La séquence du génome de la poule et ses applications en sélection

Alain Vignal, B. Besbes

▶ **To cite this version:**

Alain Vignal, B. Besbes. La séquence du génome de la poule et ses applications en sélection. Productions Animales, 2006, 19 (2), pp.109-118. hal-02664615

HAL Id: hal-02664615
https://hal.inrae.fr/hal-02664615

Submitted on 31 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.

INRA Prod. Anim.,
2006, 19 (2), 109-118

La séquence du génome de la poule et ses applications en sélection

A. VIGNAL[1], B. BESBES[2]
[1] INRA, Laboratoire de Génétique Cellulaire, F- 31326 Castanet-Tolosan Cedex
[2] ISA, 5 rue Buffon, F-22000 Saint-Brieuc
Courriel : alain.vignal@toulouse.inra.fr

La poule est la première espèce animale d'intérêt agronomique dont le génome a été séquencé. L'exploitation de données par bioinformatique permet de s'affranchir d'une bonne part des longs travaux de laboratoire nécessaires au développement de marqueurs utilisables en sélection. La connaissance de la position et de la structure des gènes change radicalement les méthodes et les perspectives de recherche en génomique.

Depuis l'avènement des techniques de séquençage de l'ADN vers le milieu des années 1970, la quantité d'information disponible dans les banques de données publiques internationales (GenBank*, EMBL*, DDBJ*) n'a cessé de croître, pour atteindre 100 milliards de bases en août 2005. Plus de 165 000 organismes sont représentés, avec des informations présentes en quantité très variable, allant de quelques fragments de séquence à des génomes complets.

La première séquence complète d'un génome bactérien : *Haemophilus influenza,* a été publiée en 1995, suivie de près en 1996 par le premier organisme eucaryote : la levure *Saccharomyces cerevisiae*, puis le premier organisme multicellulaire en 1998 : le ver nématode *Caenorhabditis elegans*. Depuis lors, la séquence de plusieurs centaines d'organismes a été publiée. Ce sont principalement des bactéries, mais aussi des organismes eucaryotes bien plus complexes comme des plantes supérieures et des vertébrés, au premier rang desquels l'homme dont la publication de la séquence date de 2001 (Lander *et al* 2001, Venter *et al* 2001). Le 1er mars 2004, la séquence du génome de la poule réalisée par une équipe du Washington University Genome Sequencing Center (WUGSC) a été rendue publique (http://www.genome.gov/11510730), puis a été publiée (Hillier *et al* 2004), ce qui constitue une avancée majeure pour l'étude de la biologie et de la génétique de cette espèce.

Afin d'analyser l'apport que représente cette première séquence du génome d'un animal d'intérêt agronomique pour la connaissance de sa biologie et son amélioration génétique, nous allons tout d'abord nous intéresser au contexte général de la génomique. Pour une description plus détaillée des outils de la génomique et de leurs applications, voir le numéro Hors-série «la génétique moléculaire» d'INRA Productions Animales (2000) ou Douaire *et al* (2001).

1 / La génomique et la séquence du génome de la poule

La génomique est une discipline scientifique récente, fruit de l'association des biologies cellulaire et moléculaire avec la génétique classique, dont l'essor tire en outre grandement parti d'une utilisation croissante de l'informatique. Son objet principal, tel qu'il avait été défini en septembre 1987 par Victor A. McKusik et Frank Ruddle dans l'éditorial du premier numéro de la revue Genomics (McKusick et Ruddle 1987), est la cartographie et le séquençage des génomes, ainsi que l'analyse de leur contenu informatif. Le bien-fondé de cette approche ne repose donc pas simplement sur la mise à disposition de cartes ou de séquences nucléotidiques, mais est aussi fortement lié aux tentatives qui sont faites pour leur donner un sens à travers des liens vers la biologie expérimentale et vers des techniques d'analyse informatique. Depuis plusieurs années, d'autres branches de la génomique se sont développées, telles que la génomique fonctionnelle, comparative, etc…

On peut considérer que la séquence d'un génome représente l'aboutissement d'une tâche descriptive, dont le niveau le plus simple de présentation est l'ordonnancement des quatre bases nucléotidiques : adénine (A), cytosine (C), guanine (G) et thymine (T) tout au long de la molécule d'ADN. Cependant, il est bien entendu qu'une vision aussi simpliste n'a pas de sens et que la séquence doit être annotée. Au minimum, les éléments importants, tels que les gènes ou leurs séquences de régulation, doivent être repérés et leurs fonctions déterminées, mais beaucoup d'autres informations utiles peuvent être indiquées, telles que la position de séquences répétées ou de régions de conservation interspécifiques plus ou moins grandes, suggérant une importance fonctionnelle. Un exemple d'information particulièrement utile pour la génétique et l'amélioration des espèces, est le repérage des régions contenant des variations interindividuelles telles que des SNP* ou des marqueurs microsatellites*, qui pourront être utilisés comme marqueurs génétiques.

* Voir glossaire.

1.1 / Génomique de la poule

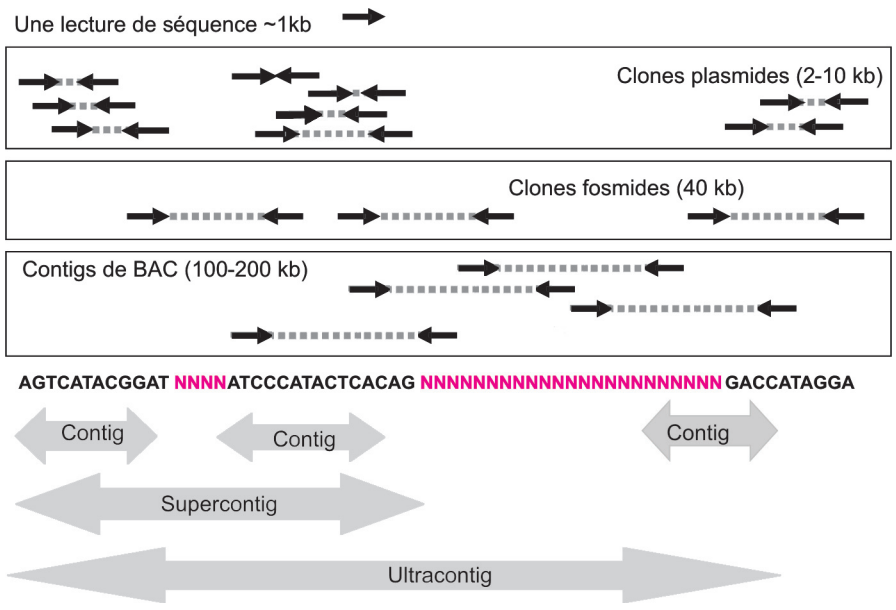
La poule a depuis longtemps été un modèle d'étude en biologie dans des domaines aussi divers que l'immunologie avec les études sur la maturation des lymphocytes B, retracées par Glick (1994) ou l'oncologie, avec des études sur les propriétés tumorigènes de virus initiées par Roux (Calnek 1992). Elle reste un modèle très pertinent pour l'étude de la fonction des gènes et de leurs séquences régulatrices, grâce au modèle des cellules DT40 à fort taux de recombinaison homologue, ou pour l'embryologie, grâce à la facilité d'accès aux embryons dans les œufs (Brown *et al* 2003). En ce qui concerne les études de la structure des gènes et de leur séquence nucléotidique, l'extension aux vertébrés du concept de morcellement des séquences codantes en introns* et exons* a été faite dans le cadre d'une étude sur le gène de l'ovalbumine de poule (Breathnach *et al* 1977). Malgré tous ces apports à la biologie fondamentale, les premières avancées dans le domaine de la génomique chez la poule, sont le fait de laboratoires travaillant dans le domaine agronomique. Ceux-ci ont développé des cartes génétiques dès le début des années 1990 (Bumstead et Palyga 1992), dans le but de localiser sur le génome des régions gouvernant des caractères, souvent quantitatifs, d'intérêts économiques (QTL* : Quantitative Trait Locus ou ETL : Economic Trait Locus). La publication une dizaine d'années plus tard d'une carte consensus contenant plus de 2000 marqueurs (Groenen *et al* 2000), ainsi que les nombreuses publications ayant trait à la localisation de QTL chez la poule ont démontré l'efficacité et l'intérêt de cette approche (Hocking 2005). Afin d'aller plus loin dans la démarche du clonage positionnel, visant à affiner la localisation des QTL jusqu'à l'identification du polymorphisme qui les sous-tend, les mêmes laboratoires se sont intéressés aux autres outils de cartographie : cytogénétique*, physique par contigage* de clones ou cartographie d'hybrides irradiés. Plus récemment, les aspects de la génomique fonctionnelle ont été abordés par le développement de banques et de séquences d'étiquettes de gènes (EST*) en grand nombre, puis la réalisation de macro réseaux* et de micro réseaux* pour l'étude de leur expression. Ces travaux ont été réalisés par des groupes de recherche en agronomie (Carre *et al* 2006), mais aussi dans d'autres domaines de recherche (Hubbard *et al* 2005). L'utilisation de ces nouveaux outils est déjà en cours dans les différents domaines d'application de la génomique de la poule, y compris pour la recherche de gènes différentiellement exprimés entre des lignées de poule divergeant pour des caractères agronomiques tels que la croissance ou l'engraissement (Cogburn *et al* 2003).

