054.xml. ISSN 1544-6115. Disp. à l'adr. DOI: 10.1515/sagmb-2017-0054.
- ZHOU, X. et al., 2013. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLOS Genetics* [en ligne]. Vol. 9, no. 2, e1003264 [visité le 2025-07-01]. ISSN 1553-7404. Disp. à l'adr. DOI: 10.1371/journal.pgen.1003264.

ZHU, L. et al., 2012. Identification of Quantitative Trait Loci for Canine Hip Dysplasia by Two Sequential Multipoint Linkage Analyses. *Journal of Applied Statistics* [en ligne]. Vol. 39, no. 8, p. 1719–1731 [visit  le 2022-11-10]. ISSN 0266-4763. Disp.   l'adr. DOI: 10.1080/02664763.2012.673121.



Titre : Recherche des bases génétiques de la Dysplasie Coxo-Fémorale chez le chien guide comme modèle de la Dysplasie Développementale de la Hanche chez l'humain

Mot clés : Dysplasie Coxo-Fémorale, Chien, Génétique, Une seule santé, Science Ouverte

Résumé : Au cours des dernières décennies, le chien (*Canis lupus familiaris*) est devenu un modèle de choix pour l'étude des maladies génétiques complexes, grâce à sa diversité phénotypique, fruit d'une domestication ancienne et d'une intense sélection conduisant à la formation de plus de 400 races modernes, véritables isolats génétiques.

Cette thèse porte sur la dysplasie coxo-fémorale (DCF), une affection multifactorielle débilatante du développement articulaire et homologue à la dysplasie développementale de la hanche (DDH) chez l'humain. Plus de 700 chiens ont été séquencés à faible couverture puis imputés afin d'identifier, par analyse d'association pangénomique (GWAS), des loci associés à la DCF.

Deux outils bio-informatiques ont été développés : le package *R Pedexplorer* pour le traitement des pedigrees complexes et le pipeline *nf-core/phaseimpute* consacré à l'imputation de génotypes.

La combinaison de GWAS intra- et inter-races avec des annotations fonctionnelles et une comparaison à la littérature sur la DDH humaine nous a permis de mettre en évidence une vingtaine de loci candidats.

Ce travail, mené en partenariat avec l'ACGAO et la Fondation VISIO, ouvre la voie à un futur test génétique de risque dont le développement se poursuivra dans une thèse ultérieure.

Title: Genetics basics of hip dysplasia in dogs as a model for the developmental dysplasia of the hip in humans

Keywords: Hip Dysplasia, Dog, Genetic, One-Health, Open Science

Abstract: Over the past few decades, dogs (*Canis lupus familiaris*) have become a model of choice for the study of complex genetic diseases, thanks to their phenotypic diversity, the result of ancient domestication and intense artificial selection leading to the formation of more than 400 modern breeds, which are true genetic isolates.

This thesis focuses on canine hip dysplasia (CHD), a debilitating multifactorial joint development disorder homologous to developmental dysplasia of the hip (DDH) in humans. More than 700 dogs were sequenced at low

coverage and then imputed in order to identify loci associated with CHD through genome-wide association studies (GWAS).

Two bioinformatics tools were developed: the *R Pedexplorer* library for exploring complex pedigrees and the *nf-core/phaseimpute* pipeline for genotype imputation.

The combination of intra- and inter-racial GWAS, functional annotations and comparison with the literature on human DDH enabled us to identify about twenty candidate loci.

This work, in partnership with the ACGAO and VISIO Foundation, paves the way for a future genetic risk test, whose development will continue in an upcoming PhD project.