



HAL
open science

Génomique des canards

Alain Vignal, Christian Diot, Caroline Molette, Thomas Faraut, Man Rao,
Frederique Pitel, Valerie Fillon, Christel Marie-Etancelin

► **To cite this version:**

Alain Vignal, Christian Diot, Caroline Molette, Thomas Faraut, Man Rao, et al.. Génomique des canards. INRA Productions Animales, 2013, 26 (5), pp.391-402. <hal-01207754>

HAL Id: hal-01207754

<https://hal.science/hal-01207754v1>

Submitted on 29 May 2020

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.



HAL Authorization

A. VIGNAL^{1,2}, C. DIOT^{3,4}, C. MOLETTE^{5,6,7}, M. MORISSON^{1,2}, T. FARAUT^{1,2}, M. RAO^{1,2},
F. PITEL^{1,2}, V. FILLON^{1,2}, C. MARIE-ETANCELIN⁸

¹ INRA, UMR444 Génétique Cellulaire, F-31326 Castanet-Tolosan, France

² Université de Toulouse, INPT ENVT, UMR444 Génétique Cellulaire, F-31076 Toulouse, France

³ INRA, UMR1348 PEGASE, F-35590 Saint-Gilles, France

⁴ Agrocampus OUEST, UMR1348 PEGASE, F-35042 Rennes, France

⁵ INRA, UMR1289 TANDEM, F-31326 Castanet-Tolosan, France

⁶ Université de Toulouse, INPT ENSAT, UMR1289 TANDEM, F-31326 Castanet-Tolosan, France

⁷ Université de Toulouse, INPT ENVT, UMR1289 TANDEM, F-31076 Toulouse, France

⁸ INRA, URO631 SAGA, F-31326 Castanet-Tolosan, France

Courriel : alain.vignal@toulouse.inra.fr

Les progrès récents en séquençage des génomes ont amené une baisse substantielle des coûts et une augmentation significative des volumes produits. Ainsi, de nombreuses espèces animales sont maintenant séquencées à la suite des espèces d'intérêt agronomique majeur, que sont la poule, le bovin et le porc. La séquence génomique du canard vient d'être publiée, permettant d'ouvrir un champ nouveau à la génomique de cette espèce.

L'animal le plus souvent utilisé, à plus de 92% dans la filière des palmipèdes gras, est le canard mulard, hybride stérile issu du croisement intergénérique de deux espèces de canards : *Anas platyrhynchos* (le canard commun) et *Cairina moschata* (le canard de Barbarie). La distance phylogénétique entre ces deux espèces est de 13,5 à 20,2 millions d'années (Gonzalez *et al* 2009). A titre de comparaison, la distance phylogénétique entre le cheval (*Equus caballus*) et l'âne (*Equus asinus*) est de 6,4 à 12,7 millions d'années (<http://www.timetree.org/>).

Historiquement, le canard de Barbarie a été le premier à être utilisé pour la production de foie gras, mais il ne représente plus que 5% de la production. Il a été supplanté par le canard mulard, dont l'intérêt réside dans une robustesse générale plus grande et une facilité accrue dans la conduite d'élevage. Cet hybride bénéficie d'un effet d'hétérosis favorable démontré sur le poids (Larzul *et al* 2006) et supposé sur la qualité (Théron *et al* 2013) du foie produit. En pratique, la production de foie gras de canard mulard en France a été rendue possible par la maîtrise de l'insémination artificielle de la cane commune par le sperme du canard de Barbarie (Tai Liu et Tai 1991). Le canard commun présentant de moindres performances, n'est jamais utilisé.

Au niveau international et hors de la filière des palmipèdes gras, le terme générique « canard » se réfère au canard commun, ou canard Pékin, qui est donc largement plus étudié que le canard de Barbarie, soit comme modèle pour la résistance aux virus aviaires, soit en raison d'intérêts agronomiques pour la production de viande et d'œufs, en particulier sur le continent asiatique. De fait, la quasi-totalité des outils de génomique développés à ce jour concernent le canard commun.

La génomique, qui s'intéresse au génome pris dans son ensemble, est une discipline provenant de l'association de la biologie moléculaire et cellulaire d'une part, et de la génétique classique d'autre part. Son essor tient cependant beaucoup aux développements importants des technologies de séquençage de l'ADN et de la bioinformatique, permettant l'analyse d'un volume croissant de données. De nos jours, elle regroupe un ensemble d'analyses qui vont de l'établissement de cartes du génome et de l'identification de nouveaux gènes, à l'étude de leurs fonctions. Ainsi, on parlera de génomique structurale pour la cartographie et le séquençage des génomes, et de génomique fonctionnelle pour l'étude de l'expression et de la fonction des gènes, notamment à travers l'étude du transcriptome (ensemble des transcrits) et du protéome (ensemble des protéines).

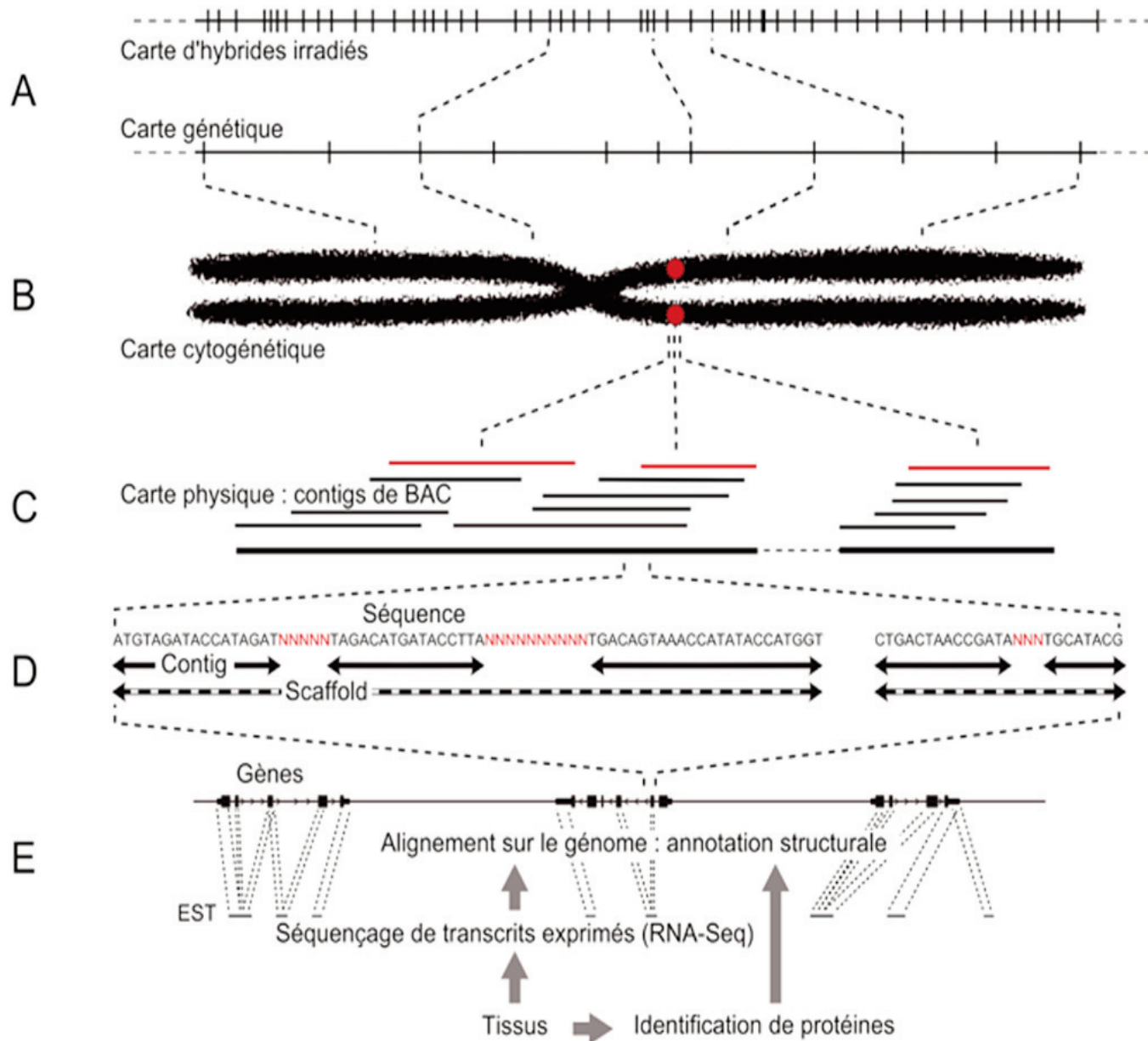
Les applications issues de la génomique structurale sont par exemple la gestion des populations et l'identification de marqueurs génétiques liés à des caractères importants par recherche de QTL (« *Quantitative Trait Locus* ») ou par association génétique. Pour la génomique fonctionnelle, nous pouvons citer la recherche de marqueurs moléculaires de phénotypes complexes ou la mise en évidence de l'implication de voies métaboliques particulières dans les phénomènes biologiques. Nous allons ici récapituler l'état des recherches en génomique du canard, avec une vision appliquée à la filière des palmipèdes gras.

1 / Connaissances sur le génome du canard commun

1.1 / Séquence du génome

Les machines à séquencer produisent des lectures (ou séquences) continues du code génétique de l'ADN allant de 100 à 1 000 paires de bases (pb^{*}). Pour séquencer un vertébré, dont le génome est de l'ordre du milliard de paires de bases (Gb^{*}), il faut donc produire des dizaines de millions de lectures qu'il faut ensuite assembler en chromosomes dont la longueur est en général de l'ordre de plusieurs dizaines de millions de paires de bases (Mb^{*}). De plus, l'assemblage

¹ Les astérisques renvoient à un glossaire en fin de document.