Le séquençage du génome de la poule a bénéficié des travaux récents visant à comprendre le fonctionnement du génome humain. En effet, une des approches utilisées pour annoter le génome humain, est de le comparer à d'autres génomes. L'importance d'une portion de séquence peut être déduite de son niveau de conservation interspécifique, ce qui peut alors apporter une aide à la prédiction de gènes ou de séquences régulatrices importantes. La séquence de plusieurs mammifères, ainsi que de plusieurs poissons étant déjà disponible, le séquençage d'un oiseau devenait très informatif de ce point de vue. Ainsi, du fait principalement de son intérêt agronomique, mais aussi de sa position comme modèle en biologie du développement, la poule était la seule espèce d'oiseau pour laquelle des cartes, des banques de clones BAC*, des séquences d'étiquettes EST et autres outils de génomique avaient été développés. Elle a donc été choisie pour représenter cette classe, qui comporte plus de 9000 espèces. C'est pourquoi, trois ans seulement après l'homme et un an avant le bovin, la poule est le premier animal de rente pour lequel la séquence du génome complet ait été réalisée et rendue publique.

1.2 / Stratégie de séquençage utilisée

La stratégie qui a été utilisée pour réaliser la séquence du génome de la poule, tirant partie des enseignements du projet génome humain, est intermédiaire entre l'approche consistant à séquencer séparément des clones BAC choisis d'après leur position sur une carte physique et le séquençage aléatoire global. Elle consiste à associer des données de séquence aléatoire avec une carte physique par contigs* de BAC obtenue par ailleurs. Un seul individu consanguin a été utilisé, afin d'éviter les problèmes d'assemblage de séquence dus au polymorphisme. Son ADN a été fractionné et cloné de manière aléatoire d'une part dans des plasmides* qui acceptent des

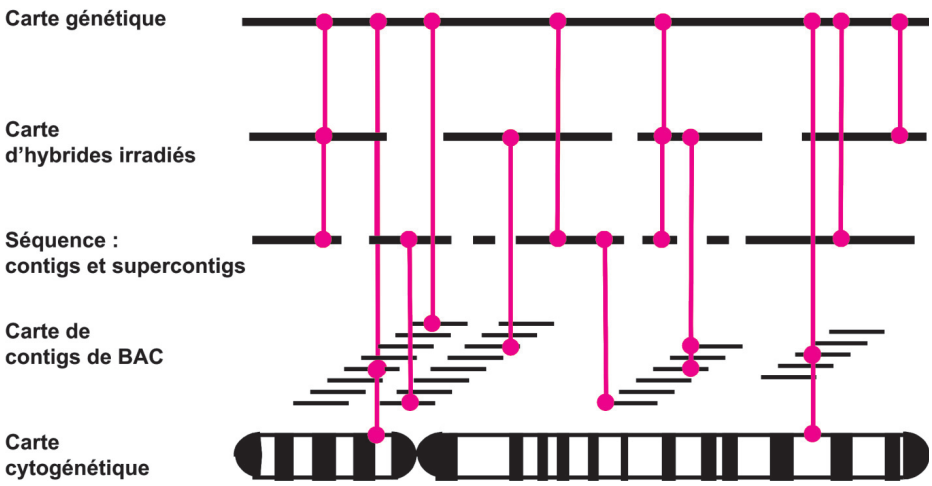
Figure 1. *Principe de l'assemblage de la séquence en contigs, supercontigs et ultracontigs**.



Les lectures de séquence brute (chromatogrammes) obtenues par séquençage des deux extrémités de plasmides* contenant des fragments aléatoires de 2 à 10 kb, sont assemblées par des algorithmes tenant compte de leur qualité. Des ensembles de séquences continues appelés contigs sont ainsi constitués. La profondeur de la séquence représente le nombre de lectures utilisées pour constituer le consensus. En raison de biais statistiques et de clonage, certaines régions sont surreprésentées et d'autres mal couvertes au point de ne pas être représentées du tout. Pour relier les contigs* de séquence entre eux et former des supercontigs*, des séquences d'extrémités de fosmides*, dont la distance est fixée par construction à 40 kb, sont utilisées. L'étape finale d'assemblage consiste à rassembler les supercontigs* en ultracontigs*, en utilisant des informations d'autres types telles que les contigs de BAC*, les cartes génétiques* ou la carte d'hybrides irradiés*.

inserts de taille variant entre 2 et 10 kb* et d'autre part dans des fosmides* qui n'acceptent que des fragments de taille voisine de 40 kb. Les clones ont alors été séquencés par les deux extrémités et plus de 10 millions de lectures (chromatogrammes produits par un séquenceur automatique) d'environ 1 000 bases (1 kb) chacune, appariées 2 à 2, ont été produites. La quantité de séquence obtenue correspond à 6,6 équivalents du génome (profondeur 6,6X). Le choix d'une profondeur de 6X, correspond à un compromis entre le coût et la qualité de la séquence : les lectures ayant été générées aléatoirement, certaines régions sont surreprésentées et d'autres non représentées, sont absentes de l'assemblage final. Les fragments de séquence sont assemblés par des algorithmes informatiques qui tiennent compte de la qualité des lectures et de l'information d'appartenance aux deux extrémités d'un même plasmide, pour construire des contigs (ensemble de lectures chevauchantes donnant une séquence continue), dont la longueur et la profondeur varient en fonction des biais de distribution statistique et de clonage (figure 1). Les contigs de séquence sont liés entre eux grâce aux séquences d'extrémité des grands plasmides et des fosmides, formant ainsi des supercontigs* (figure 1). La taille des inserts dans les fosmides étant connue (40 kb), on a une estimation assez précise de la taille des intervalles de séquence manquante. Finalement, les supercontigs sont orientés et positionnés le long des chromosomes en utilisant les informations d'autres types de cartes, plus particulièrement la carte physique de contigs de BAC (ensembles de clones BAC chevauchants, déterminés par des techniques telles que le fingerprinting*, qui permet d'estimer des similarités de séquence), mais aussi les cartes génétique*, cyto-génétique ou d'hybrides irradiés* (figures 1, 2).

