



HAL
open science

Evolutions de la recherche en génétique animale et anticipation des besoins de formation

Claire Rogel Gaillard

► **To cite this version:**

Claire Rogel Gaillard. Evolutions de la recherche en génétique animale et anticipation des besoins de formation. *Ethnozootechnie*, 2021, 109, pp.11-22. hal-03545671

HAL Id: hal-03545671

<https://hal.inrae.fr/hal-03545671>

Submitted on 27 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evolutions de la recherche en génétique animale et anticipation des besoins de formation

Claire ROGEL-GAILLARD

Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas

Contact : claire.rogel-gaillard@inrae.fr

Résumé : L'histoire de la génétique animale est ancrée dans celle des hommes et des femmes qui, depuis le début de l'agriculture, ont œuvré pour sélectionner des animaux adaptés à leurs besoins. Le développement de la génétique quantitative avec des méthodes statistiques de plus en plus performantes, le déploiement de la génétique moléculaire et des biotechnologies du génome depuis la découverte de l'ADN, l'essor de la génomique depuis le séquençage des génomes, ont progressivement dessiné des champs de recherche variés et complémentaires. La recherche en génétique animale est ainsi nourrie par l'accumulation de nouvelles connaissances et par le renouvellement des approches méthodologiques. Dans le contexte actuel de transition agroécologique et sociétale, les objectifs de sélection se diversifient, conduisant à élargir le panorama des caractères étudiés, à explorer les interactions avec des approches systémiques, à tenir compte de nouvelles échelles d'analyse, à favoriser la diversité génétique, à caractériser finement les liens en génotypes et phénotypes pour améliorer la précision des prédictions. Il sera important d'acquérir une vision intégrative à la fois des questions posées et des approches à déployer pour y répondre. Il sera attendu de favoriser des formations qui couplent science des données et biologie, avec un éclairage apporté par les enjeux sociétaux et environnementaux.

Mots-clés : *Génétique animale, transitions, interactions, formation.*

Introduction

Depuis que les humains ont commencé à élever des animaux, un défi majeur consiste à contrôler l'hérédité des caractères considérés comme d'intérêt. Dans le cas des animaux d'élevage, le gain génétique a été obtenu à l'aide de données généalogiques et de modèles statistiques. Depuis l'introduction de la sélection génomique (Meuwissen *et al.*, 2001), l'élevage est passé à des approches qui exploitent l'information génétique pour prédire la transmission des phénotypes, avec l'ambition aussi d'identifier les variabilités génomiques impliquées dans la variance des caractères mesurés. Les recherches en génétique animale alimentent ce besoin prépondérant mais ont un périmètre plus large, avec des objectifs complémentaires qui s'inscrivent dans l'histoire de l'humanité et dans le temps. L'étude des génomes permet d'explorer le passé avec, par exemple, l'analyse de la fréquence et de la ségrégation des allèles au fil des générations, l'identification de nouveaux polymorphismes. Ces recherches retracent la dynamique évolutive des génomes et apportent des informations pour améliorer notre compréhension des mécanismes biologiques mobilisés au cours de la domestication, plus généralement en réponse à une adaptation dans des environnements variés, ayant abouti dans certains cas à la définition de races. Les génomes révèlent ainsi des traces de sélection (Rubin *et al.*, 2010). Dans le temps présent, la caractérisation à la fois génétique et phénotypique des animaux en

élevage fournit les informations indispensables au choix des futurs reproducteurs (Elsen, 2020) et enrichit nos connaissances sur les associations entre génotypes et phénotypes. En lien avec le futur, les recherches doivent aussi contribuer à identifier, préserver et générer les ressources génétiques, dont la diversité constitue un carburant d'adaptation pour l'avenir. Ces questions ne sont pas nouvelles en génétique animale et sont partagées avec les généticiens des autres domaines du vivant, notamment celui des plantes.

La Figure 1 retrace de manière chronologique le développement des outils et des méthodes d'amélioration génétique des animaux à partir du XXe siècle, avec trois volets connectés, un premier sur l'étude des génomes et la génomique, un second sur les biotechnologies de la reproduction, et un troisième sur la génétique quantitative et la sélection animale. Depuis les années 2000, l'essor de la génomique a été un élément déterminant dans l'évolution des recherches, avec la mise en œuvre d'une capacité de génotypage sans précédent dans l'histoire de la génétique animale et des développements méthodologiques qui ont permis le déploiement de la sélection génomique (Verrier, 2020). Dans un contexte de transition agroécologique et sociétale, les objectifs de sélection se diversifient (Le Roy *et al.*, 2019 ; Phocas, 2020) et c'est un nouveau défi que relèvent les généticiens.