Figure 1. Cartes, séquence et annotation structurale*

Divers niveaux de cartographie sont nécessaires pour assembler des scaffolds* dont la taille varie entre quelques kb* et 5,9 Mb* (le plus grand de l'assemblage du canard), en séquences de chromosomes. Le séquençage de transcrits et de protéines est nécessaire à la détection des gènes et à la détermination de leur structure (annotation structurale*).

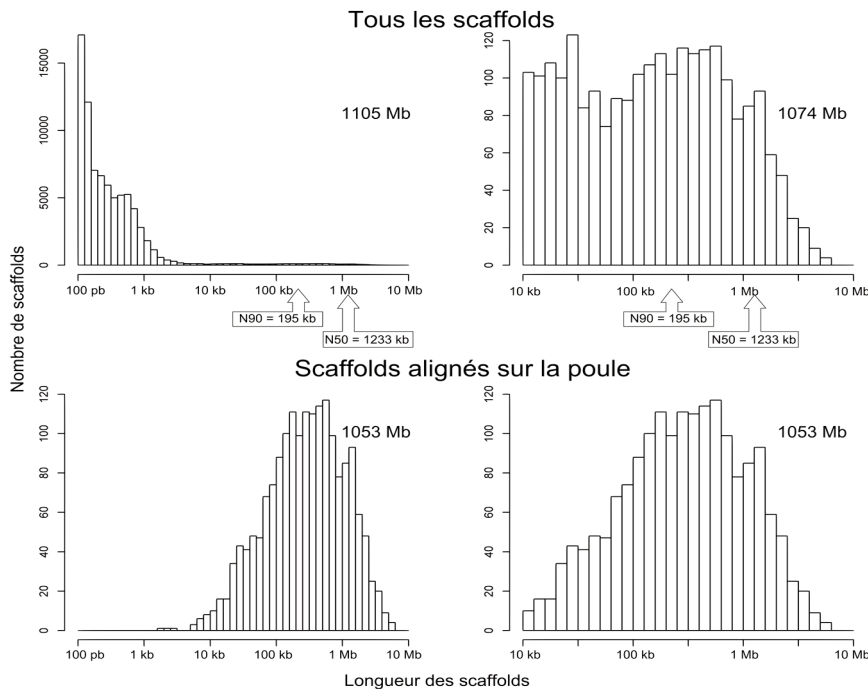
A : la carte d'hybrides irradiés et la carte génétique proviennent toutes deux du comptage d'événements dont la fréquence d'observation est fonction de la distance : nombre de cassures chromosomiques entre marqueurs pour les cartes d'hybrides irradiés (unité de distance centiRay (cR*)) et nombre de crossing-overs pour les cartes génétiques (unité de distance centiMorgan (cM*)). Le rapport avec des distances physiques le long des chromosomes va varier selon un certain nombre de facteurs, incluant l'espèce étudiée ou le chromosome pour les cartes génétiques, avec par exemple une variation allant de 2 à 12 cM*/Mb* chez la poule (Groenen *et al* 2009) ou la dose d'irradiation pour les cartes RH*, avec par exemple une moyenne de 25 cR*/Mb* chez la poule, pour une dose d'irradiation de 6 000 rads (Morisson *et al* 2007).

B : la carte cytogénétique* provient de l'hybridation de sondes ADN sur des chromosomes en métaphase observés ensuite au microscope. Sa résolution est de l'ordre de quelques Mb*, selon l'état de condensation des chromosomes utilisés ou selon les sondes choisies.

C : la carte physique en contigs* de BAC* résulte de l'estimation du chevauchement de fragments d'ADN de l'espèce à étudier, contenus dans des clones artificiels de bactéries. La résolution est très bonne, sachant que la taille moyenne des inserts dans les BAC* est de l'ordre de 100 à 200 kb*. Un autre avantage est qu'un BAC* pourra facilement être séquençé.

D : Séquence du génome.

E : L'annotation structurale des gènes utilise des données de séquences de transcrits et de protéines de l'espèce étudiée, mais peut aussi tirer parti de données provenant d'autres espèces. L'annotation fonctionnelle* des gènes nécessite des données biologiques, soit déterminées expérimentalement dans l'espèce étudiée, soit déduites de la fonction de gènes orthologues* dans une espèce proche.

Figure 2. Distribution des tailles de scaffolds* de la séquence du canard commun.

Gauche : tous les scaffolds* de plus de 100 pb*. Droite : tous les scaffolds* de plus de 10 kb*. Bas : uniquement les scaffolds* qui ont pu être alignés sur la séquence de la poule et donc pour lesquels une prédiction d'assignation chromosomique chez le canard est possible.

N50 et N90 : valeur de taille de scaffold* telle que 50 et 90% de la séquence du génome est comprise dans des scaffolds* de taille égale ou supérieure à N50 et N90. Une très grande proportion du génome du canard, 1053 Mb* au total sur 1105 Mb* séquencés, est comprise dans des scaffolds* de plus de 10 kb* avec une prédiction d'assignation chromosomique.

des nombreuses lectures produites en une séquence continue représentant un chromosome est rendu plus complexe encore à cause de caractéristiques particulières des génomes, telles que la présence de séquences répétées ou de régions se prêtant mal au séquençage. Après assemblage des lectures de séquence par voie bioinformatique, on obtient un très grand nombre de contigs* (portions de séquence continue) et de scaffolds* (ou échafaudages, portions de séquence discontinue, résultant de l'assemblage de contigs*) (Vignal 2011). Afin d'organiser ces contigs* et scaffolds* en chromosomes, il faudra donc s'appuyer sur plusieurs méthodes de cartographie complémentaires (figure 1).

Le séquençage du génome du canard commun a été réalisé au « Beijing Genomics Institute » (BGI) à l'aide de la technologie Illumina™ (Huang *et al* 2013). Cette technologie de séquençage fait partie des nouvelles générations de séquençage, dites « Next-Generation Sequencing » (NGS), ou de manière plus appropriée « séquençage parallèle ». Ces technologies ont permis ces dernières années de réaliser une importante diminution des coûts et des délais de séquençage par rapport à la méthode

« traditionnelle » de Sanger*, introduite en 1977 (Sanger *et al* 1977). Cependant, la longueur des lectures individuelles de séquence étant plus courte, il en résulte une séquence plus morcelée, avec des contigs* et scaffolds* plus nombreux et plus petits, d'où un assemblage plus complexe.

La séquence du génome du canard est sous la forme de 227 597 contigs* dont la taille moyenne est de 4,7 kb* et de 78 847 scaffolds* dont la taille moyenne est de 14,0 kb*. Le plus grand contig* fait 264 kb* et le plus grand scaffold* 5,9 Mb*, alors que la taille du plus grand chromosome (APL1*) est de l'ordre de 200 Mb*. Cependant, ces données ne permettent pas d'avoir une bonne représentation de la qualité de la séquence en raison de la distribution très particulière des tailles de scaffolds* et de contigs* (voir figure 2 pour les contigs*), avec un très grand nombre de petits contigs* et un nombre très faible de grands. Pour remédier à cela, une mesure statistique couramment utilisée pour estimer la qualité de la continuité de la séquence et de la couverture du génome est le N50 ou le N90, utilisé pour les contigs* et les scaffolds*. Ce sont les valeurs pour lesquelles au moins 50 ou 90% du génome sont couverts par des

scaffolds* ou contigs* de taille supérieure à N50 ou N90. Pour le canard commun, le N50 contig* est de 26 kb* (N90 = 3 kb*) et le N50 scaffold* de 1,2 Mb* (N90 = 195 kb*) (figure 2). A titre de comparaison, les statistiques pour la poule dont le séquençage du génome a été réalisé par la méthode de Sanger* et qui possède un génome avec des caractéristiques de taille et de nombre de chromosomes comparables à celui du canard, sont de 36 kb* pour le N50 contig* et de 7,1 Mb* pour le N50 scaffold*. Pour plus de détails sur l'assemblage d'un génome en contigs* et scaffolds* et la comparaison de l'assemblage du génome du canard avec celui de la poule, réalisé par séquençage de type Sanger*, se reporter à Vignal (2011).

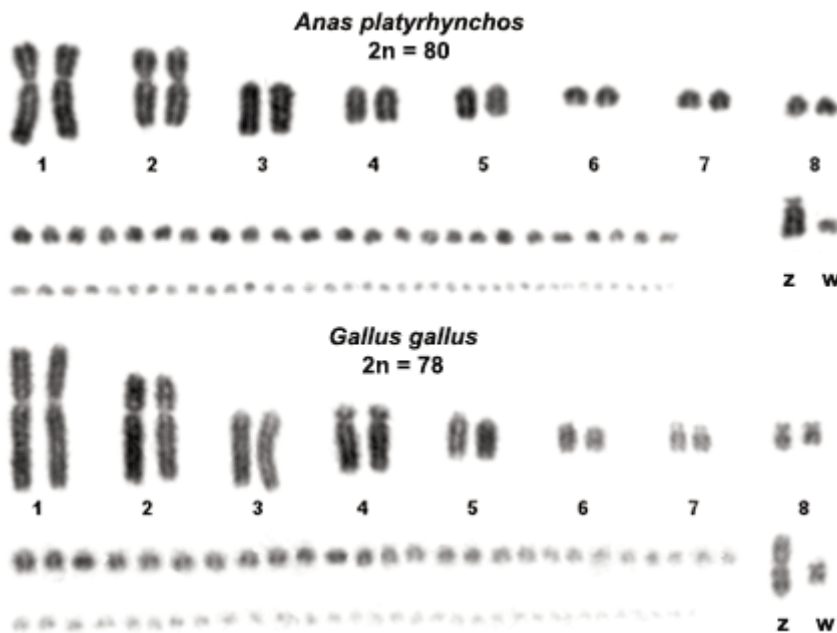
A ce stade d'avancement, nous avons donc un génome dont la taille est estimée à 1,26 Gb et morcelé en 40 paires de chromosomes (figure 3), pour lequel 1,1 Gb de séquence assemblée a été produite. De cette séquence assemblée, 90% l'est par des portions de séquence continue allant de 3 à 264 kb* et des portions de séquence discontinue de 195 à 5 900 kb* (Huang *et al* 2013). Sans l'appui de cartes, ces portions de séquence ne sont pas assignées à des chromosomes.