Figure 2. *Intégration des différents types de carte.*



Chaque méthode de cartographie présente des défauts spécifiques et permet d'obtenir une couverture du génome plus ou moins complète. La carte génétique* permet de localiser à la fois des marqueurs moléculaires et des caractères phénotypiques. Une de ses limites provient du fait que pour placer un marqueur, il faut qu'il soit polymorphe. La carte d'hybrides irradiés* permet la localisation de tout marqueur pour lequel on peut développer un test PCR. La carte de contigs* de BAC* a une résolution élevée et est composée de clones de grande taille chevauchants, contenant plusieurs gènes. Un chromosome n'est généralement pas couvert par un contig d'un seul tenant. La carte cytogénétique* a une faible résolution, mais est la seule permettant de relier les données à des chromosomes. Cependant, la localisation par FISH* n'est possible que pour des clones de grande taille, comme les BAC. Tout comme la carte de contigs de BAC, la séquence est composée de fragments (contigs, supercontigs), interrompus par intervalles. Afin de localiser les nombreux supercontigs de la séquence, il est nécessaire de les relier aux autres cartes.

1.3 / Qualité de la séquence

La qualité de la séquence de la poule est la résultante des aléas introduits par les différentes techniques et cartes utilisées pour l'assemblage, mais aussi de la structure même de son génome. Le nombre d'interruptions dans la séquence est encore conséquent et la plus grande longueur de séquence continue (contig le plus long) n'est que de 442 kb (tableau 1). Le taux de séquences répétées plus faible (10 %) que celui habituellement trouvé dans les génomes de mammifères (40-50 %), a permis un assemblage de meilleure qualité en diminuant les problèmes possibles de fusion de séquences appartenant à différents chromosomes. En revanche, des problèmes de couverture de certaines régions du génome par la séquence sont liés à la structure même du caryotype de la poule, composé de macrochromosomes et de microchromosomes comme celui de la plupart des oiseaux. En effet, dix parmi les plus petits microchromosomes n'ont pas de lien avec les autres cartes, faute d'avoir été identifiés en hybridation *in situ* fluorescente (FISH*) par des clones spécifiques de type BAC. Ces dix microchromosomes n'ont par conséquent pas de séquence attribuée (figure 3). Il est cependant vraisemblable qu'une bonne partie de la séquence leur correspondant est présente dans les 120 Mb* de séquence non attribuées à un chromosome précis, regroupées sous l'appellation chrun, pour «chromosome unknown» (figure 3). La couverture du génome est néanmoins bonne, comme le montre le fait que sur 23 212 ARNm et 485 000 EST représentant les séquences transcrites des gènes, 97 % et 96 % respectivement sont retrouvés au moins partiellement dans la séquence (Hillier *et al* 2004).

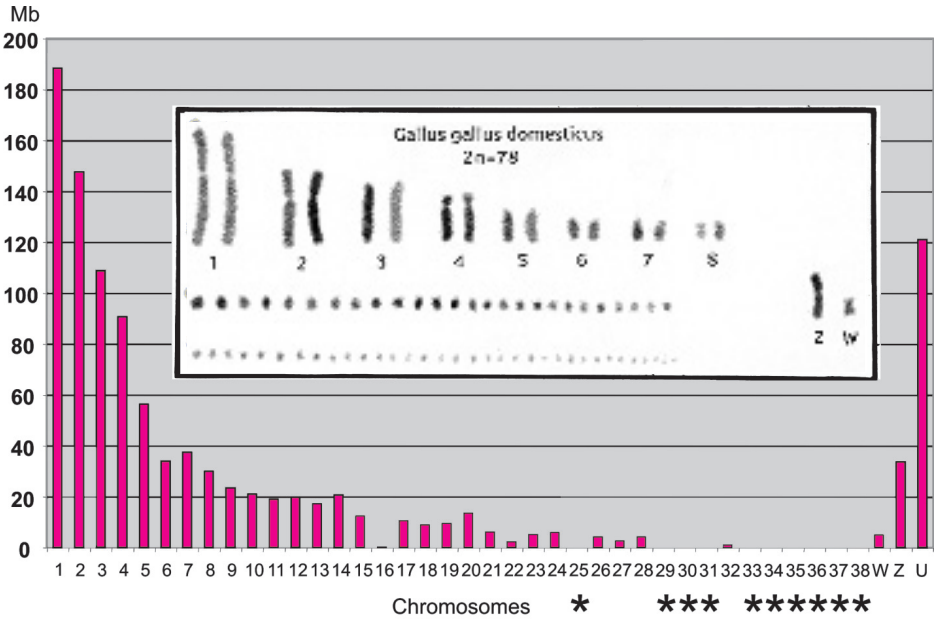
Finalement, lors de l'alignement de cartes d'hybrides irradiés avec la séquence du génome de la poule, nous avons pu établir que si l'assemblage des macrochromosomes ne soulevait pas de problèmes particuliers (Leroux *et al* 2005, Morisson *et al* 2004, Pitel *et*

Tableau 1. *Statistiques sur l'assemblage du génome (Hillier* et al *2004).*

Elément	Nombre > 1 kb[1]	Longueur N50 (kb)[2]	Nombre N50[3]	Plus long (kb)[4]
Contigs	98 612	36	7 468	442
Supercontigs	32 767	7 067	37	33 505

[1] Nombre de contigs de longueur supérieure à 1 kb ; [2] Longueur minimale des contigs pour que 50 % de la séquence du génome soit comprise dans des contigs de taille supérieure ; [3] Nombre de minimum de contigs pour couvrir 50% de la séquence du génome ; [4] Contig et supercontig le plus long. Ces données montrent que la séquence du génome est morcelée : la plus grande longueur de séquence continue (contig*) est de 442 kb et pour couvrir 50 % du génome en prenant les fragments de séquence continue les plus longs, il faut au minimum 7 468 fragments et descendre à une taille de fragment de 36 kb.

Figure 3. *Séquence assignée aux chromosomes de poule.*



Le caryotype montre les chromosomes de poule au stade métaphase de la mitose. L'histogramme présente la longueur de séquence en Mb, assignée aux différents chromosomes. Les étoiles repèrent les microchromosomes pour lesquels aucune séquence n'a pu être assignée. U : séquence "chrun", non attribuée à un chromosome. La longueur de séquence attribuée aux chromosomes est bien proportionnelle à leur taille, sauf pour le chromosome 16 et pour 10 microchromosomes repérés par une étoile. Le chromosome 16 contient les complexes B et Y du complexe majeur d'histocompatibilité et les gènes des ARN ribosomiques, mais seule une partie de la séquence du complexe B est présente dans l'assemblage. Les 10 autres microchromosomes n'ont pas pu être identifiés par FISH sur la carte cytogénétique, empêchant tout lien avec les autres cartes et n'ont donc pas de séquence attribuée.

al 2004), des segments mal placés sur un chromosome ou assignés à un mauvais chromosome peuvent parfois être détectés, en particulier pour certains microchromosomes (Morisson *et al* 2005).