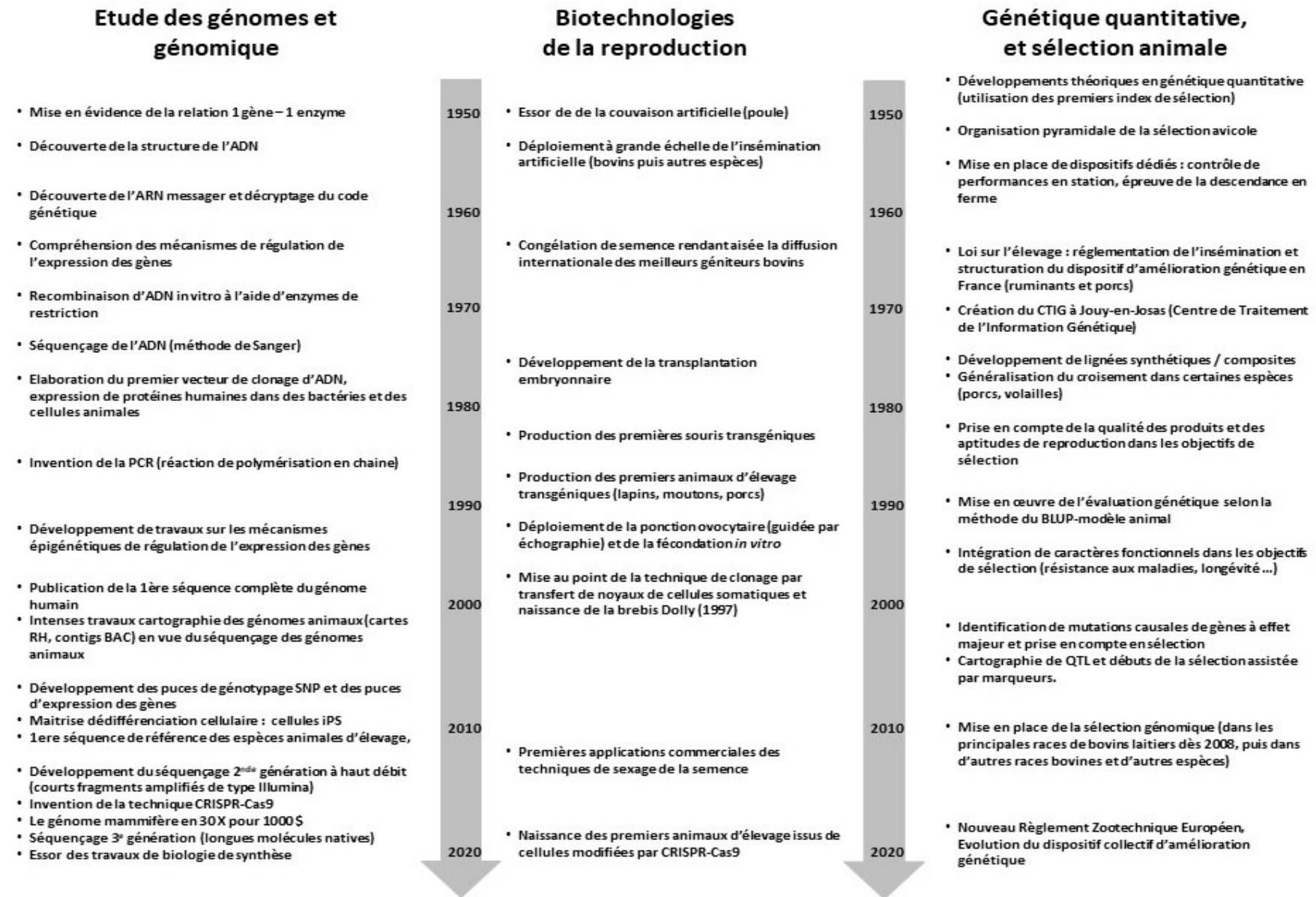


Figure 1. Chronologie du développement des outils et des méthodes de l'amélioration génétique des animaux d'élevage depuis 1950. D'après Ducos *et al.* (2020), in Verrier, Milan et Rogel-Gaillard (coords), avec l'aimable autorisation des éditions Quae.

C'est dans ce cadre d'une forte avancée des connaissances sur les génomes et leur expression, d'innovations méthodologiques et d'un champ de contraintes sociétal et environnemental nouveau qu'ont évolué et continuent d'évoluer les recherches

en génétique animale. Nous allons présenter quelques-unes de ces évolutions, sans volonté d'exhaustivité et sans reprendre toute leur histoire, afin de dégager quelques fronts de sciences et de proposer des pistes quant aux besoins de formation associés.

Diversification des objectifs de sélection dans un contexte de transition agroécologique et sociétale

La définition des objectifs de sélection est incontournable pour tout programme d'amélioration génétique. Elle a évolué et continue d'évoluer pour répondre aux besoins (Figure 2). L'impact des pratiques sur le changement climatique, sur les ressources (sol, eau, biodiversité), comme sur la santé avec ses liens entre humains, animaux, environnement, conduit depuis une dizaine d'années à une meilleure prise en compte de la durabilité, de l'empreinte environnementale et du bien-être animal. Cette transition agroécologique, au cœur de la feuille de route INRAE 2030 ([connaitre/inrae2030\), implique une diversification des objectifs de sélection : adaptation et résilience face à des fluctuations environnementales, compétence immunitaire et résistance aux maladies en limitant l'usage des antibiotiques, réduction des émissions de méthane, digestion d'aliments plus riches en fibres, bien-être animal, longévité, etc. Elle nécessite des recherches soutenues pour identifier et réaliser les mesures associées et définir les critères de sélection qui répondent aux objectifs revisités qui visent une multi-performance économique et environnementale.](https://www.inrae.fr/nous-</p>
</div>
<div data-bbox=)

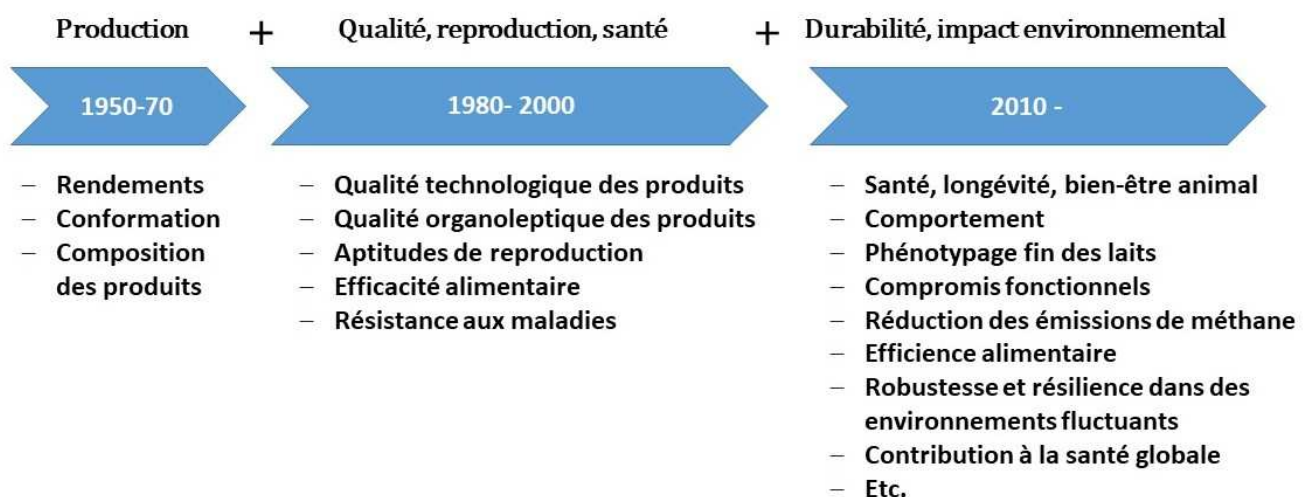


Figure 2. Evolution des objectifs de sélection en élevage pour la multi-performance, en lien avec la transition agroécologique et sociétale.

La prise en compte de la diversité génétique dans les programmes de sélection est mise en avant comme un des leviers pour favoriser la transition agroécologique (Phocas *et al.*, 2015 ; Caquet *et al.*, 2020). Optimiser la diversité des ressources génétiques sous contrainte d'amélioration génétique nécessite de faire la part entre les régions du génome sous forte pression de sélection, les régions du génome considérées comme neutres pour les caractères sous sélection (maintenir de la diversité) et les régions du génome dont il est connu que le polymorphisme est un gage d'adaptation notamment à des pathogènes variés à l'échelle d'une population (par exemple : région génomique du complexe majeur d'histocompatibilité, qui contient des gènes très polymorphes responsables de la

reconnaissance du soi et du non soi et de la présentation des antigènes aux cellules du système immunitaire).

Les objectifs de sélection doivent répondre aux besoins identifiés sur toute la chaîne de valeurs, depuis par l'amont jusqu'à l'aval, « de la ferme à la table » (*From farm to fork*) pour faire référence à l'un des huit axes stratégiques du Pacte Vert européen (*Green deal*), nouvelle feuille de route dont l'objectif est de rendre l'économie de l'Union européenne durable (lien : https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_fr).

La transition agroécologique est couplée avec une transition sociétale qui interroge sur les régimes alimentaires à base de viande, le droit à l'usage des animaux, la place de l'élevage dans nos sociétés. Elle invite à réfléchir aux limites en sélection (cas de la race bovine Blanc Bleu Belge dont l'hypertrophie musculaire qui entraîne des césariennes systématiques lors des naissances), et de manière plus générale au nécessaire couplage entre éthique des fins et des moyens pour l'amélioration génétique (Heams, 2020).