1.2 / Assemblage en chromosomes

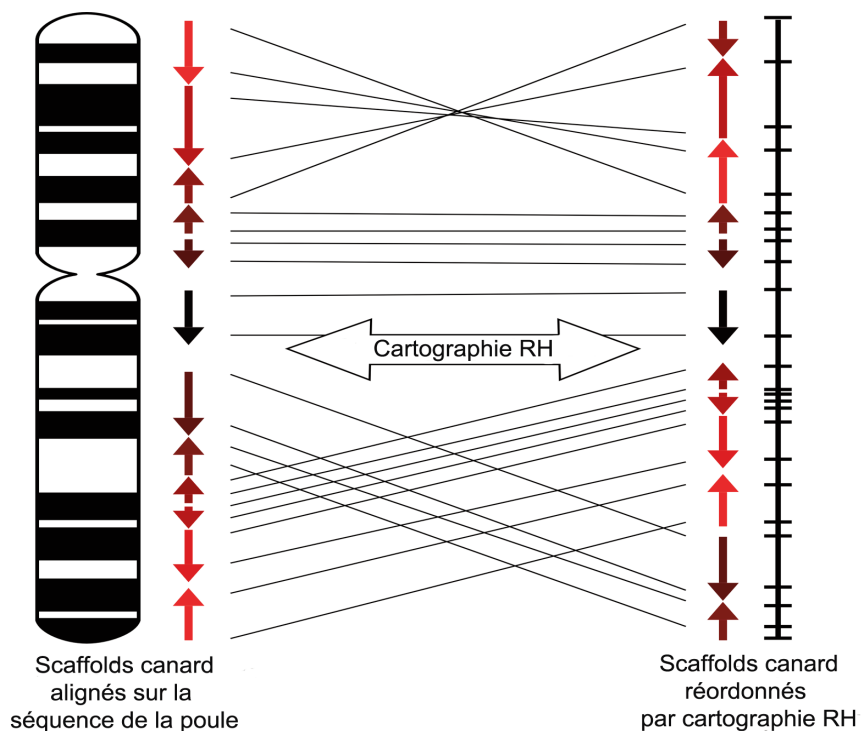
Grâce à la grande similarité de leurs caryotypes respectifs, une première approximation de l'ordre des scaffolds* le long des chromosomes de canard peut être obtenue par comparaison avec la séquence de la poule.

Le génome de la poule est maintenant très documenté, puisque son séquençage date de 2004 (Hillier *et al* 2004) et que de nombreuses études concernant sa structure et son fonctionnement ont été publiées depuis. Il est de ce fait considéré comme génome de référence en génomique aviaire. De plus, les études de cartographie cytogénétique comparées par hybridation de sondes moléculaires (FISH* : « Fluorescent In Situ Hybridization ») ont montré une grande stabilité de l'organisation chromosomique chez les oiseaux, les réarrangements interchromosomiques étant particulièrement rares (Kayang *et al* 2006, Fillon *et al* 2007, Skinner *et al* 2009, Zhang *et al* 2011). En se basant sur ces faits, une première approximation de l'ordre des scaffolds* du canard a été obtenue par alignement de ces derniers sur la séquence de la poule. Cet ordre est utilisé en première instance, par exemple pour développer des marqueurs génétiques pour un chromosome donné.

Cependant, quelques différences existent entre les génomes de la poule et du canard. La poule comporte 39 paires de

Figure 3. Caryotype de canard et de poule.

Les caryotypes du canard (haut) et de la poule (bas) sont très similaires, comportant tous deux des macrochromosomes et des microchromosomes. La seule différence visible est une paire de microchromosomes supplémentaire chez le canard. Cette paire supplémentaire est le chromosome 10 (APL10), qui correspond au bras court du chromosome 4 de la poule (GGA4*) (Skinner *et al* 2009).

Figure 4. Principe de l'ordonnement des scaffolds*, basé sur la cartographie RH*.

Ordonner directement plusieurs milliers de marqueurs par la cartographie RH* pose de nombreux problèmes et l'obtention de cartes par cette méthode est longue et difficile. Dans notre cas, les marqueurs sont les scaffolds* obtenus par assemblage de la séquence du canard. Une manière de simplifier le problème, consiste à proposer un premier ordre basé sur l'alignement de ces scaffolds* sur génome de la poule (gauche), puis d'utiliser les données de cartographie RH* afin de déterminer l'ordre chez le canard (droite). Cette approche a été utilisée pour réaliser des premières cartes RH* chez le canard (Rao *et al* 2012).

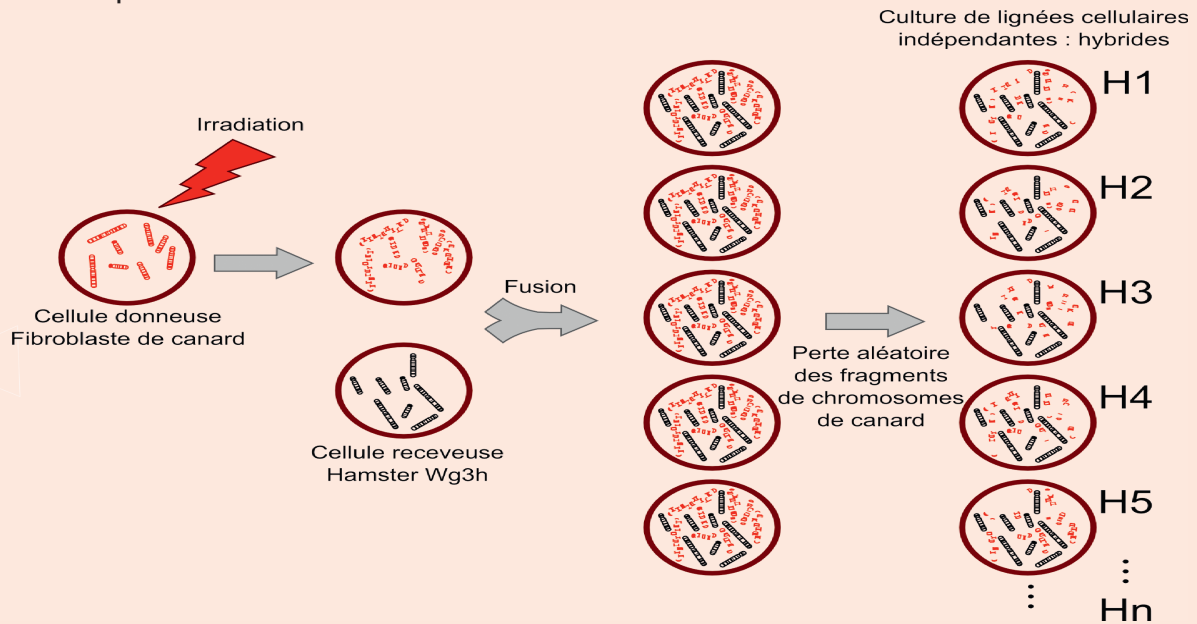
chromosomes, alors que le caryotype du canard commun présente 40 paires de chromosomes. Cette différence de nombre est due au seul réarrangement inter-chromosomique connu à ce jour entre les génomes de la poule et du canard : l'homologue du chromosome 4 de la poule est présent sous la forme de 2 chromosomes chez le canard : les chromosomes 4 et 10 (Skinner *et al* 2009). De plus, seul un petit nombre de réarrangements intra-chromosomiques ont été détectés (Fillon *et al* 2007, Skinner *et al* 2009).

Afin de mieux déterminer l'ordre des scaffolds* le long des chromosomes, ceux-ci ont été alignés sur des cartes spécifiques du canard, mais qui ont l'inconvénient à ce jour, d'être rudimentaires. En effet, seuls 225 scaffolds* couvrant 289 Mb* ont pu être assignés à des chromosomes (Huang *et al* 2013), en les alignant sur les premières cartes produites, l'une génétique (Huang *et al* 2006) et l'autre cytogénétique (Skinner *et al* 2009).