1.4 / Annotations*

Le fait d'avoir à notre disposition la carte ultime du génome de la poule sous la forme de la succession des 4 bases du code génétique, représente à la fois une avancée significative et une porte ouverte vers de nouvelles possibilités d'analyse. L'annotation du génome de la poule est déjà bien avancée et présente des données aussi diverses que des estimations locales de qualité de l'assemblage, la structure de gènes déjà connus ou prédits par des algorithmes informatiques, les zones conservées avec d'autres génomes tels que celui de l'homme et de la souris ou l'existence de polymorphisme utilisable pour des études de génétique.

Les données d'assemblage et d'annotation du génome de la poule peuvent être consultées sur plusieurs sites, dont les plus informatifs sont celui de l'University of California, Santa Cruz (UCSC Genome Browser : http://genome.ucsc.edu/cgi-bin/hgGateway), le système Ensembl de l'European Bioinformatics Institute : http://atlas.cnio.es/Gallus_gallus/ ou le présentateur de cartes du National Center for Biotechnology Information (NCBI Map Viewer : http://www.ncbi.nlm.nih.gov/mapview/). Ces sites, ou navigateurs de génome, présentent l'état actuel de l'assemblage du génome, avec un grand nombre d'informations supplémentaires.

Le premier type d'informations concerne l'assemblage lui-même, avec des données sur la qualité de la séquence consensus. Celle-ci ressort de trois paramètres : (i) l'estimation de la probabilité d'erreur sur la lecture des bases, laquelle dépend principalement de la qualité locale des chromatogrammes résultant des réactions de séquence, encore appelés lectures, et de la profondeur ou nombre de lectures pour chaque base, (ii) la position des trous entre contigs de séquence, (iii) la localisation de clones de type BAC, permettant de relier des contigs entre eux.

Le deuxième type d'informations concerne la structure des gènes. Ces annotations ont deux origines très différentes, expérimentale ou résultant d'une analyse bioinformatique. Les annotations d'origine expérimentale proviennent le plus souvent de données de séquence sur des ARNm résultant d'analyses détaillées de gènes bien connus ou découlant de travaux de séquençage d'un grand nombre d'EST, séquences alignées sur la séquence génomique. Les exons ainsi repérés correspondent donc au mieux à des gènes connus, au pire à des transcrits réels mais inconnus, qui restent donc à étudier plus en détail. Dans le cas d'annotations résultant d'analyses bioinformatiques, les exons repérés sur la séquence génomique sont le résultat d'une estimation effectuée à l'aide de programmes de prédiction de gènes. Il en existe plusieurs, utilisant des méthodes variées et produisant des résultats différents. Certaines méthodes sont basées sur l'analyse des signaux contenus dans la séquence, tels que les cadres ouverts de lecture codant pour les protéines, les signaux d'épissage, etc. D'autres vont intégrer en plus une comparaison de séquences inter espèces, en supposant que les séquences les mieux conservées correspondent à des gènes. D'ailleurs, une bonne part de l'analyse du génome de la poule s'appuie sur les connaissances étendues existant pour le génome humain, de la souris ou de poissons. Chaque algorithme de prédiction de gènes utilisé prédit des faux positifs et des faux négatifs et tout résultat devra être vérifié expérimentalement.

Le troisième type d'informations découle de la comparaison des segments conservés entre la poule et l'homme ou la souris. Ces segments peuvent être de courtes séquences potentiellement importantes d'un point de vue fonctionnel : gènes, séquences de régulation, ARN non codants ou autres. Mais ils peuvent aussi être des segments de plus grande taille, fragments de chromosomes conservés au cours de l'évolution des vertébrés. La connaissance de ces segments conservés permet de se référer au génome de l'homme ou de la souris, dont l'annotation est beaucoup plus avancée. La comparaison des séquences de génomes de vertébrés, montre que le nombre de réarrangements inter chromosomiques est plus faible entre la poule et l'homme, qu'entre la souris et l'homme (Hillier *et al* 2004).

Finalement, un quatrième type d'informations concerne les variations nucléotidiques interindividuelles, telles que les microsatellites ou les SNP. Cette information est très utile pour

le développement de marqueurs génétiques moléculaires supplémentaires.

2 / Applications en biologie et génomique

Du fait d'un accès facile aux données à travers les browsers de génome, une bonne part des travaux de biologie moléculaire autrefois longs et fastidieux, sont remplacés par des analyses «*in silico*».

2.1 / Sélection de gènes candidats, structure des gènes

Avec la séquence, la position dans le génome et la structure d'une grande part des gènes de la poule est maintenant soit connue, soit prédite par analyse bioinformatique. Ceci représente bien entendu un gain important dans la recherche et l'analyse des gènes candidats. Après la primo localisation d'un QTL, on peut dresser une liste aussi complète que possible des gènes de la région déterminée et choisir dans cette première liste ceux présentant une activité connue ou prédite intéressante. Le plus souvent chez la poule, cette fonction sera déduite de celle connue de gènes orthologues* étudiés dans d'autres espèces, le plus souvent l'homme ou la souris. Dans bien des cas, les structures de gènes obtenues par les différents programmes de prédiction ne sont pas cohérentes entre-elles et les comparaisons interspécifiques aideront à se rapprocher de la bonne. Par exemple, l'alignement des séquences de la poule et de l'homme ont permis de proposer pour près de 2000 gènes des départs de traduction en protéine différents des prédictions précédentes faites sur le génome humain (Hillier *et al* 2004). De telles données ont donc une très grande importance pour la prédiction de la structure de ces protéines et par extension de la compréhension de leur activité. Les deux espèces bénéficient donc de ces travaux.

2.2 / Développement de marqueurs génétiques

Pour affiner la position d'un QTL, il est nécessaire de densifier la région en marqueurs génétiques. Ceux-ci peuvent être recherchés directement dans la séquence du génome, où il est possible de faire des requêtes pour trouver des séquences de type microsatellite, qui présentent l'avantage d'être hautement polymorphes. Une fois que tous les marqueurs microsatellites d'une région auront été testés, le développement pourra porter sur les marqueurs de type SNP. Ces derniers sont moins informatifs mais présents en densité plus importante. Ils présentent en plus l'intérêt d'être éventuellement responsables de l'altération qualitative ou quantitative de l'expression d'une protéine.

L'individu choisi pour réaliser la séquence a été une poule de jungle hautement consanguine, afin que son polymorphisme ne cause pas de problèmes lors de l'assemblage. Afin de mettre en évidence un grand nombre de SNP, une recherche spécifique a donc été réalisée par le Bejing Genomics Institute (http://chicken.genomics.org.cn/index.jsp). Pour ce faire, un séquençage partiel aléatoire couvrant 1 fois le génome au total, a été réalisé en utilisant 3 individus génétiquement très différents : une poule pondeuse White Leghorn, une poule de chair Cornish et une poule chinoise de race Silkie. Cette séquence a été alignée sur la séquence génomique assemblée et une analyse, à l'aide d'algorithmes tenant compte de la qualité des séquences et permettant donc de différencier autant que possible erreurs de séquençage de polymorphismes réels, a permis, de mettre en évidence plus de 2,8 millions de SNP positionnés dans le génome (Wong *et al* 2004). Ces données permettent de disposer de SNP candidats localisés et pouvant donc être choisis à l'aide des navigateurs de génome, pour être utilisés dans d'autres populations pour des tests d'association avec des phénotypes.