Ce contexte de double transition est stimulant pour les généticiens car il impose une approche plus holistique

L'apport des données liées à l'essor de la génomique

La génétique est un champ disciplinaire basé sur les lois de l'hérédité et la connaissance des génomes. C'est une science de la prédiction qui permet d'estimer, pour un caractère ou un ensemble de caractères héréditaires, leurs valeurs dans une descendance, sur la base des informations des parents. Avec l'essor de la génomique, la prédiction s'appuie sur l'exploitation du polymorphisme des génomes, et des outils de génotypage d'une puissance remarquable ont pu être élaborés grâce au séquençage des génomes. Au-delà du repérage des variabilités interindividuelles basées sur l'étude des polymorphismes, l'analyse des génomes et de leur expression fournit des informations fonctionnelles qui permettent de préciser les liens entre génotypes et phénotypes et d'accélérer l'identification de mutations causales de caractères monogéniques

des objectifs de sélection et une intégration des attentes sociales et environnementale pour leur définition. Il conduit notamment à une interdisciplinarité de deux ordres : d'une part avec des scientifiques de disciplines plus ou moins proches (physiologie, médecine vétérinaire, microbiologie, écologie, technologies du numérique avec le développement de l'élevage connecté et des capteurs) et, d'autre part, avec des acteurs en sciences humaines et sociales pour avancer dans la prise en compte d'objectifs multiples à valeurs marchandes comme non marchandes (Phocas, 2020).

avec quelques exemples : chez le porc pour le gène « RN » (rendement napole) dont une mutation dominante est responsable d'une teneur élevée en glycogène dans le muscle squelettique (Milan *et al.*, 2000), chez les bovins pour des anomalies pathologiques (Bourneuf *et al.*, 2017), chez la poule pour des caractères morphologiques visibles comme la couleur chocolat du plumage (Li *et al.*, 2019) et la crête en forme de rose (Imsland *et al.*, 2012).

Cette connexion entre structure du génome et annotation fonctionnelle définit un champ de recherche complexe qui resserre les liens entre généticiens quantitatifs et moléculaires, avec l'émergence de la biologie computationnelle dans ce domaine de recherche.

Des génomes de référence aux pan-génomes

L'assemblage et le séquençage des génomes ont constitué des étapes fondatrices pour l'évolution récente des recherches en génétique, avec par exemple le séquençage du génome de la poule en 2004, du porc en 2009, de la vache en 2012, du lapin en 2014. Des études de diversité génétique ont en parallèle conduit à l'identification de polymorphismes nucléotidiques (SNP pour *single nucleotide polymorphism* ou SNV pour *single nucleotide variant*) répartis sur l'ensemble du génome et des puces de génotypage à basse (environ 10 000 SNP), moyenne (environ 50 000 SNP) et haute densité (> 500 000 SNP) ont été conçues avec des partenariats entre les communautés scientifiques, des sélectionneurs et des entreprises, selon des modalités variables par espèce.

L'accès à des infrastructures de séquençage de plus en plus nombreuses et performantes a permis la production de nouvelles données issues du séquençage de génomes d'individus différents d'une même espèce. Le terme de reséquençage de génome et souvent employé dans ce cas, pour faire la distinction

avec le séquençage initial du génome de référence pour l'espèce concernée. Le projet 1000 génomes bovins illustre de telles initiatives (Hayes et Daetwyler, 2018) et un projet 1000 génomes est engagé chez la poule (Tixier-Boichard *et al.*, 2021). Ces efforts de reséquençage mettent en évidence, d'une part la diversité des polymorphismes potentiellement exploitables, au-delà des SNP et, d'autre part, les insuffisances d'un unique génome de référence pour bien représenter l'ensemble du génome d'une espèce donnée.

Le reséquençage à grande échelle des génomes montre la présence d'une variabilité notable liée à des insertions et délétions de séquences d'ADN, avec des variations du nombre de copies de certains motifs d'ADN (CNV pour *copy number variant*). Les microsatellites largement utilisés avant l'exploitation massive des SNP, entrent dans cette catégorie des CNV mais d'autres motifs d'ADN sont en nombre variable de copies selon les génomes, sous forme de répétitions en tandems (ex : microsatellites) ou

dispersée dans les génomes (ex : séquence endovirales). Ces informations ne sont pas toutes répertoriées parmi celles qui sont associées à un génome de référence. Or, issues du reséquençage des génomes, elles sont à connecter à celles du génome de référence pour enrichir et rendre plus exhaustives les données structurales des génomes de chaque espèce. Un point complémentaire peut aussi être souligné ici : la séquence du génome de référence ne reflète pas l'assemblage par haplotype des chromosomes parentaux. Grâce aux nouvelles technologies de séquençage, il est possible de lire la séquence nucléotidique directement à partir de molécules uniques sur des longueurs de l'ordre de 104 bases ou plus, les taux d'erreur (5 à 15%) étant compensés partiellement par l'établissement de consensus partir de plusieurs séquences indépendantes. Il devient donc envisageable de reconstituer la séquence des deux haplotypes d'un même génome sur de grands segments.

A la notion de génome de référence succède maintenant celle de pangénome qui correspond à toute

collection de séquences génomiques à analyser conjointement ou à utiliser comme référence pour une espèce donnée (Computational Pan-Genomics Consortium, 2018). On passe d'une représentation linéaire du génome de référence à une représentation graphique du pan-génome. Chez les bovins, la séquence du génome de référence, assemblée à partir d'une vache Hereford (ARS-UCD1.2), a été récemment augmentée avec des variants filtrés selon la fréquence des allèles dans des races laitières (Brown Swiss, Holstein) et mixtes (Fleckvieh, Original Braun) pour construire des graphes de référence spécifiques aux échelles de chaque race ou du pan-génome bovin (Crysnanto et Pausch, 2020). Un consortium pour élaborer un pan-génome pour les bovins a été initié fin 2019 et il est attendu que de tels efforts se déploient pour d'autres espèces d'élevage.

Ces approches dédiées aux pan-génomes requièrent une expertise bioinformatique nouvelle identifiée comme un sous-domaine de recherche en biologie computationnelle (Computational Pan-Genomics Consortium, 2018).

L'annotation fonctionnelle des génomes

Un objectif important en génétique est d'établir des liens entre génome et fonction, afin d'identifier les informations portées par le génome qui sont impliquées dans les variations des caractères mesurés et identifier des mutations causales permettant de comprendre les mécanismes biologiques en jeu. L'effort conséquent de séquençage et assemblage des génomes est couplé à un besoin accru d'annotation fonctionnelle, complémentaire de celui de répertorier les variations structurales et polymorphismes.

Il s'agit ici de donner un sens biologique aux données de séquences, en s'appuyant sur la démultiplication des méthodologies d'analyse des génomes, fondées d'une part sur l'acquisition de données expérimentales variées qui renseignent sur l'expression des génomes (ex : RNA-Seq), l'accessibilité et les sites d'interaction de la chromatine (ex : ATAC-Seq), la méthylation de l'ADN et les modifications de la chromatine, etc. et, d'autre part, sur des prédictions *in silico*. L'enjeu est ici de caractériser les gènes qui codent des protéines mais aussi tous les autres éléments fonctionnels : transcrits qui ne codent pas de protéines (ex : ARN longs non codants, micro-ARN), éléments régulateurs exerçant leurs fonctions à proximité des gènes ou à distance (ex : promoteurs, *enhancers/silencers*).