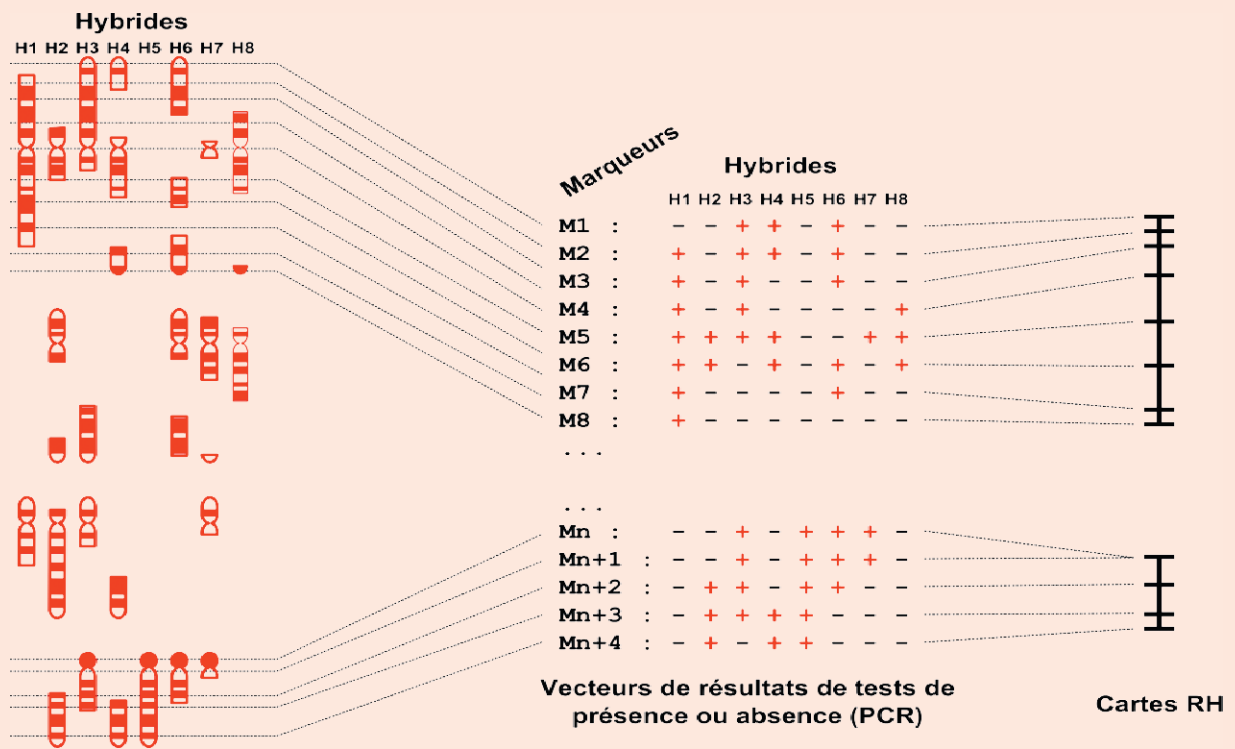
Afin de mieux déterminer l'ordre des scaffolds* le long des chromosomes, un panel d'hybrides irradiés* a été réalisé, permettant de produire des premières cartes denses de chromosomes. Pour le principe de la cartographie d'hybrides irradiés* (RH*), voir encadré 1. Plusieurs réarrangements intra-chromosomiques inédits entre les chromosomes de poule et de canard ont pu ainsi être mis en évidence (Rao *et al* 2012). Le principe de l'ordonnement et de l'orientation des scaffolds*, basé sur la séquence de la poule, puis revu par cartographie RH* chez le canard est présenté dans la figure 4. Des cartes pour deux premiers chromosomes ont été réalisées par test classique de présence ou absence de fragments d'ADN courts (marqueurs) en PCR, tel que décrit dans l'encadré 1B (Rao *et al* 2012). Afin de réaliser des cartes pour l'ensemble du génome, la présence ou l'absence de fragments chromosomiques de canard est maintenant testée par séquençage des hybrides et alignement des séquences obtenues sur les scaffolds* de la séquence du génome. De la sorte, c'est la présence ou l'absence de la séquence des scaffolds* dans les hybrides qui est testée : les scaffolds* servent eux-mêmes de marqueurs pour construire les cartes et sont ainsi ordonnés le long du génome.

Une carte génétique* de nouvelle génération est en cours de réalisation à l'INRA, dans le cadre d'une recherche de QTL. Cette carte viendra en appui à la carte RH* en apportant un degré de résolution qui lui est propre, intermédiaire entre la carte RH* et le niveau chromosomique (figure 1).

Un niveau de cartographie plus fin nécessiterait la réalisation d'une carte de

Encadré 1. Panel RH* et cartographie RH*.**A : production du panel RH.**

Le but est de produire des lignées cellulaires possédant chacune une portion du génome du canard, sous forme fragmentée. Les cellules donneuses de canard subissent une forte dose de rayons gamma provoquant la cassure de leurs chromosomes, ce qui leur est létal. Ces cellules sont ensuite fusionnées avec des cellules de hamster dans le but de récupérer les fragments de chromosomes cassés. Au cours des divisions cellulaires, une fraction aléatoire des morceaux de chromosomes de la cellule donneuse de canard sont perdus. En partant de plusieurs événements de fusion, une collection de lignées cellulaires est ainsi construite, chacune conservant une représentation différente du génome de départ.

B : cartographie de marqueurs.

La présence ou l'absence de fragments de chromosomes de canard dans les lignées d'hybrides irradiés est testée, généralement par PCR (marqueurs PCR). Plus une paire de marqueurs est proche dans le génome, plus la probabilité de co-rétention dans les hybrides sera élevée. La fraction d'hybrides ayant retenu une paire de marqueurs donne une estimation de leur distance exprimée en cR* (centi Ray). En testant un grand nombre de marqueurs, des cartes de chromosomes peuvent être ainsi établies.

contigs* de BAC* (« *Bacterial Artificial Chromosome* ») ou chromosome artificiel de bactérie) (figure 1). A notre connaissance, bien que des banques de BAC* (Yuan *et al* 2006) et de fosmidés* (Moon et Magor 2004) aient été produites, aucun projet de cartographie faisant appel à ces outils n'existe à ce jour.

1.3 / Annotation structurale et fonctionnelle*

La séquence assemblée, qu'elle soit ou non ordonnée le long des chromosomes, n'est qu'une succession des 4 lettres A, C, G et T composant le code génétique. La première étape élémentaire de compréhension fonctionnelle d'un génome est de repérer les gènes. Cette étape peut être réalisée dès l'assemblage en scaffolds* et l'obtention de séquences suffisamment longues à l'échelle de la taille des gènes. La position des scaffolds* le long des chromosomes est un plus, mais n'est pas indispensable à ce stade.

Une source de données essentielle pour l'annotation est la séquence de transcrits et de protéines. Un effort important de séquençage de transcrits a été réalisé, à partir de plusieurs tissus, dont le cerveau, le muscle, la rate, l'intestin et le poumon. Pour cette étude préliminaire, l'INRA s'est concentré sur les tissus permettant de détecter un grand nombre de gènes (cerveau) ou correspondant à des caractères agronomiques d'intérêt (muscle). D'autres projets de RNA-Seq sont en cours à l'INRA sur le foie (cf. § 3.1). D'autres membres du consortium international de séquençage du canard ont séquencé des tissus d'animaux infectés par des virus influenza hautement ou faiblement pathogènes (Huang *et al* 2013). Ces données ont été combinées à d'autres sources de données, principalement des séquences de transcrits de canard déjà publiées, des séquences de protéines d'autres espèces d'oiseaux, de mammifères et d'autres vertébrés, et des prédictions de gènes obtenues par bioinformatique. On peut déjà toucher du doigt l'aspect exploratoire de ces approches, en comparant le nombre de gènes codant pour des protéines trouvés par l'un (19 144 gènes : BGI) ou l'autre (15 634 gènes : Ensembl) des membres du consortium. Une annotation structurale* du génome du canard, comportant 16 450 modèles de gènes, 17 169 modèles de transcrits et 16 353 modèles de protéines est disponible sur le site de Ensembl (http://pre.ensembl.org/Anas_platyrhynchos/Info/Index).

L'annotation fonctionnelle* consiste à spécifier la fonction des gènes détectés. Comme assez peu de gènes ont été étudiés chez le canard, cette annotation est réalisée par similarité de séquence avec

des gènes identifiés dans d'autres espèces. Concrètement, des recherches de gènes orthologues* ont été faites dans plusieurs bases de données, incluant des bases de protéines annotées, de domaines fonctionnels protéiques ou la base de données d'annotations Gene Ontology*.

2 / Variabilité interindividuelle, cartographie génétique et recherche de QTL

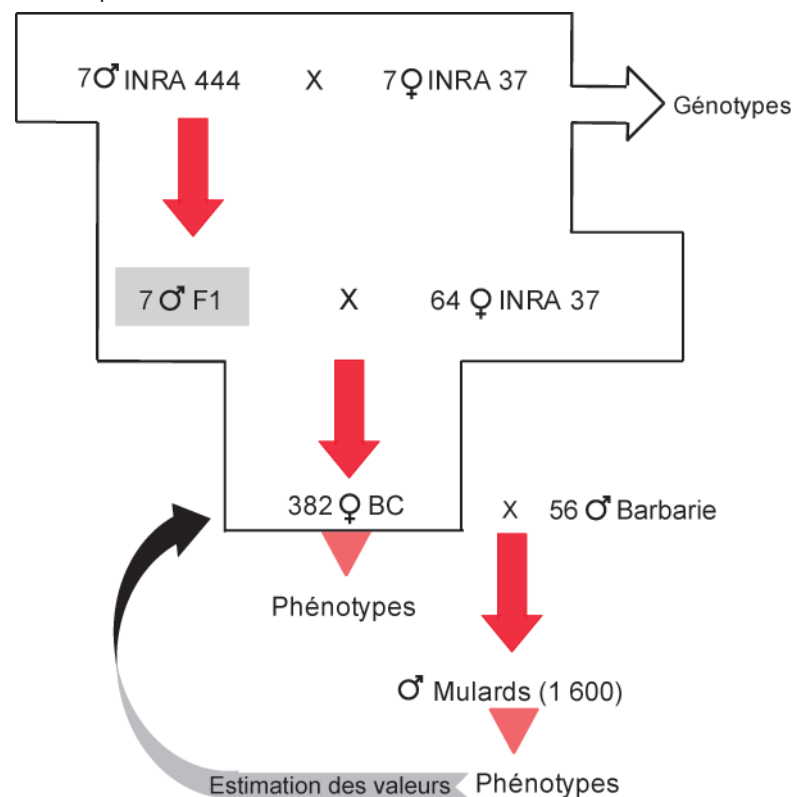
2.1 / Cartes et marqueurs génétiques

La variabilité génétique interindividuelle est exploitée pour l'amélioration génétique des espèces d'élevage, dont le canard. Une des approches de la génomique consiste à rechercher des associations entre des marqueurs génétiques moléculaires de l'ADN et des caractères quantitatifs : il s'agit de l'identification de QTL. Les marqueurs génétiques les plus couramment utilisés sont les microsatellites et les SNP* (« *Single Nucleotide Polymorphism* ») ou polymorphisme de base nucléotidique ponctuel. Les microsatellites sont des variations interindi-

viduelles du nombre de copies d'un motif nucléotidique court (généralement di-, tri- ou tétra-nucléotidiques), qui peuvent être détectées par PCR et mesure de la longueur du produit d'amplification sur gel d'électrophorèse (Litt et Luty 1989, Weber et May 1989). L'intérêt de ces marqueurs microsatellites est d'être multi-alléliques. Les SNP* sont une simple substitution d'une base en une position du génome et sont donc di-alléliques (Vignal 2011).