2.3 / Transposition à d'autres espèces d'oiseaux

Plusieurs tentatives d'utilisation de marqueurs génétiques de poule pour étudier d'autres espèces telles que la dinde (Reed *et al* 2000) ou la caille (Inoue-Murayama *et al* 2001) ont été faites. Cependant, les taux de succès obtenus sont faibles, suggérant qu'au plus 20 % et 17 % de marqueurs microsatellites de poule sont utilisables chez la dinde et la caille respectivement. Devant le nombre trop faible de marqueurs microsatellites de poule disponible pour réaliser des cartes satisfaisantes, des marqueurs spécifiques de la dinde (Burt *et al* 2003) et de la caille (Kayang *et al* 2004) ont été développés et cartographiés. Malgré ces efforts, les cartes obtenues pour ces espèces n'ont qu'une couverture partielle du génome et le développement de marqueurs dans des régions ciblées pour en augmenter la densité, est très lourd.

Le fait de disposer de la séquence de la poule facilite maintenant l'étude d'autres espèces qui nous intéressent comme la caille ou le canard. Des études de cytogénétique ont montré que la conservation de synténie est élevée entre ces trois espèces (Fillon *et al* 2003) et il est donc possible de cibler une région spécifique du génome de la poule, pour développer des marqueurs dans les autres espèces. Un très grand nombre de marqueurs microsatellites choisis dans la séquence peut ainsi être testé et si ils sont en nombre insuffisant, des SNP peuvent être recherchés dans les gènes. Pour ce faire, des amorces sont choisies d'après la séquence du génome de la poule, de préférence dans les exons mieux conservés, pour faciliter l'amplification par PCR. La connaissance de la structure des gènes chez la poule permettra d'éviter les introns ou de cibler ceux dont la taille est suffisamment petite pour être compatible avec une PCR. Dans le cas d'une recherche de polymorphisme, l'amplification d'un petit intron sera recherchée, le taux de variation nucléotidique y étant plus élevé que dans les introns.

2.4 / Connaissance des génomes et intégration des données

La comparaison de la séquence de la poule avec celle d'autres génomes de vertébrés a permis de mettre en évidence des séquences d'ADN aussi bien conservées que des exons, mais ne codant pas pour des protéines. Ces séquences sont parfois regroupées dans des zones pauvres en gènes, appelées déserts de gènes stables (Ovcharenko *et al* 2005) et ont donc vraisemblablement un rôle biologique, souvent de régulation, qui pourra être déterminé dans le futur.

La connaissance non seulement de la structure complète des gènes et de leur position sur le génome, mais aussi de leurs séquences régulatrices ou d'autres séquences d'ADN de fonction encore inconnue, est essentielle pour la mise en relation de variations nucléotidiques avec des phénotypes. En effet, s'il semble naturel de s'intéresser aux modifications de séquence qui altèrent la structure primaire d'une protéine par changement d'acide aminé lorsque l'on veut mettre en évidence des modifications pouvant influencer sur l'expression d'un phénotype, on ne peut jamais exclure l'implication d'autres types de modifications, dans d'autres classes de séquences, qui représentent la plus grande part de la séquence du génome. Plus précisément, les exons de gènes,

seules séquences codant directement pour les acides aminés des protéines, ne représentent que 1 % de la séquence de l'homme. Les introns de gènes en représentent 25 %, le reste étant composé de séquence intergénique. Autre fait marquant : 50 % des génomes de mammifères sont composés de séquences répétées, présentes aussi bien dans les régions intergéniques que dans les introns. Il est donc important de repérer toutes ces classes de séquences, afin de pouvoir s'intéresser en priorité à celles présentant la plus forte probabilité de jouer un rôle dans la variabilité des phénotypes. A titre d'exemple, un SNP situé en dehors d'une séquence codante et expliquant la variation phénotypique d'un QTL, a été décrit pour la croissance musculaire chez le porc (Van Laere *et al* 2003). Ce SNP se trouve dans une région de l'intron 3 du gène *IGF2*, repérable par plusieurs caractéristiques de séquence, dont une richesse élevée en nucléotides (G+C) et une forte conservation entre espèces. La mutation empêche la fixation d'un facteur nucléaire, modulateur du taux d'expression (Van Laere *et al* 2003). D'autres types d'éléments distants des gènes, tels que les LCR (Locus Control Region) peuvent être responsables de variations simultanées de l'expression de plusieurs gènes, à travers des modifications de la structure de la chromatine (Levings et Bungert 2002).

3 / Applications en sélection

Le terme «gène» peut avoir des significations différentes en fonction de l'époque et du contexte dans lequel on l'utilise. On distingue en particulier deux conceptions principales. La conception fonctionnelle, basée sur l'observation de phénotypes chez des individus résultant de croisements est pratiquée depuis les débuts de la génétique classique, tandis que la conception structurale est basée sur la biologie moléculaire et utilise l'ADN comme support. Les applications en sélection animale nécessitent de relier les disciplines de la génétique quantitative et de la biologie moléculaire, qui utilisent chacune l'une de ces deux conceptions.

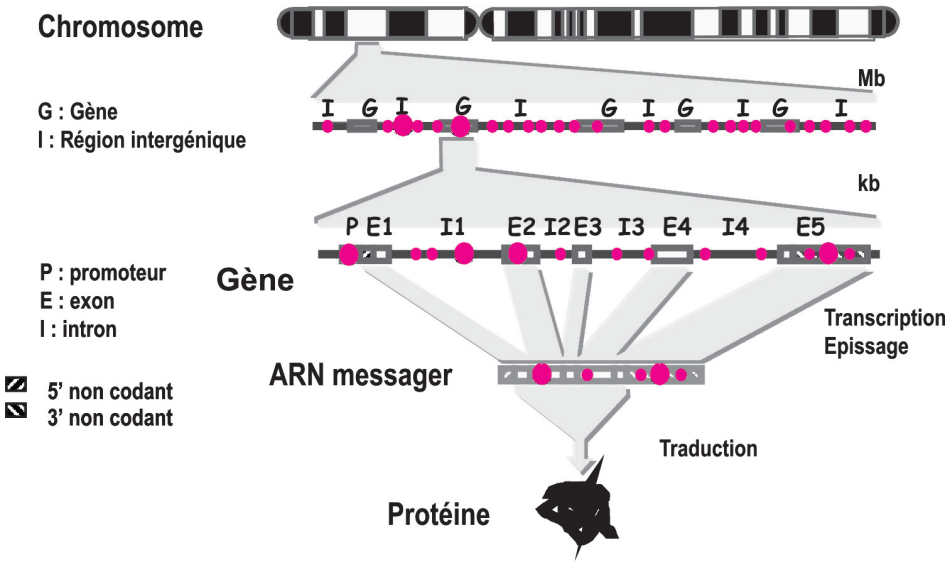
3.1 / Les marqueurs moléculaires à utiliser en sélection

La génétique quantitative est basée sur le modèle polygénique (aussi dit infinitésimal) qui suppose que les caractères quantitatifs sont sous la dépendance d'un grand nombre (strictement parlant infini) de gènes indépendants ayant chacun un faible effet additif. Ce modèle qui continue à dominer le monde de la génétique quantitative est maintenant clairement mis en défaut. En effet, des études réalisées sur la souris (Peripato *et al* 2004), mais aussi dans d'autres espèces animales y compris la poule (Carlborg *et al* 2004, Carlborg *et al* 2003), indiquent que les interactions entre gènes non allèles, ou épistasie, constituent une source importante de variations phénotypiques. La littérature fait mention à présent «d'architecture génétique» des caractères, qui se décrit en terme de nombre, localisation, effet et relations épistatiques des gènes/QTL expliquant la variabilité génétique. Le séquençage du génome de la poule permettra certainement d'améliorer notre compréhension de cette «architecture génétique», en fournissant un support concret sous la forme de la séquence des gènes et de son polymorphisme.