Ces approches sont complexes car l'annotation fonctionnelle requiert de générer des données

d'expression pour une gamme de tissus et de conditions physiologiques la plus large possible. Elle repose donc souvent sur l'étude de peu d'individus, contrairement aux études de diversité génétique, avec comme enjeu de connecter ces annotations aux polymorphismes d'intérêt.

Afin de progresser spécifiquement sur l'annotation des génomes des animaux d'élevage, l'initiative internationale FAANG (*Functional Annotation of Animal Genomes*) a démarré en 2014 et rassemble maintenant une large communauté scientifique avec plusieurs dizaines de structures de recherche impliquées sur tous les continents (<https://www.faang.org/>). Cette initiative a mis en avant l'importance du partage des données, des méthodes et des ressources biologiques (Giuffra *et al.*, 2019). Trois projets européens ont démarré en 2019 pour quatre ans, dédiés aux espèces bovine (projet H2020 BovReg), aquacoles (projet H2020 AQUA-FAANG), ainsi qu'au porc et à la poule (projet H2020 GENE-SWITCH).

Les généticiens progressent dans l'annotation des génomes, avec un couplage nécessaire entre expérimentation, données multi-omiques, et prédictions *in silico*.

Le déploiement des approches multi-omiques

Les évolutions technologiques ont permis le développement d'approches à haut débit pour caractériser l'expression des génomes dans leur globalité. Or, autant l'ADN d'un individu est fixé à sa naissance (en faisant abstraction des mutations somatiques au cours de la vie), autant le programme d'expression des génomes varie selon les tissus et les conditions environnementales. Les données omiques donnent accès aux transcrits (transcriptome), aux protéines (protéome), aux métabolites (métabolome). Ces données multi-omiques sont à vocations multiples, avec quatre grands types de résultats associés : i) des informations sur la co-expression des gènes qui mettent en évidence des voies biologiques mobilisées dans les conditions étudiées ; ii) des analyses différentielles qui illustrent les contrastes entre deux groupes d'échantillons ; iii) des

biomarqueurs candidats ; iv) des informations pour enrichir l'annotation fonctionnelle des génomes.

Il existe ici un fort enjeu d'intégration des données car les d'informations sont de natures distinctes et le plus souvent dissociées des informations de polymorphismes génétiques. Des approches se développent pour collecter sur de grandes cohortes les informations de transcriptome couplées à celles du polymorphisme des SNP, permettant des études d'association qui utilisent les niveaux d'expression des gènes comme des phénotypes. Ce sont des approches dites « eQTL » (QTL d'expression) qui aident à la priorisation de mutations causales candidates (van den Berg et al., 2019). Ces données omiques fournissent des phénotypes qualifiés d'intermédiaires, qui viennent compléter les phénotypes issus d'autres types de mesures (Figure 3).

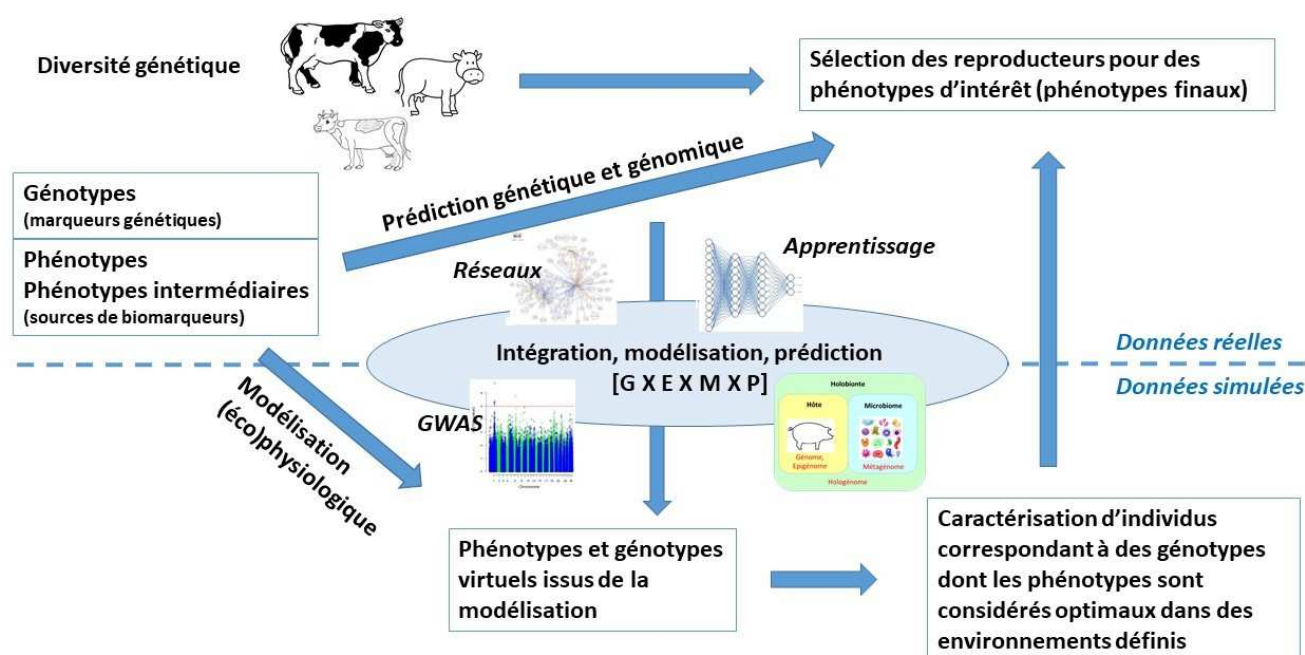


Figure 3. Représentation des liens entre données réelles et données simulées pour la sélection génomique intégrative. Les compétences sur l'intégration des données, la modélisation et la prédiction sont centrales et s'appuient sur les théories et méthodes de la génétique quantitative avec les apports attendus de l'annotation fonctionnelle des SNP et de leurs causalités, de la biologie de systèmes (réseaux) et des opportunités à explorer offertes par l'apprentissage profond et l'intelligence artificielle. Il est attendu des avancées méthodologiques pour tenir compte conjointement des informations de génotypes (*G*), environnement (*E*), microbiome (*M*), pratiques (*P*), et de leurs interactions. GWAS : *genome wide association studies* (études d'association).

Ces données multi-omiques ont contribué à élargir la définition d'un gène, au-delà de séquences d'ADN qui codent des protéines. L'analyse des transcriptomes a mis en évidence l'importance des transcrits non codants petits (ex : microARN) et longs, et l'existence de gènes qui ne codent pas de protéines. Les mutations causales candidates sont ainsi réparties dans des gènes

codant des protéines mais également dans des gènes non codants, ce qui complexifie leur identification et explique aussi pourquoi nombre d'entre elles sont difficiles à identifier à proximité de SNP en déséquilibre de liaison.

L'apport de ces données multi-omiques requiert, outre les bases en biostatistique pour les exploiter, d'avoir des connaissances sur la structure et l'organisation des

génomomes et des compétences pour consulter les bases de données dédiées.