Des cartes génétiques rudimentaires ont été publiées : l'une réalisée en Chine comporte 115 marqueurs microsatellites répartis en 19 groupes de liaison, dont 10 assignés à des chromosomes et couvrant 1 353 cM* (Huang *et al* 2006) et l'autre en France comporte 91 marqueurs microsatellites répartis en 16 groupes de liaison couvrant 778 cM* (Kileh-Wais *et al* 2013). Une troisième carte a été réalisée à Taïwan, comportant 260 marqueurs AFLP (« *Amplification Fragment Length Polymorphism* ») ou Polymorphisme de longueur des fragments d'amplification, répartis en 32 groupes de liaison (Huang *et al* 2009). Si cette dernière carte semble plus complète, il faut noter que les marqueurs AFLP sont

Figure 5. Dispositif de recherche de QTL GeneCan.



Le dispositif GeneCan vise à détecter des QTL en ségrégation chez le canard commun, qui influenceraient les caractéristiques du produit final, le canard mulard. Les valeurs phénotypiques des 382 femelles backcross, issues du croisement entre deux lignées de canard commun INRA 444 et INRA 37, sont estimées d'après des mesures effectuées chez leurs fils mulards. Les mâles Barbarie sont issus d'une seule lignée et ne servent qu'à produire les descendants mulards.

d'un emploi beaucoup moins aisé que les microsatellites, posant des problèmes de standardisation et donc de transfert entre laboratoires. Sachant que le caryotype du canard comporte 40 paires de chromosomes, on estime que moins de la moitié du génome du canard est couvert par les cartes génétiques actuelles.

Les marqueurs microsatellites, utilisés depuis plus de vingt ans, sont de plus en plus remplacés par les SNP* dans les espèces animales d'intérêt agronomique majeur. Par exemple, la poule bénéficie maintenant d'une puce à SNP* permettant le génotypage simultané de plus de 580 000 SNP* (Kranis *et al* 2013), tandis qu'auparavant quelques centaines de microsatellites seulement étaient utilisées dans le meilleur des cas.

Cependant, les puces à SNP* comportant une densité aussi élevée de marqueurs ont un coût élevé. Ainsi, même pour des puces ne comportant que 6 000 marqueurs, le coût reste important et le nombre minimum d'individus à étudier est trop élevé (de l'ordre de 1 200) pour l'ordre de grandeur des projets de génétique en cours chez le canard. De ce fait, les marqueurs microsatellites restent parfois encore une bonne option, car utilisés en faible nombre, ils permettent de mettre en œuvre rapidement et à relativement bas coût, une étude restreinte à une région spécifique. La seule restriction à l'utilisation des microsatellites concerne le développement préalable des marqueurs qui nécessite des données de séquence. La disponibilité de la séquence du génome du canard a maintenant permis d'accéder directement à plus de 6 000 marqueurs microsatellites potentiellement très polymorphes, directement utilisables pour une étude ciblée. Une carte du chromosome APL12*, nécessaire à une étude QTL en cours (cf. § 2.3) et faisant appel à 19 marqueurs microsatellites, a ainsi pu être réalisée très récemment.

2.2 / Dispositif de recherche de QTL et nouvelle carte génétique*

Le dispositif de recherche de QTL GeneCan consiste en 7 familles de pères F1, permettant d'étudier la ségrégation de QTL et de marqueurs génétiques chez le canard commun. Les individus dont on veut étudier les performances sont les canes communes backcross, d'une part pour leurs performances propres, mais aussi surtout en tant que mères de mulards. Afin d'évaluer ces caractères maternels, chaque femelle backcross a été inséminée avec de la semence d'un mâle de Barbarie pour réaliser un test sur sa descendance mulard (figure 5). Les premiers résultats de détection de QTL ont déjà été publiés (Kileh-Wais *et al* 2013), mais comme la carte génétique*

Encadré 2. QTL, eQTL, pQTL, mQTL.

Les caractères agronomiques finaux d'intérêt pour les filières sont des phénotypes dont la définition est assez vague d'un point de vue biologique. Pour la filière des palmipèdes à foie gras, il s'agit par exemple de l'indice de consommation ou de la qualité du foie gras d'un point de vue organoleptique. Comme la sélection génétique nécessite de pouvoir utiliser des quantités mesurables afin de construire des index et que les caractères traités ne sont pas toujours aussi simples que des poids d'animaux ou leur taille, des analyses plus fines sont souvent réalisées, telles que la mesure du pH ultime ou de la vitesse de décroissance du pH *post-mortem*, comme indicateurs de la qualité de la viande. De fait, de très nombreux phénotypes intermédiaires entre le génome et les phénotypes d'intérêt peuvent être mesurés, chacun ayant une interaction d'ordre différent avec l'environnement. Ces phénotypes intermédiaires - niveaux de métabolites, de protéines, de transcrits de gènes - pourraient aussi servir de marqueurs en vue de la sélection génétique (Lagarrigue et Tixier-Boichard 2011).

Les recherches de QTL visent à définir des régions du génome gouvernant des caractères d'intérêt. Pour un caractère donné, l'idée est de mettre en évidence un ou plusieurs polymorphismes de l'ADN, permettant de réaliser une sélection assistée par marqueur. Comme pour la sélection en génétique quantitative, l'analyse des phénotypes intermédiaires peut aider dans ces recherches. On parlera donc de eQTL pour des QTL d'expression de gènes (niveaux de transcrits), de pQTL pour les protéines et de mQTL pour les métabolites.

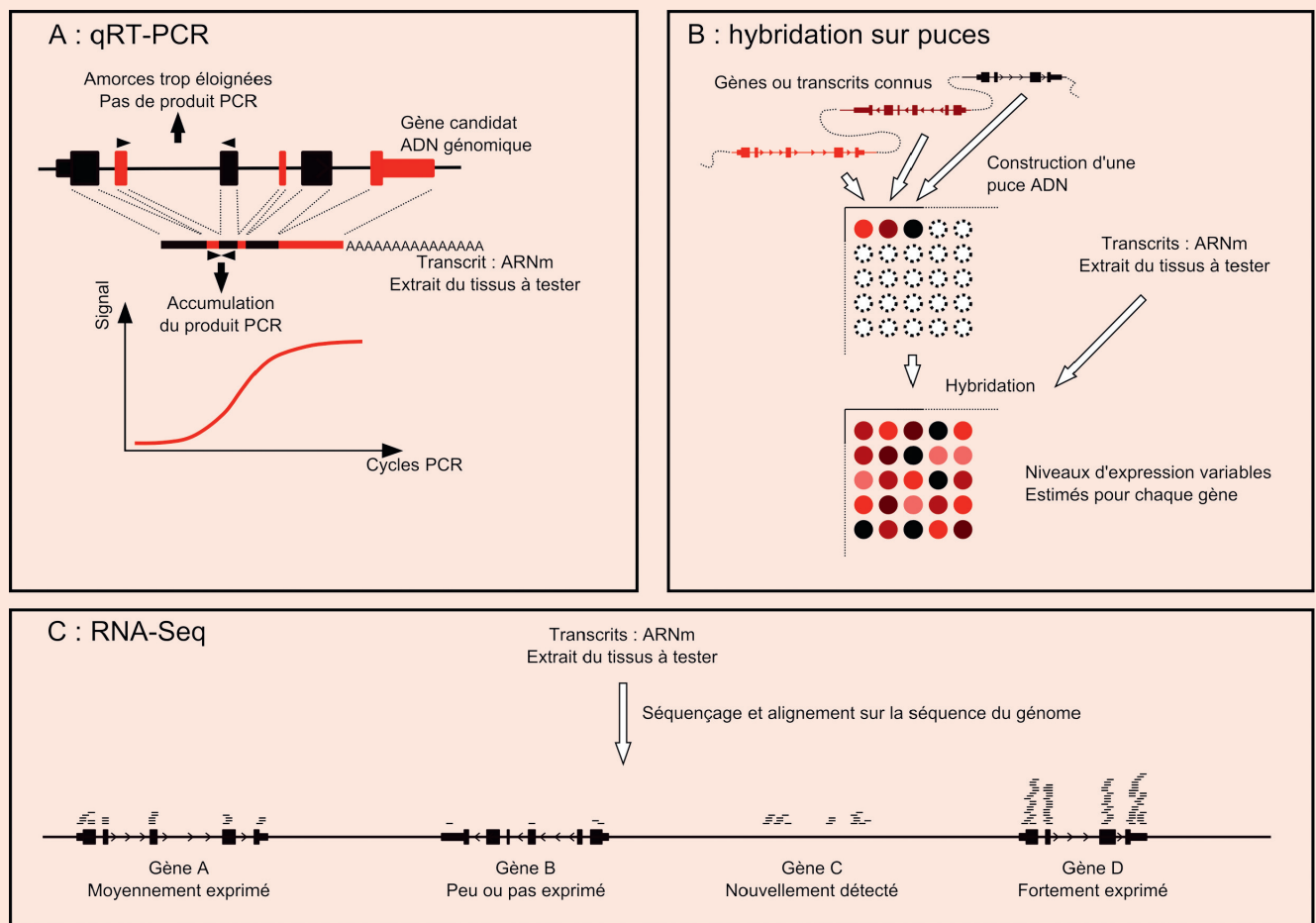
utilisée ne couvrirait que très partiellement le génome (moins de la moitié), une nouvelle détection de QTL sera entreprise lorsqu'une carte génétique* de qualité sera disponible.