A l'opposé, pour la biologie moléculaire, le gène est une séquence d'ADN qui code pour un produit selon une correspondance linéaire : à la séquence des gènes correspond celle d'ARN messagers, puis celle de la structure primaire de protéines. Ici aussi, le modèle a ses défauts. Par exemple, des régions régulatrices non transcrites distantes des séquences codantes qu'elles régulent, peuvent ségréger indépendamment d'elles *via* le processus de la recombinaison et donc se comporter comme des éléments mendéliens séparés, sans pour autant être des gènes distincts au sens moléculaire. De même, des SNP situés dans deux exons distincts d'un même gène, peuvent provoquer des phénotypes similaires, tout en étant indépendants génétiquement au moins à l'échelle d'une population et en terme de déséquilibre de liaison, du fait d'une distance physique non négligeable.

Quelles que soient les limites de la biologie moléculaire, elle débouche sur la détection de marqueurs génétiques et leur utilisation en sélection. Ces marqueurs correspondent à des variations entre individus (microsatellites, SNP) présentes tout au long du génome, sans que la plupart d'entre elles n'ait d'influence directe sur l'expression des gènes (figure 4). Il est à noter que le polymorphisme à l'origine de variations phénotypiques peut se trouver aussi bien dans des régions codantes (exons) que dans des régions régulatrices (figure 4). Dans certaines conditions, telles qu'en présence d'un fort déséquilibre de liaison ou quand les généalogies sont connues, le polymor-

Figure 4. *Polymorphisme et phénotype*.



Les points rouges représentent le polymorphisme (SNP ou autre) sans influence sur l'expression des gènes et donc sur le phénotype. Les gros points rouges représentent le polymorphisme pouvant entraîner une variation phénotypique à travers une altération de la structure des protéines ou une variation du niveau d'expression des gènes. Le polymorphisme moléculairement neutre semble majoritaire dans le génome. Cependant, il peut servir de marqueurs pour un QTL.

phisme moléculaire neutre peut cependant servir de marqueur pour un QTL.

Une des applications principales de l'utilisation des données de génomique dans le domaine des productions animales, est la Sélection Assistée par Marqueurs (SAM), ou *Marker Assisted Selection* (MAS), qui consiste à utiliser des marqueurs moléculaires pour sélectionner les animaux reproducteurs. Afin de réaliser une SAM, il est nécessaire de disposer de marqueurs positionnés sur le génome le plus près possible du polymorphisme nucléotidique dont on veut sélectionner un allèle. Cependant, l'existence de recombinaisons entre un marqueur et le locus que l'on veut sélectionner peut poser des problèmes importants, allant jusqu'à provoquer l'inverse de l'effet escompté en cas d'inversion de phase. Ceci peut être en partie résolu par l'utilisation de deux marqueurs positionnés de part et d'autre du locus à sélectionner. L'inconvénient de cette approche est que l'on sélectionne alors une large portion d'ADN, pouvant entraîner une diminution importante de la diversité génétique autour du locus dans la population. Les meilleurs marqueurs à utiliser sont donc ceux qui sont directement responsables de la variation phénotypique que l'on veut sélectionner. Il est clair que le fait d'avoir la séquence du génome de la poule va grandement nous aider à obtenir de tels marqueurs, soit directement, soit grâce aux progrès accélérés réalisables en génomique et en clonage positionnel.

3.2 / Les applications pour l'aviculture : exemple de la résistance aux maladies

a) Contexte économique et scientifique

Les applications les plus judicieuses de la SAM sont celles pour lesquelles l'estimation des performances des reproducteurs est difficile, voire impossible à déterminer directement, ou pour lesquelles les mesures sont trop coûteuses. La sélection d'individus d'un sexe sur la base d'un caractère exprimé dans l'autre sexe est, du fait de la caractéristique des oiseaux, moins précise qu'une sélection directe sur les performances de l'individu. C'est le cas par exemple des caractéristiques de ponte, pour lesquelles les mâles sont sélectionnés sur la performance de leurs sœurs. Une meilleure évaluation génétique de ces mâles pourrait être obtenue en incluant l'information de marqueurs moléculaires. De plus, cette évaluation pourrait être réalisée plus précocement sans attendre l'évaluation des sœurs, qui peut parfois être tardive comme c'est le cas par exemple de la mesure de la persistance de la ponte. La sélection sur la résistance aux maladies et la résistance au portage, qui pose des problèmes de sécurité alimentaire, implique des mesures coûteuses et les individus testés ne peuvent être utilisés en reproduction. En effet, les lignées pures sont sélectionnées dans des élevages protégés et bien contrôlés, notamment sur le plan sanitaire. Sauf accident, les animaux ne sont pas en contact avec les agents pathogènes. Par conséquent, la sélection naturelle pour une plus grande résistance ne s'exerce plus. Or, on observe sur le terrain une progression de la virulence de certains agents pathogènes. A ceci, il faut ajouter une demande de plus en plus forte de la part des consommateurs pour une limitation de l'utilisation d'antibiotiques et autres médicaments afin d'éviter la présence de résidus chimiques dans l'alimentation humaine. Pour ces raisons, l'idée de sélectionner pour une plus grande résistance aux maladies regagne du terrain. Cependant, une telle sélection par exposition aux agents pathogènes d'individus choisis dans les lignées commerciales, suivie de la sélection de collatéraux, est très coûteuse et imprécise. L'identification de QTL ou de gènes candidats, apparaît donc comme la seule voie envisageable et économiquement viable pour sélectionner les espèces avicoles pour une plus grande résistance aux maladies.

La plupart des études se sont focalisées sur les maladies engendrant des conséquences sévères. Or, l'appréciation de la sévérité varie dans l'espace et le temps, nécessitant des changements fréquents d'objectifs. Par exemple, la salmonellose considérée auparavant comme une cause de morbidité et de mortalité, apparaît aujourd'hui davantage comme une maladie zoonotique du fait d'une part des progrès sanitaires réalisés, d'autre part de l'importance des produits d'origine avicole dans les contaminations humaines. Le passage de la production d'œufs de la cage au sol (systèmes alternatifs) engendre une recrudescence des contaminations par *Escherichia coli*. Enfin, l'émergence de souches hyper virulentes comme c'est le cas pour le virus de la maladie de Marek, ou nouvelle comme le variant J de la leucose lymphoïde, constitue autant de modèles pour tester la sélection assistée par marqueurs.

Pour le sélectionneur, l'idéal est donc de sélectionner pour une résistance accrue à plusieurs agents pathogènes. Malheureusement, cette idée apparaît de plus en plus comme une utopie. Il semble en effet qu'aux différentes maladies correspondent différents gènes et mécanismes de résistance, au lieu d'un seul mécanisme commun tel qu'une plus grande activité immunologique. Ceci dit, l'influence du complexe majeur d'histocompatibilité (CMH), appelé aussi le complexe *B*, sur la résistance à plusieurs agents pathogènes ou maladies : virus induisant des tumeurs, maladies auto-immunes ou infections microbiennes est bien prouvée (Bacon 1987). L'association entre certains haplotypes du CMH et la résistance à la maladie de Marek est probablement le meilleur exemple connu, et ce parmi l'ensemble des espèces d'animaux de rente. Cependant, le mécanisme sous-jacent à cette résistance reste inconnu et des gènes indépendants du CMH interviennent également (Vallejo *et al* 1998, Yonash *et al* 1999).

b) Conditions pour l'utilisation des marqueurs

La mise en place d'une sélection assistée par marqueurs pour améliorer la résistance aux maladies dépend :

– de l'importance du gène ou QTL dans le contrôle génétique de la résistance et par la suite du progrès à attendre de cette sélection assistée par marqueurs,

– de la corrélation entre l'amélioration observée au sein de la lignée sélectionnée et celle observée sur les produits de croisement impliquant cette lignée,

– de l'effet de cette sélection sur les caractères de production,

– du risque d'une sensibilité accrue vis-à-vis d'autres maladies : il est important de s'assurer que la sélection contre un pathogène spécifique n'engendre pas une sensibilité accrue à d'autres pathogènes,

– de l'interaction entre l'hôte, l'agent pathogène et le vaccin.