Des recherches à la croisée de plusieurs fronts de sciences

Interactions entre génétique, épigénétique et environnement

L'épigénétique au sens large recouvre l'ensemble des phénomènes susceptibles de modifier l'expression d'un génome sans en altérer sa séquence. L'épigénome d'une cellule est l'ensemble complet des marques épigénétiques, telles que la méthylation de l'ADN, les modifications post-traductionnelles des histones (méthylation ou acétylation), le remodelage de la chromatine, les ARN non codants et autres molécules qui peuvent transmettre des informations à travers la mitose en régulant l'expression génique. Les marques épigénétiques sont à la fois stables et héréditaires au cours des divisions cellulaires : elles se transmettent de cellules mères à cellules filles, voire d'une génération à une autre. La variabilité génétique n'explique qu'une partie de la variance des phénotypes et les recherches récentes visent à estimer la part de cette variance portée par les modifications épigénétiques. Ces recherches sont d'une grande complexité car l'épigénome varie selon les cellules et organes (spécificité d'expression des gènes en lien avec la spécialisation fonctionnelle des organes) et aussi selon les variations d'environnement. De plus, les marques épigénétiques apposées sur le génome sont modifiables et/ou réversibles selon l'environnement, et ces modifications peuvent aussi avoir des conséquences à long terme (Beaujean *et al.*, 2020).

Les recherches en génétique sur le sujet visent, d'une part à cartographier et annoter l'ensemble des transcrits non codants et marques épigénétiques et, d'autre part, à étudier la fraction héréditable des marques épigénétiques transmises à la descendance, avec des travaux notables sur la génétique de l'épigénétique chez les oiseaux (Guerreo-Bosagna *et al.*, 2018). Chez la truite, il a été montré récemment que des variations précoces de température de l'eau modifient les profils de méthylation de l'ADN et que ces changements, observés dans des lignées isogéniques de truites, dépendent du fond génétique (Lallias *et al.*, 2020). Des travaux sont en cours chez les bovins également et il est anticipé qu'une étape importante vers l'élevage de précision sera franchie le jour où des technologies normées, permettant de quantifier les variations pour un sous-ensemble de marques épigénétiques, à moindre coût et en routine, pourront être déployées à grande échelle chez les animaux d'élevage (revue par Beaujean *et al.*, 2020).

Aborder ce front de sciences requiert une excellente connaissance de la structure des génomes et des méthodologies pour analyser et cartographier les marques épigénétiques. Il s'appuie également sur des approches expérimentales et des connaissances fines en physiologie animale pour aborder les mécanismes biologiques sous-jacents.

Emergence du rôle des microbiotes et de leurs microbiomes

Les microbiotes sont définis comme les écosystèmes microbiens ou ensemble des communautés microbiennes qui vivent en interactions étroites et permanentes avec leurs organismes hôtes, leur assurant des fonctions vitales. La définition du terme microbiome a été récemment précisée comme étant le « théâtre d'activité » d'un microbiote, incluant l'ensemble des molécules (ADN, ARN, protéines, métabolites, etc.) produites et présentes dans l'environnement du microbiote (Berg *et al.*, 2020). Les travaux chez les animaux d'élevage ciblent actuellement la composition en ADN des microbiotes, afin d'identifier les communautés microbiennes présentes (qui est là) et leur abondance relative (proportion dans l'écosystème global). La plupart des travaux chez les animaux d'élevage s'appuient sur la caractérisation du microbiote par séquençage de la

région variable du gène de l'ARNr 16S. Des approches plus exhaustives mais aussi plus coûteuses visent le séquençage complet de l'ADN du microbiote, avec alignement des séquences sur un catalogue de gènes établi au préalable. Ainsi, aux côtés des génomes des hôtes se construisent progressivement des catalogues de gènes du microbiote intestinal, permettant d'étudier conjointement les génomes de l'hôte et de son microbiote. Les premiers catalogues ont été publiés pour le porc (Xiao *et al.*, 2016), la poule (Huang *et al.*, 2018), la vache (Li *et al.*, 2020). Il est intéressant de noter qu'il y a environ une dizaine d'années entre la publication des génomes des hôtes et de leurs microbiotes : 14 ans pour la poule (génome publié en 2004), 8 ans pour la vache (génome publié en 2012), 7 ans pour le porc (génome publié en 2009).

Les microbiotes les plus étudiés chez les animaux d'élevage sont ceux du tube digestif (Calenge *et al.*, 2014) qui ont des rôles multiples : fonctions trophiques et métaboliques (ex : dégradation de la cellulose, production de vitamines, d'acides gras à chaîne courte, de nutriments), fonction intestinale (ex : régénération et limitation de la perméabilité intestinale). A ces fonctions variées s'ajoute la mise en évidence d'un axe microbiote-intestin-cerveau associé à des phénomènes d'anxiété, de perception de la douleur.

Les recherches en génétique animale dans ce domaine en pleine effervescence sont de quatre ordres : 1) étudier la covariation de la composition du microbiote intestinal et de phénotypes d'intérêt (efficacité alimentaire, efficacité vaccinale, émission de méthane, croissance, etc.) ; 2) analyser le contrôle génétique de la variabilité de composition du microbiote intestinal (paramètres génétiques, études d'association) ; 3)

estimer la part de variance des phénotypes due aux variations du microbiote, avec la proposition du mot microbiabilité pour cette estimation (Camarinha-Silva *et al.*, 2017).

Le lien entre l'hôte et son microbiome est à l'origine de la notion d'holobionte et d'hologéome, une nouvelle échelle biologique qui formalise cette association fonctionnelle (Theis *et al.*, 2016). Il va devenir important d'élargir la notion d'individu à celle de l'holobionte et de populations d'holobiontes qui interagissent entre eux et avec leur environnement (Vignal *et al.*, 2020).

Ce front de science, s'appuie sur des compétences fortes en bioinformatique et biologie computationnelle et requiert de l'interdisciplinarité avec des biostatisticiens et des microbiologistes.

Trier les informations pour identifier les mutations causales

La disponibilité en grand nombre de génomes séquencés (cf. « projets 1000 génomes ») et l'accès à du génotypage dense a démultiplié la capacité de conduire efficacement des études d'associations (identification des SNP en déséquilibre de liaison avec un phénotype) et d'identifier des variants causaux avec cadence soutenue en particulier chez les bovins (Bourneuf *et al.*, 2017). Cela a permis aussi le développement d'approches fondées sur le repérage des régions génomiques avec perte d'homozygotie,

qui peuvent révéler des associations avec des pathologies ou anomalies congénitales limitant le développement embryonnaire ou la survie à la naissance (Michot *et al.* 2017).

L'exploration des données de séquence *in silico* en lien avec leur annotation va enrichir la caractérisation fonctionnelle des SNP et aider à leur tri et priorisation pour l'identification de variants causaux (Xiang *et al.*, 2020, Groß *et al.*, 2020a).