Pour développer cette carte, la stratégie consiste à séquencer les mâles F1 du dispositif QTL, qui doivent être hétérozygotes pour qu'un marqueur soit informatif, et à aligner leur séquence sur le génome de référence. Près de 10 millions de SNP* ont été ainsi détectés, dont environ 100 000 hétérozygotes pour les 7 mâles F1 du dispositif GeneCan et donc informatifs pour les 7 familles. Faute d'assemblage de la séquence du génome en chromosomes, les marqueurs ont été choisis en fonction de leur position prédite par leur positionnement sur le génome de la poule en utilisant la similarité de séquence (Vignal *et al* 2013). La réalisation d'une carte génétique* comportant plus de 350 marqueurs SNP*, choisis de manière à optimiser la couverture du génome, est prévue à très court terme. Cette carte étant construite à partir du dispositif de détection de QTL GeneCan (Kileh-Wais *et al* 2013), la détection de QTL, eQTL et pQTL (encadré 2) pourra être réalisée dans des conditions optimales (François *et al* 2013).

2.3 / Premiers résultats de détection de QTL

Dans le cadre du dispositif animal décrit ci-dessus et à l'aide de la carte génétique composée de 91 marqueurs microsatellites répartis en 16 groupes de liaison, une première détection de QTL (Kileh-Wais *et al* 2013) a été réalisée sur un grand nombre de caractères recueillis sur les mulards, allant de leur croissance à leur métabolisme hépatique en début, milieu et fin de gavage, en

passant par leur aptitude au gavage et à la qualité de leurs produits (foie gras et magret). L'analyse caractère par caractère a mis en évidence 22 QTL significatifs à 1% au niveau du chromosome, auxquels 52 QTL putatifs (à 5% au niveau du chromosome) peuvent être ajoutés. Rapporté au nombre de caractères mesurés, il y a très peu de QTL détectés concernant la croissance ou le métabolisme hépatique des mulards. En effet, si 2 zones chromosomiques de co-localisation de QTL apparaissent sur APL2* et APL3* pour le poids des animaux, les QTL relatifs au métabolisme sont faiblement significatifs et repartis sur 12 chromosomes distincts. Par rapport à la quantité de caractères enregistrés, beaucoup de QTL d'aptitude au gavage et de qualité des foies et magrets ont été identifiés : sur un total de 34 QTL, 8 sont liés aux poids de la carcasse et de ses différents morceaux, dont 3 QTL liés au poids de foie gras (significatifs à 1% au niveau du chromosome sur APL2*, APL9* et APL20*) et 2 QTL liés au poids du muscle du magret (sur APL2* et APL19*). Étonnamment, aucun QTL de gras abdominal ou de poids de la peau des magrets n'a été détecté, alors que ces caractères d'engraissement périphérique présentent de nombreux QTL chez le poulet (Nadaf *et al* 2009). Enfin, sur les caractères de qualité des produits, 14 QTL sont associés à la qualité des foies gras et 12 QTL à la qualité des magrets. Parmi ces derniers, on soulignera le chromosome APL2*, dont des marqueurs sont associés à de fortes variations du taux de lipides intramusculaires et des pertes à la cuisson, et le chromosome APL6*, qui modifie le taux de pertes à la cuisson, et qui avait déjà été repéré sur du muscle de canard Pékin non gavé (Huang *et al* 2007). Pour les caractères phares

Encadré 3. *qRT-PCR, puces et RNA-seq pour les études d'expression de gènes.*

Les études d'expression de gènes consistent à déterminer le niveau de leur transcrits (ARN) dans des conditions différentes, des variations de niveaux pouvant être dues à l'environnement, à la génétique ou plus simplement à des comparaisons entre tissus d'un même animal. Trois familles de méthodes peuvent être utilisées : la PCR quantitative pour travailler gène à gène (A), l'utilisation de puces d'expression (B) ou plus récemment le RNA-Seq (C), dont l'essor est lié à celui des techniques de séquençage de nouvelle génération.

de la filière, à savoir la qualité du foie gras, 2 groupes de co-localisation de QTL se dégagent : d'une part, 5 QTL associés aux taux de lipides, de protéines et de collagène des foies, à leur indice de jaune et leurs taux de fonte se situent sur APL2* et d'autre part, 4 QTL associés à l'indice de rouge, aux taux de lipides, de protéines et de fonte apparaissent sur APL9*.

Les zones du génome potentiellement intéressantes mises en lumière par cette première analyse, doivent maintenant être étudiées d'autres manières pour tenter de cerner le ou les gènes qui influencent les caractères d'intérêt. Pour les caractères de qualité de foie, nous avons choisi d'utiliser des outils de transcriptomique et de protéomique pour identifier les mécanismes sous-jacents à ces QTLs, dans le cadre d'une thèse en cours (François *et al* 2013). Enfin, la carte utilisée dans cette détection étant très imparfaite, de nouvelles détections seront mises en œuvre avec la carte SNP* prochainement,

afin de poursuivre le balisage du génome du canard.

3 / Génomique fonctionnelle du canard commun

3.1 / Etude du transcriptome

Des progrès indéniables ont été réalisés en génomique fonctionnelle mais, tout comme en génétique, les stratégies doivent être adaptées aux moyens et outils disponibles. L'approche récente de RNA-Seq (encadré 3) est très prometteuse, mais pour des raisons de coûts elle ne peut être réalisée pour le moment que sur un nombre restreint d'échantillons. Pour des études nécessitant un plus grand nombre d'échantillons, telles que celles conduites en génétique pour la recherche de eQTL (QTL d'expression génique), cette approche par séquençage est donc pour le moment exclue. L'utilisation de puces d'expression semble donc nettement préférable, bien qu'encore

couteuse pour de si grands nombres d'échantillons. L'idée qui prévaut encore actuellement est donc de s'orienter sur l'analyse de l'expression par RT-PCR quantitative en temps réel (qRT-PCR), la miniaturisation des volumes réactionnels ayant permis de réduire les coûts. Une telle étude est en cours, concernant l'analyse de l'expression d'une centaine de gènes sur plusieurs centaines d'individus avec comme objectif de réaliser des détections de eQTL à partir du dispositif GeneCan précédemment décrit (François *et al* 2013). L'inconvénient de l'approche par qRT-PCR est que l'on doit définir a priori les gènes d'intérêt. C'est ce qui a été fait pour les travaux de François *et al* (2013) à partir de la littérature mais aussi à partir de notre propre expérience sur le foie. Pour des études pour lesquelles un nombre plus restreint d'individus (de l'ordre de 20 à 30) est possible, nous envisageons de concevoir une puce d'expression spécifique au canard. L'une des difficultés à laquelle nous serons certaine-

ment confrontés sera l'annotation fonctionnelle* incomplète du génome. Pour pallier cette difficulté, nous pourrions nous appuyer sur le génome de la poule, qui est mieux annoté.

Finalement, un projet en cours vise à analyser le transcriptome hépatique par RNA-Seq (séquençage d'ARN, pour estimer le niveau d'expression de l'ensemble des gènes après alignement sur la séquence de référence) des deux espèces de canards et de leurs produits hybrides (mulards, issus du croisement entre une mère cane commune et un père canard de Barbarie et hinnies, issus du croisement réciproque et non utilisés pour la production de foie gras) dans deux conditions : gavés et non gavés (Diot *et al* 2013).

3.2 / Etude du protéome

L'étude des profils d'expression protéique dans les tissus d'intérêt (principalement le foie) revêt un intérêt particulier pour la compréhension des mécanismes physiologiques. L'étape limitante de ces analyses reste l'identification des protéines par spectrométrie de masse. En effet, lorsque les bases de données utilisées pour effectuer les identifications sont NCBI Inr Aves (<http://www.ncbi.nlm.nih.gov>) et SwissProt Chordata (<http://www.uniprot.org>), seuls 17 spots issus de gels d'électrophorèse bidimensionnelle obtenus à partir de foie gras de canard mulard sont identifiés parmi 31 spots différenciellement exprimés entre différents stades de la stéatose hépatique (Bax *et al* 2012). Ce faible taux d'identification (55%) des protéines chez le canard mulard s'explique par une moindre annotation des génomes de canard et conduit par conséquent à l'obligation de travailler par homologie de séquences avec d'autres espèces animales telle que la poule. Ainsi, dans l'étude de Bax *et al* (2012), la presque totalité des protéines a été identifiée grâce à l'annotation du génome de la poule (sauf 2 spots identifiés à partir du porc et du bovin). En revanche, dès lors que les bases de données précédemment citées sont complétées par une banque d'EST* (« *Expressed Sequence Tags* ») de canard commun, le taux de succès de l'identification est nettement amélioré. Lors d'une étude portant sur la variabilité de la perte à la cuisson, 47 protéines parmi 70 (soit 67%) différenciellement exprimées dans du foie gras de canard mulard ont pu être identifiées grâce à l'utilisation additionnelle de la banque d'EST*. Ainsi, 28% des protéines identifiées l'ont été dans l'espèce *Anas platyrhynchos*, 47% dans l'espèce *Gallus gallus* et 25% avec d'autres espèces (Théron *et al* 2011).

4 / Perspectives et applications

4.1 / Gestion des populations

Un premier jeu de marqueurs microsatellites a été défini pour de l'assignation de parenté à la fois chez le canard commun et le canard de Barbarie (Chapuis *et al* 2010). Cependant, il n'avait pu être réalisé qu'à partir d'un nombre restreint de marqueurs disponibles à l'époque et nécessitait de tester au préalable tous les marqueurs dans les deux espèces. Le séquençage en cours de canards de Barbarie va permettre de repérer les marqueurs microsatellites communs aux deux espèces, commune et Barbarie, et donc d'améliorer ce panel. De plus, ce séquençage de canards de Barbarie va aussi permettre la détection de SNP*, en référence à l'assemblage de la séquence du canard commun. La limite de ces recherches de marqueurs génétiques chez le canard de Barbarie vient du fait qu'un nombre limité d'individus a été séquencé et surtout que la détection est faite en alignant sur la séquence du canard commun. Cette difficulté pourrait être levée par la réalisation de la séquence du canard de Barbarie.