Grâce à la protéomique, il est possible d'identifier des gènes de l'hôte qui interagissent avec ceux du pathogène. C'est ainsi qu'on a découvert que l'hormone de croissance (GH) s'associe à une protéine (SORF2) du virus de la maladie de Marek (Liu et al 2001). Ces études peuvent être réalisées avec d'au-

tant plus de facilités qu'en plus de celui de la poule, les génomes de beaucoup d'agents pathogènes impliqués en aviculture ont été séquencés et que les fonctions de leurs gènes sont en cours de caractérisation. Certaines puces produites pour des études de génomique fonctionnelle chez la poule contiennent en plus la séquence de transcrits viraux.

Conclusion

La disponibilité de la séquence du génome de la poule va permettre d'accélérer grandement les connaissances sur la biologie de cette espèce. Un point important à retenir, est que les progrès dans le domaine de la génomique peuvent être très rapides. En effet, alors qu'il y a une dizaine d'années à peine, des cartes génétiques de haute densité n'étaient entrevues qu'à un horizon lointain pour la poule, sa séquence a été obtenue trois ans seulement après celle de l'homme, qui avait elle-même devancé les prévisions. Du fait de sa situation phylogénétique (seul oiseau séquencé) et de son intérêt en biologie du développement, on peut s'attendre à des progrès rapides dans le domaine de l'annotation du génome de la poule et par voie de conséquence dans la compréhension des mécanismes génétiques. La mise au point de marqueurs et leur utilisation en sélection pourrait suivre plus rapidement que prévu en aviculture, grâce à la découverte de SNP directement impliqués dans les mécanismes biologiques d'intérêt, ainsi qu'à la généralisation de méthodes de génotypage rapides et peu coûteuses.

GLOSSAIRE

Annotation : indication sur l'information biologique contenue dans une séquence, allant de la présence de gènes et leur fonction à des caractéristiques de la séquence elle-même, telles que sa conservation entre espèces ou le pourcentage de chaque base.

BAC : Bacterial Artificial Chromosome. Clone bactérien 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 : représente la localisation de fragments d'ADN sur les chromosomes au stade métaphase de la mitose.

Carte d'hybrides irradiés : réalisée en testant par PCR la présence ou l'absence de fragments d'ADN dans une collection de clones cellulaires contenant chacun des fragments différents de chromosomes cassés par irradiation.

Carte génétique : obtenue par l'étude de la ségrégation de marqueurs polymorphes, soit moléculaires, soit phénotypiques.

Contig : ensemble de clones (souvent BAC) ou de lectures de séquence chevauchants.

Contigage : réalisation d'un contig.

DDBJ : DNA Database of Japan. Base de données de séquences nucléotidiques (Japon). Données échangées avec EMBL et GenBank.

EMBL nucleotide sequence database : European Molecular Biology Laboratory. Base de données de séquences nucléotidiques (Europe). Données échangées avec GenBank et DDBJ.

EST : Expressed Sequence Tag. Séquence nucléotidique courte et partielle, généralement obtenue après le séquençage global d'une banque d'ADNc (copie d'ARNm) et pouvant servir d'étiquette pour un gène, soit pour l'étude de son expression, soit pour sa cartographie.

Exon : portion de séquence de gène, présente dans l'ARNm mature, dans laquelle se trouvent les séquences d'ADN traduites en protéines.

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.

Fingerprinting : technique permettant d'estimer très grossièrement la similarité entre des séquences d'ADN sans les séquencer.

Fosmide : vecteur de clonage contenant des fragments d'ADN de 40 kb.

GenBank : base de données de séquences nucléotidiques (USA). Données échangées avec EMBL et DDBJ.

Intron : séquence d'ADN dans un gène, séparant deux exons*.

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

QTL : Quantitative Trait Locus. Portion du génome contenant un polymorphisme impliqué dans la variation d'un caractère quantitatif.

Macro réseaux et Micro réseaux : étiquettes de gènes déposées en grand nombre sur un support solide, afin d'en étudier l'expression par hybridation.

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

Microsatellite : marqueur génétique hautement informatif, dont les allèles correspondent à une variation du nombre de répétitions d'un motif nucléotidique court.

Orthologues : gènes de deux espèces différentes, ayant une origine évolutive commune.

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

Plasmide : vecteur de clonage contenant de courts fragments d'ADN, d'au plus 10 kb.

Promoteur : Séquence d'ADN régulatrice, située en amont d'un gène.

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

Supercontig : ensemble de contigs de séquence reliés par des séquences d'extrémités de clones plasmides ou fosmides.

Ultracontig : ensemble de supercontigs de séquence reliés par des informations de cartographie.

Références

Bacon L.D., 1987. Influence of the major histocompatibility complex on disease resistance and productivity. Poult. Sci., 66, 802-811.

Breathnach R., Mandel J.L., Chambon P., 1977. Ovalbumin gene is split in chicken DNA. Nature, 270, 314-319.

Brown W.R., Hubbard S.J., Tickle C., Wilson S.A., 2003. The chicken as a model for large-scale analysis of vertebrate gene function. Nat. Rev. Genet., 4, 87-98.

Bumstead N., Palyga J., 1992. A preliminary linkage map of the chicken genome. Genomics, 13, 690-697.

Burt D.W., Morrice D.R., Sewalem A., Smith J., Paton I.R., Smith E.J., Bentley J., Hocking P.M., 2003. Preliminary linkage map of the turkey (*Meleagris gallopavo*) based on microsatellite markers. Anim. Genet., 34, 399-409.

Calnek B.W., 1992. Gordon Memorial Lecture. Chicken neoplasia-a model for cancer research. Br. Poult. Sci., 33, 3-16.

Carlborg O., Hocking P.M., Burt D.W., Haley C.S., 2004. Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. Genet. Res., 83, 197-209.

Carlborg O., Kerje S., Schutz K., Jacobsson L., Jensen P., Andersson L., 2003. A global search reveals epistatic interaction between QTL for early growth in the chicken. Genome Res., 13, 413-421.

Carre W., Wang X., Porter T.E., Nys Y., Tang J., Bernberg E., Morgan R., Burnside J., Aggrey S.E., Simon J., Cogburn L.A., 2006. Chicken Genomics Resource: Sequencing and Annotation of 35,407 ESTs from Single and Multiple Tissue cDNA Libraries and CAP3 Assembly of a Chicken Gene Index. Physiol Genomics, sous presse.

Cogburn L.A., Wang X., Carre W., Rejto L., Porter T.E., Aggrey S.E., Simon J., 2003. Systems-wide chicken DNA microarrays, gene expression profiling, and discovery of functional genes. Poult. Sci., 82, 939-951.

Douaire M., Vignal A., Duclos M., Dambrine G., 2001. Les outils de la génomique et leurs applications en sélection avicole. 4èmes Journ. Rech. Avicole, Nantes, France, 27-29 mars. 357-364.