Méthodologies pour la sélection

La mise en œuvre de l'évaluation génétique selon la méthode du BLUP (*Best Linear Unbiased Predictor* ou « meilleur prédicteur linéaire non biaisée ») modèle animal a constitué une étape majeure à partir des années 1990. A partir des années 2000, c'est le début de la sélection assistée par marqueurs avec l'identification de mutations causales pour des caractères monogéniques pris en compte en sélection (Figure 1). La sélection génomique est l'innovation suivante, à l'origine de la réorganisation en profondeur des étapes requises pour les programmes d'amélioration génétique. Démarrée en France en 2009 chez les trois principales races laitières bovines Holstein, Montbéliarde et Normande (Boichard, 2015), elle s'est progressivement étendue à d'autres races, notamment à viande, et à d'autres espèces (porc, petits ruminants, volailles, poissons). Elle s'appuie sur l'existence d'une population de référence pour laquelle des informations de phénotypage et de génotypage sont disponibles, permettant ensuite d'inférer les phénotypes sur la simple information des génotypes dans la population sélectionnée. Cette

approche a réduit considérablement l'intervalle de génération pour l'amélioration génétique notamment des caractères laitiers à mesurer chez les descendantes des taureaux à qualifier comme reproducteurs, et les coûts associés.

La sélection génomique s'appuie sur la connaissance des polymorphismes à chaque position testée (de plusieurs milliers à plusieurs centaines de milliers). Pour un caractère sous contrôle génétique complexe (contrairement aux anomalies mono-génétique par exemple), une hypothèse courante est de considérer que son déterminisme est polygénique et résulte de l'effet d'un très grand nombre de gènes à effet très petit. La sélection génomique, fondée sur une couverture dense et répartie sur tout le génome répond remarquablement bien à cette hypothèse. Afin d'améliorer la précision des prédictions génomiques, l'ambition est actuellement de s'appuyer sur l'annotation des génomes (Clark *et al.*, 2020) pour établir une typologie des SNP sur la base de leur signification biologique : dans un gène codant une

protéine (exon, intron, régions de régulation), dans un gène non codant, sans annotation connue, etc. Cette typologie fonctionnelle permet d'introduire une pondération d'ordre biologique dans les équations de prédiction, attendue pour en améliorer la précision. Des avancées dans ce domaine sont en cours chez les bovins (Xiang *et al.*, 2020), le porc (Groß *et al.*, 2020a)

et la poule (Groß *et al.*, 2020b). Outre ces informations issues du génome, un front de science connexe est lié à la prise en compte des informations d'épigénome et de microbiome dans ces équations et à la réflexion sur l'apport du *deep learning* et de l'intelligence artificielle dans le domaine de la sélection.

Anticipation des besoins de formation

Les recherches en génétique animale ont profondément évolué depuis le démarrage des programmes d'amélioration génétique en 1950, marquées par des avancées majeures dues notamment aux méthodes statistiques utilisées et aux capacités d'investigation fournies par la génomique. Les verrous actuels ne sont plus liés aux capacités de séquençage et génotypage. En effet, les besoins sont dorénavant orientés vers : i) la caractérisation et le déploiement à grande échelle des mesures pertinentes pour les nouveaux phénotypes ciblés, avec des méthodes standardisées et reproductibles et des ontologies alignées (vocabulaire partagé) ; ii) l'annotation fonctionnelle des séquences génomiques et des polymorphismes ; iii) la prise en compte d'informations biologiques dans les équations de prédiction génomique (ex : annotation fonctionnelle des SNP) ; iv) l'étude de nouvelles échelles

biologiques (cellules uniques, holobionte et populations d'holobiontes), v) l'étude des interactions avec la combinatoire entre génome (en incluant le génome mitochondrial) et épigénome de l'hôte, microbiomes, environnement et pratiques liées au système d'élevage, vi) l'intégration de l'ensemble des données de toutes natures pour comprendre la transmission de la variabilité des phénotypes et leur plasticité au cours de la vie et en réponse à l'environnement.

La figure 3 illustre de manière schématique l'interface entre données réelles et données simulées pour l'amélioration génétique. L'intégration des données et la modélisation pour la prédiction nourrissent cette interface, avec l'ambition d'avancer vers une plus forte intégration de l'ensemble des données produites et disponibles.

Connaissance et annotation fonctionnelle des génomes

Ce domaine de recherche requiert des compétences en bioinformatique et en biologie computationnelle. Il devra s'appuyer sur une excellente connaissance de l'organisation et de la structure des génomes à l'échelle des séquences (visualisation linéaire d'un génome de référence et visualisation avec des graphes d'un pan-génome) comme de son organisation biophysique (organisation en chromosomes, conformation tridimensionnelle de la chromatine,

boucles de transcription, etc.). Les bases de données génomiques nécessitent un apprentissage dédié pour leur utilisation et aussi le dépôt de nouvelles informations. On peut ici anticiper l'importance grandissante de l'alignement des ontologies pour faciliter la fouille de données et de texte et capitaliser sur ce qui est disponible aussi bien dans les bases de données publiques que dans les publications (*data mining*).

Sélection génomique intégrative et identification de prédicteurs

Ce domaine de recherche nécessite un ressourcement des compétences disciplinaires, d'une part en génétique-génomique quantitative (théorie, concepts, méthodologies statistiques, plans de croisement, etc.) et, d'autre part, en biostatistiques appliquées à l'intégration de données (inférence de réseaux, recherche de causalités, exploitation conjointe de mesures phénotypiques et omiques, etc.). La génétique-génomique quantitative et la biologie des systèmes se joignent ici et seront nourries par les

efforts sur l'annotation fonctionnelle des génomes. On peut anticiper qu'elles le seront aussi par les approches basées sur l'apprentissage profond et l'intelligence artificielle. Les prédicteurs attendus ici incluent des marqueurs génétiques suivis lors de la transmission des caractères héritable associés, et des biomarqueurs qui prédisent par exemple des réponses aux échelles des individus et des groupes et leur dynamique temporelle au cours de la vie.

Big data et numérique, plateformes et ressources

Les avancées en génétique animale s'appuient sur des infrastructures partagées et ouvertes qui font maintenant parties intégrantes du paysage de la recherche : centres de ressources biologiques (à vocation reproductive ou non), plateformes analytiques (séquençage, génotypage, métabolomique protéomique, immunophénotypage, etc.), bases de données pour l'organisation et la mise à disposition des informations associées, centres de stockage données (*data center*) et de calcul.

Ces activités nécessitent, outre l'expertise pour générer les données et les interpréter, des compétences en informatique pour construire les bases et organiser leur interopérabilité (ex : recueil des données présentes dans diverses bases pour un même individu). Il est

fortement recommandé de s'inscrire dans une démarche de science ouverte comme le formalise le concept de « FAIRisation » des données, FAIR étant l'acronyme de *Findable, Accessible, Interoperable, Reusable*. La science ouverte favorisera la réutilisation des données pour des méta-analyses notamment.

Deux aspects sont à souligner ici : l'importance de renseigner les métadonnées associées aux données (origine des échantillons, informations sur la race et les conditions d'élevage, données d'environnement, etc.) et de concevoir un plan de gestion des données (complémentaire d'un plan expérimental), dont la définition et la mise en œuvre nécessitent une sensibilisation et des formations dédiées.

Analyse fonctionnelle des polymorphismes, liens de causalités entre génotypes et phénotypes

La génétique est une science des données mais également (faut-il le rappeler ?) une science biologique. Elle a besoin de mesurer des phénotypes pour la sélection. Elle s'appuie sur de l'expérimentation pour valider les associations identifiées et caractériser les relations de causalité (analyse fonctionnelle de mutations candidates), contribuant ainsi à la compréhension des mécanismes biologiques impliqués. La biologie est indispensable au déploiement et au renouvellement des mesures à réaliser pour le phénotypage des animaux (imagerie avec observations associées, suivi du comportement des animaux avec la conception de capteurs, données moléculaires à l'échelle de sous-populations cellulaires voire de cellules uniques, etc.).