4.2 / Connaissance des génomes

L'analyse des résultats du programme de RNA-Seq impliquant les deux espèces et leurs hybrides permettra d'avoir d'une part, une première vue exhaustive des variations de l'expression génique entre le canard commun, le canard de Barbarie et leurs descendants hybrides et, d'autre part, une bonne idée de la divergence nucléotidique (SNP*) des régions codantes et de structure primaire des protéines. L'analyse de ces données devrait également être riche d'enseignements sur le fonctionnement du génome des deux hybrides réciproques, le mulard et le hinny.

Ces travaux sur le canard s'inscrivent également dans le cadre d'un projet international qui porte sur l'étude de l'évolution de génomes d'oiseaux impliquant le séquençage de près de 50 espèces différentes. Dans ce contexte, le génome du canard commun, récemment séquencé et assemblé avec la contribution de l'INRA, permettra de compléter la liste des génomes d'oiseaux pour lesquels on dispose d'un assemblage en chromosomes du génome. Ce nombre est actuellement de trois pour lesquelles l'assemblage est publié et disponible : poule, dinde, diamant mandarin et

devrait augmenter bientôt quoique assez peu, car la plupart des espèces séquencées ne bénéficient pas de cartes du génome. Ces travaux vont donc contribuer à l'étude de l'évolution des caryotypes chez les oiseaux.

4.3 / Liens entre phénotype et génotype

Les connaissances accumulées en génomique sont bien plus importantes pour le canard commun que pour le canard de Barbarie. Cependant, les outils proposés par la génomique n'ont d'intérêt pour l'analyse des caractères que s'il existe en regard des dispositifs expérimentaux sur les animaux correspondants, souvent lourds à mettre en place. En matière de recherche de QTL, le seul dispositif existant ne permet d'aborder que la voie du canard commun pour les caractères exprimés chez le mulard. Cependant, il n'est pas inconcevable de considérer que le prix du génotypage à haute densité puisse diminuer dans un avenir proche, si bien que les recherches de zones du génome impliquées dans la variabilité génétique d'un caractère pourraient être réalisées par tests d'association directement dans des populations en sélection.

Conclusion

Bien que ne disposant pas des moyens pouvant être mis en œuvre pour les espèces majeures, le canard commence néanmoins à bénéficier, avec un décalage dans le temps, des progrès réalisés sur les coûts et la rapidité des technologies de la génomique et notamment du séquençage. Un point important à souligner est que la profusion de données générées par les technologies à haut débit est à mettre en regard des problèmes nouveaux posés par leur analyse, qui devient souvent l'étape critique. L'acquisition des données phénotypiques est également un point à ne pas négliger.

Pour la génomique, en plus de la difficulté liée au décalage technologique, une difficulté additionnelle, liée à la filière foie gras, est due au fait que le produit final exploité est le fruit d'un croisement interspécifique, nécessitant une connaissance approfondie des génomes des deux espèces parentes. Les avancées technologiques permettent toutefois d'envisager sereinement l'avenir des études réalisées sur ce modèle de croisement original et particulièrement intéressant, tant du point de vue de la filière que de l'avancée des questions scientifiques plus académiques.

Glossaire

Annotation structurale : repérage des coordonnées des diverses structures dans le génome, telles que les gènes.

Annotation fonctionnelle : renseignements sur les fonctions des séquences, le plus souvent pour les gènes.

APL : *Anas platyrhynchos*. Préfixe utilisé pour désigner un chromosome de canard.

BAC : Bacterial Artificial Chromosome. Vecteur de clonage permettant l'obtention de clones bactériens contenant un grand fragment d'ADN génomique (taille > 100 kb*). Les BAC assemblés en contigs* sont à la base des cartes physiques du génome.

Carte cytogénétique : carte des chromosomes. Réalisée par localisation visuelle (FISH*) au microscope de fragments d'ADN sur les chromosomes au stade métaphase de la mitose.

Carte d'hybrides irradiés : réalisée en testant par PCR ou par séquençage, la présence ou l'absence de fragments d'ADN dans une collection de clones d'hybrides irradiés (RH)*. Deux fragments d'ADN sont proches sur le génome s'ils sont trouvés fréquemment dans les mêmes clones.

Carte génétique : obtenue par l'étude de la ségrégation dans des familles ou des populations, de marqueurs polymorphes, soit moléculaires, soit phénotypiques, deux séquences étant d'autant plus proches qu'elles sont souvent transmises ensemble lors de la méiose.

Contig : ensemble de lectures de séquences ordonnées grâce à des informations sur leur parties chevauchantes.

cM : centiMorgan. Unité de distance en cartographie génétique.

cR : centiRay. Unité de distance en cartographie d'hybrides irradiés.

EST : Expressed Sequence Tag : séquences étiquettes (partielles) de transcrit, obtenues par séquençage aléatoire d'ARN.

FISH : Fluorescent *In Situ* Hybridisation. Hybridation de sondes d'ADN marquées à l'aide d'un fluorochrome, sur des chromosomes au stade métaphase de la mitose. Permet la réalisation de la carte cytogénétique*.

Fosmide : vecteur de clonage permettant l'obtention de clones bactériens contenant des fragments d'ADN génomique de taille déterminée et égale à 40 kb*.

Gb : gigabase ; séquence d'un million de paires de bases (pb*) de longueur.

Gene Ontology : projet de bioinformatique visant à standardiser la représentation et les attributs des gènes et de leurs produits à travers les espèces et les bases de données.

GGA : *Gallus gallus*. Préfixe utilisé pour désigner un chromosome de poule.

Hybride irradié : cellule hybride obtenue par fusion entre cellules hôte d'une espèce et donneuse d'une autre espèce, contenant une fraction aléatoire du génome de l'espèce donneuse, après cassures par irradiation, reconstitution aléatoire de chromosomes ou insertion dans des chromosomes de la cellule hôte et rétention partielle. Deux séquences proches sur le génome sont en probabilité dans les mêmes clones RH*, tandis que deux séquences distantes ont une probabilité faible d'être conservées ensemble.

Kb : kilobase ; séquence de mille paires de bases (pb*) de longueur.

Mate-pair : séquences appariées (1 à 10 kb* de distance), produites en circularisant les fragments d'ADN, puis par séquençage à travers le point de jointure.

Mb : mégabase ; séquence d'un million de paires de bases (pb*) de longueur.

Orthologues : deux gènes présents dans deux espèces différentes et ayant un ancêtre commun.

Paired end : séquences appariées produites par la lecture des deux extrémités de courts fragments d'ADN (moins de 500 pb*) dans le cas des nouvelles technologies de séquençage.

Pb : paire de base ; unité de séquence d'ADN, représentée par une base et sa complémentaire-inverse sur l'autre brin.

QTL : Quantitative Trait Locus. Région du génome gouvernant un caractère quantitatif.

RH : Radiation Hybrid (hybride irradié*).

Sanger (méthode de) : Méthode de séquençage publiée en 1977 et encore utilisée de nos jours avec les séquenceurs à électrophorèse capillaire.

Scaffold : ensemble de contigs* de séquence reliés entre eux par des informations apportées par des lectures appariées (mate-pairs* ou paired-ends*).

SNP : Single Nucleotide Polymorphism ; polymorphisme ponctuel de nucléotide : substitution d'une base par une autre dans une séquence.

Références

- Bax M.L., Chambon C., Marty-Gasset N., Remignon H., Fernandez X., Molette C., 2012. Proteomic profile evolution during steatosis development in ducks. *Poult. Sci.*, 91, 112-20.
- Chapuis H., Faugas R., Rossignol M.N., Feve K., Genestout L., Chantry-Darmon C., Guemene D., 2010. Vers l'utilisation de marqueurs moléculaires pour l'assignation des parents en canard grâce au programme Apache. 9^{èmes} Journ. Rech. Palmipèdes à Foie Gras, Bordeaux, France, 1-4.
- Diot C., Houée-Bigot M., Demeure O., Baéza E., Vignal A., Pitel F., Marie-Etancelin C., Robert-Granié C., Molette C., Bouchez O., Esquerré D., Marsaud N., Klopp C., Peterlongo P., Lemaître C., 2013. Analyse de transcriptomes de foies de canards par séquençage d'ARN (RNASeq). 10^{èmes} Journ. Rech. Palmipèdes à Foie Gras, La Rochelle, France, 562-565.
- Fillon V., Vignoles M., Crooijmans R.P., Groenen M.A., Zoorob R., Vignal A., 2007. FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes. *Anim. Genet.*, 38, 303-307.
- François Y., Marie-Etancelin C., Molette C., Vignal A., Davail S., 2013. Optimisation d'une population expérimentale de canes communes (*Anas platyrhynchos*) et de leurs descendants mulards : utilisation de données de QTL. 10^{èmes} Journ. Rech. Palmipèdes à Foie Gras, La Rochelle, France, 1-5.
- Gonzalez J., Düttmann H., Wink M., 2009. Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *J. Zool.*, 279, 310-318.
- Groenen M.A.M., Wahlberg P., Foglio M., Cheng H.H., Megens H.J., Crooijmans R.P., Besnier F., Lathrop M., Muir W.M., Wong G.K., Gut I., Andersson L., 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome. Res.*, 19, 510-519.
- Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A.M., Delany M.E., Dodgson J.B. et al, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432, 695-716.
- Huang C.W., Cheng Y.S., Rouvier R., Yang K.T., Wu C.P., Huang H.L., Huang M.C., 2009. Duck (*Anas platyrhynchos*) linkage mapping by AFLP fingerprinting. *Genet. Select. Evol.*, 41, 28.
- Huang Y., Zhao Y., Haley C.S., Hu S., Hao J., Wu C., Li N., 2006. A Genetic and Cytogenetic Map for the Duck (*Anas platyrhynchos*). *Genetics*, 173, 287-296.
- Huang Y., Haley C.S., Wu F., Hu S., Hao J., Wu C., Li N., 2007. Genetic mapping of quantitative trait loci affecting carcass and meat quality traits in Beijing ducks (*Anas platyrhynchos*). *Anim. Genet.*, 38, 114-119.
- Huang Y., Li Y., Burt D.W., Chen H., Zhang Y., Qian W., Kim H., Gan S., Zhao Y., Li J., Yi K., Feng H., Zhu P., Li B., Liu Q., Fairley S., Magor K.E., Du Z., Hu X., Goodman L., Tafer H., Vignal A., Lee T., Kim K.W., Sheng Z., An Z., Searle S., Herrero J., Groenen M.A.M., Crooijmans R.P.M.A., Faraut T., Cai Q., Webster R.G., Aldridge J.R., Warren W.C., Bartschat S., Kehr S., Marz M., Stadler P., Smith J., Kraus R.H.S., Zhao Y., Ren L., Fei J., Morisson M., Kaiser P., Griffin D.K., Rao M., Pitel F., Wang J., Li N., 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genet.*, 45, 776-788.
- Kayang B.B., Fillon V., Inoue-Murayama M., Miwa M., Leroux S., Feve K., Monvoisin J.L., Pitel F., Vignoles M., Moullhayrat C., Beaumont C., Ito S., Minvielle F., Vignal A., 2006. Integrated maps in quail (*Coturnix japonica*) confirm the high degree of synteny conservation