Fillon V., Vignoles M., Garrigues A., Pitel F., Morisson M., Crooijmans R.P., Groenen M.A., Gellin J., Vignal A., 2003. The chicken cytogenetic map: an aid to microchromosome identification and avian comparative cytogenetics. Br. Poult. Sci., 44, 795-797.

Glick B., 1994. The bursa of Fabricius: the evolution of a discovery. Poult. Sci., 73, 979-983.

Groenen M.A.M., Cheng H.H., Bumstead N., Benkel B.F., Briles W.E., Burke T., Burt D.W., Crittenden L.B., Dodgson J., Hillel J., Lamont S., de Leon A.P., Soller M., Takahashi H., Vignal A., 2000. A consensus linkage map of the chicken genome. Genome Res., 10, 137-147.

Hillier L.W., Miller W., Birney E. *et al,* 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature, 432, 695-716.

Hocking P.M., 2005. Review on QTL mapping results in chickens. Worlds Poult. Sci. J., 61, 215-226.

Hubbard S.J., Grafham D.V., Beattie K.J. *et al,* ,2005. Transcriptome analysis for the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags. Genome Res., 15, 174-183.

Inoue-Murayama M., Kayang B.B., Kimura K., Ide H., Nomura A., Takahashi H., Nagamine Y., Takeda T., Hanada H., Tatsuda K., Tsudzuki M., Matsuda Y., Mizutani M., Murayama Y., Ito S., 2001. Chicken microsatellite primers are not efficient markers for Japanese quail. Anim. Genet., 32, 7-11.

INRA Productions Animales, 2000. Génétique moléculaire : principes et application aux populations animales. Numéro hors série. 262p.

Kayang B.B., Vignal A., Inoue-Murayama M., Miwa M., Monvoisin J.L., Ito S., Minvielle F., 2004. A first-generation microsatellite linkage map of the Japanese quail. Anim. Genet., 35, 195-200.

Lander E.S., Linton L.M., Birren B. *et al,* 2001. Initial sequencing and analysis of the human genome. Nature, 409, 860-921.

Leroux S., Dottax M., Bardes S., Vignoles F., Feve K., Pitel F., Morisson M., Vignal A., 2005. Construction of a radiation hybrid map of chicken chromosome 2 and alignment to the chicken draft sequence. BMC Genomics, 6, 12.

Levings P.P., Bungert J., 2002. The human beta-globin locus control region. Eur. J. Biochem., 269, 1589-1599.

Liu H.C., Kung H.J., Fulton J.E., Morgan R.W., Cheng H.H., 2001. Growth hormone interacts with the Marek's disease virus SORF2 protein and is associated with disease resistance in chicken. Proc. Natl. Acad. Sci. USA, 98, 9203-9208.

McKusick V.A., Ruddle F.H., 1987. A new discipline, a new name, a new journal [editorial]. Genomics, 1-2.

Morisson M., Jiguet-Jiglaire C., Leroux S., Faraut T., Bardes S., Feve K., Genet C., Pitel F., Milan D., Vignal A., 2004. Development of a gene-based radiation hybrid map of chicken Chromosome 7 and comparison to human and mouse. Mamm. Genome, 15, 732-739.

Morisson M., Leroux S., Jiguet-Jiglaire C., Assaf S., Pitel F., Lagarrigue S., Bardes S., Feve K., Faraut T., Milan D., Vignal A., 2005. A gene-based radiation hybrid map of chicken microchromosome 14: comparison to human and alignment to the assembled chicken sequence. Genet. Sel. Evol., 37, 229-251.

Ovcharenko I., Loots G.G., Nobrega M.A., Hardison R.C., Miller W., Stubbs L., 2005. Evolution and functional classification of vertebrate gene deserts. Genome Res., 15, 137-145.

Peripato A.C., De Brito R.A., Matioli S.R., Pletscher L.S., Vaughn T.T., Cheverud J.M., 2004. Epistasis affecting litter size in mice. J. Evol. Biol., 17, 593-602.

Pitel F., Abasht B., Morisson M., Crooijmans R.P., Vignoles F., Leroux S., Feve K., Bardes S., Milan D., Lagarrigue S., Groenen M.A., Douaire M., Vignal A., 2004. A high-resolution radiation hybrid map of chicken chromosome 5 and comparison with human chromosomes. BMC Genomics, 5, 66.

Reed K.M., Mendoza K.M., Beattie C.W., 2000. Comparative analysis of microsatellite loci in chicken and turkey. Genome, 43, 796-802.

Vallejo R.L., Bacon L.D., Liu H.C., Witter R.L., Groenen M.A., Hillel J., Cheng H.H., 1998. Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens. Genetics, 148, 349-360.

Van Laere A.S., Nguyen M., Braunschweig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M., Andersson L., 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature, 425, 832-836.

Venter J.C., Adams M.D., Myers E.W. *et al*, 2001. The sequence of the human genome. Science, 291, 1304-1351.

Wong G.K., Liu B., Wang J. *et al*, 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature, 432, 717-722.

Yonash N., Bacon L.D., Witter R.L., Cheng H.H., 1999. High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. Anim. Genet., 30, 126-135.

Résumé

En décembre 2004, soit trois ans seulement après celle de l'homme, la première version de la séquence du génome de la poule a été publiée. Ce travail est le résultat de plus de dix années de recherches en génomique. Celles-ci ont débuté par la réalisation des premières cartes génétiques, puis de banques de clones BAC*, de cartes cytogénétiques incluant les microchromosomes, de cartes d'hybrides irradiés et la production de séquences d'EST. En effet, c'est la mise en commun de l'ensemble de ces données disponibles qui a contribué à l'assemblage final de la séquence. Bien entendu, comme la première version de celle du génome humain, la séquence de la poule n'est pas parfaite et des travaux de vérification et de corrections, notamment des microchromosomes, seront nécessaires. De plus, un effort d'annotation* est maintenant à réaliser. Cependant, le fait de disposer de la séquence du génome d'un nombre croissant de vertébrés va permettre d'affiner nos connaissances grâce aux études comparatives. La disponibilité de la séquence de la poule va permettre de remplacer une bonne partie des longs et fastidieux travaux de biologie moléculaire nécessaires à la détermination de la structure, de la fonction et du polymorphisme des gènes par des analyses *in silico*. Ceci devrait accélérer la mise au point de marqueurs moléculaires utilisables pour la sélection de phénotypes intéressants pour les productions animales.

Abstract

The chicken genome sequence and its utility for breeding

In December 2004, the first version of the chicken genome sequence was published three years only after the human one. This work is the result of more than ten years of research in the field of chicken genomics, started by the realisation of the first genetic maps, followed by that of BAC clone libraries, cytogenetic maps including the microchromosomes, radiation hybrid maps and the production of EST. Indeed, all the mapping data available help in the assembly of the final sequence. Of course, similar to the first version of the human sequence assembly, the chicken one is not perfect and improvements, in particular for the microchromosomes, will be necessary. Also, the task of annotating the genome will have to be pursued. However, the fact of having the genome sequence of an increasing number of vertebrate animals will enable to refine our knowledge through comparative studies. The availability of the chicken genome sequence will enable to replace much of the molecular biology work that was necessary for the determination of the structure, function and polymorphism of genes by performing in silico analyses. This should accelerate the development of molecular markers usable for the selection of phenotypes of interest in livestock production systems.

VIGNAL A., BESBES B., 2006. La séquence du génome de la poule et ses applications en sélection. INRA Prod. Anim., 19, 109-118.