Dans ce domaine, il est nécessaire de développer des modèles expérimentaux *in vivo*, *in vitro* ou *ex vivo*, soit sur l'espèce cible soit sur des espèces dites modèles comme les rongeurs, le poisson-zèbre. De solides connaissances en biologie et physiologie animales sont attendues, aux côtés de formations de pointe en biologie cellulaire et en génétique moléculaire. Il sera important de mobiliser des compétences pour maîtriser les méthodologies de modification ciblée du génome (édition ou ré-écriture du génome), afin de construire les outils ad hoc pour modifier spécifiquement un gène ou un SNP candidat à valider, et ensuite en mesurer les effets. Les experts dans l'édition-réécriture des génomes anticipent la possibilité de modifier simultanément plusieurs dizaines de locus. La maîtrise de ces techniques pour la recherche semble incontournable.

Conclusion : savoir ouvrir de nouvelles boîtes

Quelle sera la couleur scientifique de la génétique du XXI^e siècle ? Quelle place occuperont les races et ressources génétiques dites anciennes, conservées pour leur intérêt potentiel, à une époque où la transition agroécologique n'était pas engagée ? La biodiversité est identifiée comme un levier d'action majeur pour la résilience des systèmes en élevage et leur durabilité. La préserver voire l'augmenter sous contrainte de sélection pour des caractères d'intérêt est un enjeu fort qui doit mobiliser les chercheurs en génétique, dans un contexte de fortes attentes sur le bien-être animal, la réduction de l'empreinte environnementale et l'ancrage dans la bioéconomie avec notamment une contribution de l'élevage pour

améliorer le bouclage des cycles (Peyraud *et al.*, 2020).

Le numérique s'impose dans de très nombreux domaines d'activité, avec le déploiement de cursus pédagogiques adaptés. Or, le vivier de compétences sur la science des données appliquée à génomique animale pour la transition agroécologique est actuellement insuffisant et ne couvre pas l'ensemble des besoins des secteurs public et privé. Les jeunes diplômés en science des données ou data scientists méconnaissent les opportunités offertes par ce domaine qui s'est largement renouvelé dans les années passées. Il sera intéressant de considérer qu'un.e

généticien.nne nouvelle génération en sélection génomique est un data scientist qui maîtrise les concepts de la génétique-génomique quantitative et qui va s'employer à exploiter conjointement et de manière innovante des jeux complexes de données pour préciser les prédictions génétiques et optimiser l'exploitation de la diversité génétique. Des compétences complémentaires seront attendues pour identifier et valider des causalités et contribuer à la compréhension des mécanismes biologiques en jeu.

La recherche en général requiert des bases à la fois théoriques et pratiques, et la multidisciplinarité est devenue indispensable. Nourrie par la curiosité, elle teste des hypothèses avec comme corollaire l'acceptation des incertitudes associées, et doit laisser la place à la sérendipité. La transition agroécologique et sociétale nécessite aussi de sortir de son laboratoire pour communiquer sur le renouvellement des enjeux en génomique animale et sur l'éthique en expérimentation animale.

Références

- Beaujean N., Boutinaud M., Devinoy E., Jammes H., Le Guillou S., Le Provost F., Leroux C., Mobuchon L., Pannetier M., Sellem E., Kiefer H. (2020) L'épigénétique et la construction du phénotype chez le bovin. *INRAE Prod Anim.* 33, 109-124.
- Berg G., Rybakova D., Fischer D., Cernava T., Vergès M.C., Charles T., Chen X., Cocolin L., Eversole K., Corral G.H., Kazou M., Kinkel L., Lange L., Lima N., Loy A., Macklin J.A., Maguin E., Mauchline T., McClure R., Mitter B., Ryan M., Sarand I., Smidt H., Schelkle B., Roume H., Kiran G.S., Selvin J., Souza R.S.C., van Overbeek L., Singh B.K., Wagner M., Walsh A., Sessitsch A., Schloter M. (2020) Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 8, 103.
- Boichard D. (2015) Sélection génomique chez les bovins laitiers : les raisons de son succès. *Sélectionneur Français* 66, 11-20.
- Bourneuf E., Otz P., Pausch H., Jagannathan V., Michot P., Grohs C., Piton G., Ammermüller S., Deloche M.C., Fritz S., Leclerc H., Péchoux C., Boukadiri A., Hozé C., Saintilan R., Créchet F., Mosca M., Segelke D., Guillaume F., Bouet S., Baur A., Vasilescu A., Genestout L., Thomas A., Allais-Bonnet A., Rocha D., Colle M.A., Klopp C., Esquerré D., Wurmser C., Flisikowski K., Schwarzenbacher H., Burgstaller J., Brüggmann M., Dietschi E., Rudolph N., Freick M., Barbey S., Fayolle G., Danchin-Burge C., Schibler L., Bed'Hom B., Hayes B.J., Daetwyler H.D., Fries R., Boichard D., Pin D., Drögemüller C., Capitan A. (2017) Rapid Discovery of De Novo Deleterious Mutations in Cattle Enhances the Value of Livestock as Model Species. *Sci Rep.* 7, 11466.
- Calenge F., Martine C., Le Floch N., Phocas F., Morgavi D., Rogel-Gaillard C., Quéré P. (2014) Intégrer la caractérisation du microbiote digestif dans le phénotypage de l'animal de rente : vers un nouvel outil de maîtrise de la santé en élevage ? *INRA Prod. Anim.* 27, 209-222.
- Camarinha-Silva A., Maushammer M., Wellmann R., Vital M., Preuss S., Jörn Bennewitz J. (2017) Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. *Genetics* 206, 1637-1644.
- Caquet T. (ed.), Tixier-Boichard M. (ed.), Gascuel C. (ed.) (2020). *Agroécologie : des recherches pour la transition des filières et des territoires*. QUAE éditions, collection Matière à débattre et décider, 104 p.
- Clark E.L., Archibald A.L., Daetwyler H.D., Groenen M.A.M., Harrison P.W., Houston R.D., Kühn C., Lien S., Macqueen D.J., Reecy J.M., Robledo D., Watson M., Tuggle C.K., Giuffra E. (2020) From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol.* 21, 285.
- Computational Pan-Genomics Consortium (2018) Computational pan-genomics: status, promises and challenges. *Brief Bioinform* 19, 118-135.
- Crysnanto D., Pausch H. (2020) Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery. *Genome Biology* 21, 184.
- Ducrocq V. (2020) Conception et réalisation des évaluations génétiques ou génomiques des reproducteurs. In : *Génétique des animaux d'élevage, diversité et adaptation dans un monde changeant* (E. Verrier coord., D. Milan coord., C. Rogel-Gaillard coord.) QUAE édition, collection Savoir faire, 71-87
- Elsen J.M. (2020) Le choix des reproducteurs et l'efficacité de la sélection. In : *Génétique des animaux d'élevage, diversité et adaptation dans un monde changeant* (E. Verrier coord., D. Milan coord., C. Rogel-Gaillard coord.) QUAE édition, collection Savoir faire, 89-106.
- Giuffra E., Tuggle C.K., the FAANG Consortium (2019) Current achievements and roadmap. *Annual Review of Animal Biosciences* 7, 65-88.
- Groß C., Derks M., Megens H.J., Bosse M., Groenen M.A.M., Reinders M., de Ridder D. (2020a) pCADD: SNV prioritisation in *Sus scrofa*. *Genet Sel Evol.* 52, 4.
- Groß C., Bortoluzzi C., de Ridder D., Megens H.J., Groenen M.A.M., Reinders M., Bosse M. (2020b) Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. *PLoS Genet.* 16, e1009027.
- Hayes B.J., Hans D., Daetwyler H.D. (2018) 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annual Review of Animal Biosciences* 7, 89-102.
- Heams T. (2020) Ethique et amélioration génétique animale. In : *Génétique des animaux d'élevage, diversité et adaptation dans un monde changeant* (E. Verrier coord., D. Milan coord., C. Rogel-Gaillard coord.) QUAE édition, collection Savoir faire, 241-256.