with chicken (*Gallus gallus*) despite 35 million years of divergence. *BMC Genomics*, 7, 101.

Kileh-Wais M., Elsen J.M., Vignal A., Feves K., Vignoles F., Fernandez X., Manse H., Davail S., Andre J.M., Bastianelli D., Bonnal L., Filangi O., Baéza E., Guemene D., Genet C., Bernadet M.D., Dubos F., Marie-Etancelin C., 2013. Detection of QTL controlling metabolism, meat quality, and liver quality traits of the overfed interspecific hybrid mule duck. *J. Anim. Sci.*, 91, 588-604.

Kranis A., Gheyas A.A., Boschiero C., Turner F., Yu L., Smith S., Talbot R., Pirani A., Brew F., Kaiser P., Hocking P.M., Fife M., Salmon N., Fulton J., Strom T.M., Haberer G., Weigend S., Preisinger R., Gholami M., Qanbari S., Simianer H., Watson K.A., Woolliams J.A., Burt D.W., 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*, 14, 59.

Lagarrigue S., Tixier-Boichard M., 2011. Nouvelles approches de phénotypage pour la sélection animale. In : Numéro spécial, Amélioration Génétique. Mulsant P., Bodin L., Coudurier B., Deretz S., Le Roy P., Quillet E., Perez J.M. (Eds). *INRA Prod. Anim.*, 24, 377-386.

Larzul C., Imbert B., Bernadet M.D., Guy G., Rémignon H., 2006. Meat quality in an intergeneric factorial crossbreeding between Muscovy (*Cairina moschata*) and Pekin (*Anas platyrhynchos*) ducks. *Anim. Res.*, 55, 1-11.

Litt M., Luty J.A., 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.*, 44, 397-401.

Moon D.A., Magor K.E., 2004. Construction and characterization of a fosmid library for

comparative analysis of the duck genome. *Anim. Genet.*, 35, 417-418.

Morisson M., Denis M., Milan D., Klopp C., Leroux S., Bardes S., Pitel F., Vignoles F., Gerus M., Fillon V., Douaud M., Vignal A., 2007. The chicken RH map: current state of progress and microchromosome mapping. *Cytogenet. Genome Res.*, 117, 14-21.

Nadaf J., Pitel F., Gilbert H., Duclos M.J., Vignoles F., Beaumont C., Vignal A., Porter T.E., Cogburn L.A., Aggrey S.E., Simon J., Le Bihan-Duval E., 2009. QTL for several metabolic traits map to loci controlling growth and body composition in an F2 intercross between high- and low-growth chicken lines. *Physiol. Genomics*, 38, 241-249.

Rao M., Morisson M., Faraut T., Bardes S., Feve K., Labarthe E., Fillon V., Huang Y., Li N., Vignal A., 2012. A duck RH panel and its potential for assisting NGS genome assembly. *BMC Genomics*, 13, 513.

Sanger F., Nicklen S., Coulson A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci., USA*, 74, 5463-5467.

Skinner B.M., Robertson L.B., Tempest H.G., Langley E.J., Ioannou D., Fowler K.E., Crooijmans R.P., Hall A.D., Griffin D.K., Volker M., 2009. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics*, 10, 357.

Tai Liu J.J., Tai C. 1991. Mule duck production in Taiwan. I. Artificial insemination of ducks. *Abbrev. Journal*, 328, 1-6.

Théron L., Fernandez X., Marty-Gasset N., Pichereaux C., Rossignol M., Chambon C., Viala D., Astruc T., Molette C., 2011. Identifi-

fication by proteomic analysis of early post-mortem markers involved in the variability in fat loss during cooking of mule duck "foie gras". *J. Agric. Food Chem.*, 59, 12617-12628.

Théron L., Fernandez X., Marty-Gasset N., Chambon C., Viala D., Pichereaux C., Rossignol M., Astruc T., Molette C., 2013. Proteomic analysis of duck fatty liver during post-mortem storage related to the variability of fat loss during cooking of "foie gras". *J. Agric. Food Chem.*, 61, 920-930.

Vignal A., 2011. Etat actuel du séquençage et de la connaissance du génome des espèces animales. In : Numéro spécial, Amélioration Génétique. Mulsant P., Bodin L., Coudurier B., Deretz S., Le Roy P., Quillet E., Perez J.M. (Eds). *INRA Prod. Anim.*, 24, 387-404.

Vignal A., Rué O., Marie-Etancelin C., Klopp C., 2013. Développement de marqueurs SNP chez le canard commun pour la recherche de QTL. *10^{èmes} Journ. Rech. Palmipèdes à Foie Gras*, La Rochelle, France, 591-595.

Weber J.L., May P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, 44, 388-396.

Yuan X., Zhang M., Ruan W., Song C., Ren L., Guo Y., Hu X., Li N., 2006. Construction and characterization of a duck bacterial artificial chromosome library. *Anim. Genet.*, 37, 599-600.

Zhang Y., Zhang X., O'Hare T.H., Payne W.S., Dong J.J., Scheuring C.F., Zhang M., Huang J.J., Lee M.K., Delany M.E., Zhang H.B., Dodgson J.B., 2011. A comparative physical map reveals the pattern of chromosomal evolution between the turkey (*Meleagris gallopavo*) and chicken (*Gallus gallus*) genomes. *BMC Genomics*, 12, 447.

Résumé

La démocratisation des outils de la génomique et plus particulièrement du séquençage à haut débit a permis le séquençage du génome du canard commun. Des projets complémentaires sont déjà initiés pour prolonger et exploiter au mieux ces premiers acquis. En tout premier lieu, il s'agit de poursuivre la description de la structure du génome et d'en exploiter les connaissances : cartes d'hybrides irradiés pour ordonner la séquence le long des chromosomes ; génomique comparée avec le génome de la poule pour bénéficier des connaissances sur ce génome modèle ; recherche de SNP (Single Nucleotide Polymorphism) pour les études et la gestion de populations ; carte génétique pour la détection des QTL (Quantitative Trait Locus). La première détection de QTL influençant les performances du mulard, réalisée à l'aide de marqueurs microsatellites chez la cane commune, sera complétée par une seconde étude utilisant des marqueurs SNP développés spécifiquement et permettant une bien meilleure couverture du génome. Par ailleurs, il est important de réaliser une annotation fonctionnelle du génome, ce qui peut être abordé par le séquençage de transcrits. A terme, le génome annoté sera utilisé pour analyser son expression dans différents tissus et/ou conditions d'élevage, la connaissance des modèles de transcrits et de protéines facilitant les études en transcriptomique et protéomique. Le canard mulard, produit du croisement de la cane commune avec le canard de Barbarie, devra également être étudié en raison de son intérêt agronomique, lié à ses performances exceptionnelles dans la filière des palmipèdes gras.

Abstract

Duck genomics

As DNA sequencing throughputs increase and genomics becomes commonplace, more and more animal species are studied. Thus, the duck genome sequence was published recently. However, although this is an important step and will drastically increase our knowledge on the biology of this species, complementary projects are needed. Some have been initiated, amongst which SNP (Single Nucleotide Polymorphism) detection for population studies and management, radiation hybrid maps for ordering the sequence along chromosomes, genetic maps for QTL (Quantitative Trait Locus) detection and transcript sequencing for structural and functional annotation. The first QTL detection for traits influencing the performances of the mule duck, performed using microsatellite markers in the common duck, will be completed by a second study using specifically designed SNP, allowing for an improved coverage of the genome. Transcript and protein models constructed from the sequences will facilitate proteomics and transcriptomics studies.

The mule duck, the product of a cross between the common duck and the Muscovy duck, has an important agronomic interest for the foie gras industry and will also have to be studied.

VIGNAL A., DIOT C., MOLETTE C., MORISSON M., FARAUT T., RAO M., PITEL F., FILLON V., MARIE-ETANCELIN C., 2013. Génomique des canards. In : Palmipèdes à foie gras. Fortun-Lamothe L. (Ed). Dossier, INRA Prod. Anim., 26, 391-402.