- Huang P., Zhang Y., Xiao K., Jiang F., Wang H., Tang D., Liu D., Liu B., Liu Y., He X., Liu H., Liu X., Qing Z., Liu C., Huang J., Ren Y., Yun L., Yin L., Lin Q., Zeng C., Su X., Yuan J., Lin L., Hu N., Cao H., Huang S., Guo Y., Fan W., Zeng J. (2018). The chicken gut metagenome and the modulatory effects of plant-derived benzyloisoquinoline alkaloids. *Microbiome* 6, 211.
- Imsland F., Feng C., Boije H., Bed'hom B., Fillon V., Dorshorst B., Rubin C.J., Liu R., Gao Y., Gu X., Wang Y., Gourichon D., Zody M.C., Zecchin W., Vieaud A., Tixier-Boichard M., Hu X., Hallböök F., Li N., Andersson L. (2012) The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet.* 8, e1002775.
- Lallias D., Bernard M., Ciobotaru C., Dechamp N., Labbé L., Goardon L., Le Calvez J.M., Bideau M., Fricot A., Prézélin A., Charles M., Moroldo M., Cousin X., Bouchez O., Roulet A., Quillet E., Dupont-Nivet M. (2020) Sources of variation of DNA methylation in rainbow trout: combined effects of temperature and genetic background. *Epigenetics* 30, 1-22.
- Le Roy P., Ducos A., Phocas F. (2019) Quelles performances pour les animaux de demain ? Objectifs et méthodes de sélection. *INRA Prod Anim* 32, 233-246.
- Li J., Bed'hom B., Marthey S., Valade M., Dureux A., Moroldo M., Péchoux C., Coville J.L., Gourichon D., Vieaud A., Dorshorst B., Andersson L., Tixier-Boichard M. (2019) A missense mutation in TYRP1 causes the chocolate plumage color in chicken and alters melanosome structure. *Pigment Cell Melanoma Res.* 32, 381-390.
- Li J., Zhong H., Ramayo-Caldas Y., Terrapon N., Lombard V., Potocki-Veronese G., Estellé J., Popova M., Yang Z., Zhang H., Li F., Tang S., Yang F., Chen W., Chen B., Li J., Guo J., Martin C., Maguin E., Xu X., Yang H., Wang J., Madsen L., Kristiansen K., Henrissat B., Ehrlich S.D., Morgavi D.P. (2020) A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. *Gigascience* 9, g1aa057.
- Meuwissen T.H., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819-1829.
- Michot P., Chahory S., Marete A., Grohs C., Dagios D., Donzel E., Aboukadiiri A., Deloche M.C., Allais-Bonnet A., Chambrial M., Barbey S., Genestout L., Boussaha M., Danchin-Burge C., Fritz S., Boichard D., Capitan A. (2016) A reverse genetic approach identifies an ancestral frameshift mutation in RP1 causing recessive progressive retinal degeneration in European cattle breeds. *Genet Sel Evol.* 48, 56.
- Milan D., Jeon J.T., Looft C., Amarger V., Robic A., Thelander M., Rogel-Gaillard C., Paul S., Iannuccelli N., Rask L., Ronne H., Lundström K., Reinsch N., Gellin J., Kalm E., Roy P.L., Chardon P., Andersson L. (2000) A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science* 288, 1248-1251.
- Peyraud J.L., Aubin J., Barbier M., Baumont R., Berri C., Bidanel J.P., Citti C., Cotinot C., Ducrot C., Dupraz P., Faverdin P., Friggens N., Houot S., Nozières-Petit M.O., Rogel-Gaillard C., Santé-Lhoutellier V. (2019) Quelle science pour les élevages de demain ? Une réflexion prospective conduite à l'INRA. *INRA Prod. Anim.* 32, 323-338.
- Phocas F. (2020) La définition des objectifs de sélection. (2020). In : *Génétique des animaux d'élevage, diversité et adaptation dans un monde changeant* (E. Verrier coord., D. Milan coord., C. Rogel-Gaillard coord.) QUAE édition, collection Savoir faire, 55-70.
- Rubin C.J., Zody M.C., Eriksson J., Meadows J.R., Sherwood E., Webster M.T., Jiang L., Ingman M., Sharpe T., Ka S., Hallböök F., Besnier F., Carlborg O., Bed'hom B., Tixier-Boichard M., Jensen P., Siegel P., Lindblad-Toh K., Andersson L. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587-591.
- Theis K.R., Dheilly N.M., Klassen J.L., Brucker R.M., Baines J.F., Bosch T.C., Cryan J.F., Gilbert S.F., Goodnight C.J., Lloyd E.A., Sapp J., Vandenkoornhuysen P., Zilber-Rosenberg I., Rosenberg E., Bordenstein S.R. (2016) Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. *mSystems* 1, e00028-16.
- Tixier-Boichard M., Lecerf F., Hérault F., Bardou P., Klopp C. (2021). Le projet « Mille Génomes Gallus » : partager les données de séquences pour mieux les utiliser. *INRAE Prod. Anim.* 33, 189-202.
- van den Berg I., Hayes B.J., Chamberlain A.J., Goddard M.E. (2019). Overlap between eQTL and QTL associated with production traits and fertility in dairy cattle. *BMC Genomics*, 20 (1) : 29
- Verrier E. (2020) De la domestication à la sélection génomique : une brève histoire de la sélection animale. In : *Génétique des animaux d'élevage, diversité et adaptation dans un monde changeant* (E. Verrier coord., D. Milan coord., C. Rogel-Gaillard coord.) QUAE édition, collection Savoir faire, 29-46.
- Vignal A., Monget P., Rogel-Gaillard C. (2020) Evolution des connaissances sur les génomes, épigénomes et microbiomes. In : *Génétique des animaux d'élevage, diversité et adaptation dans un monde changeant* (E. Verrier coord., D. Milan coord., C. Rogel-Gaillard coord.) QUAE édition, collection Savoir faire, 169-191.
- Xiang R., van den Berg I., MacLeod I.M., Hayes B.J., Prowse-Wilkins C.P., Wang M., Bolormaa S., Liu Z., Rochfort S.J., Reich C.M., Mason B.A., Vander Jagt C.J., Daetwyler H.D., Lund M.S., Chamberlain A.J., Goddard M.E. (2019) Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc. Natl. Acad. Sci. USA* 116, 19398-19408.