

## Section C

# Marqueurs moléculaires – outil d'exploration de la diversité génétique

### 1 Introduction

Les marqueurs d'ADN sont utiles pour la recherche de base (par ex. l'analyse phylogénétique et la recherche de gènes utiles) et pour la recherche appliquée (par exemple, la sélection assistée par marqueurs, les tests de paternité et la traçabilité alimentaire). Cette section se concentre principalement sur leur application dans le cadre de la caractérisation de la diversité des ressources zoogénétiques et de la recherche des variants fonctionnels des gènes pertinents. Il est important de remarquer que l'ARN et les protéines contiennent aussi les informations clés et, par conséquent, méritent une étude parallèle; leur rôle dans la recherche des variants fonctionnels est également abordé ci-après.

La diversité des organismes est le résultat des variations des séquences d'ADN et des effets de l'environnement. La variation génétique est considérable et chaque individu d'une espèce, à l'exception des jumeaux monozygotes, possède une séquence d'ADN unique. Les variations d'ADN sont des mutations résultant de la substitution des nucléotides simples (polyphormismes d'un seul nucléotide – SNP), l'insertion ou l'élimination de fragments d'ADN de différentes longueurs (d'un seul nucléotide à plusieurs milliers de nucléotides) ou la duplication ou l'inversion de fragments d'ADN. Les variations d'ADN sont classifiées comme «neutres» si elles ne produisent aucun changement aux caractères métaboliques ou phénotypiques et, par conséquent, ne sont pas sujets à une sélection positive, négative ou équilibrée; autrement, elles sont appelées

«fonctionnelles». Les mutations des nucléotides clés d'une séquence de codage peuvent changer la composition aminoacidee d'une protéine et conduire à de nouveaux variants fonctionnels. Ces variants peuvent avoir une efficience métabolique accrue ou décrue par rapport au «type sauvage» d'origine et peuvent perdre leur fonctionnalité complètement ou même gagner une nouvelle fonction. Les mutations dans les régions régulatrices peuvent affecter les niveaux et les modèles d'expression génétique; par exemple, elles peuvent désactiver ou activer les gènes ou surexprimer ou sous-exprimer les protéines des tissus spécifiques aux différentes phases de développement ou physiologiques.

Bien que l'analyse des types simples de biomolécules soit extrêmement utile pour la compréhension des phénomènes biologiques, la recherche parallèle de grande échelle sur l'ADN, l'ARN et les protéines ouvre de nouvelles perspectives dans l'interprétation et le modelage de la complexité des organismes vivants. De nouvelles disciplines scientifiques à suffixe «-omique» sont en train de se développer. Dans ces domaines, les avancées récentes en matière de préparation, d'identification et séquençage d'ADN, d'ARN et de protéines et de stockage et d'analyse des données à grande échelle révolutionnent notre compréhension. Une vue globale et intégrée d'un ensemble complet de molécules biologiques impliquées dans des procédés biologiques complexes commence à apparaître. La génomique structurelle, la

## PARTIE 4

### Cadre 70 ADN, ARN et protéines

L'ADN (acide désoxyribonucléique) est organisé en paires de chromosomes, chacun hérité d'un des parents. Chaque gène d'un individu a par conséquent deux copies, appelées allèles, une dans chaque chromosome d'une paire. Chez les mammifères, les gènes sont éparpillés le long des chromosomes, séparés par de longues séquences d'ADN, souvent répétitives. Les gènes sont formés par des séquences codantes (exons) séparées par les introns. Ces derniers n'ont aucune information de codage des protéines, mais parfois jouent un rôle dans la régulation de l'expression génique. Les instructions codées par les gènes sont mises en place par deux procédés. Le premier est la transcription (copie) des informations génétiques dans un autre type d'acide nucléique, l'ARN (acide ribonucléique). Les exons et les introns sont transcrits dans une molécule primaire d'ARN messager (ARNm). Cette molécule est ensuite éditée, un procédé qui implique l'élimination des introns, l'union des exons et l'ajout de caractéristiques uniques à chaque bout de l'ARNm. Une molécule d'ARN mature est ainsi créée et est ensuite transportée aux structures appelées ribosomes localisés dans le cytoplasme cellulaire. Les ribosomes sont composés d'ARN ribosomal (ARNr) et de protéines et fournissent les sites du deuxième procédé – la traduction de l'information génétique,

précédemment copiée dans l'ARNm, dans un polypeptide (une protéine entière ou une des chaînes d'un complexe de protéines). La molécule d'ARNm est lue ou traduite en trois nucléotides (un codon) à la fois. La complémentarité entre le codon d'ARNm et l'anticodon d'une molécule ARN de transfert (ARNt) qui porte l'aminoacide correspondant au ribosome, assure que le polypeptide nouvellement formé contienne la séquence spécifique des acides aminés requis.

Les gènes ne sont pas tous traduits en protéines; certains expriment leur fonction en tant que molécules d'ARN (comme l'ARNr et l'ARNt impliqués dans la traduction). Récemment, de nouveaux rôles d'ARN ont été découverts dans le processus d'édition de l'ARNm et dans la régulation de l'expression génique (Storz et al., 2005; Aravin et Tuschl, 2005; Wienholds et Plasterk 2005). En fait, les ARN non codants semblent être les responsables clés des différents procédés régulateurs (Bertone et al., 2004; Clop et al., 2006). Trois types de molécules sont ainsi disponibles pour la recherche en matière de caractéristiques génétiques aux niveaux des cellules, des tissus et de l'organisme entier: l'ADN, qui contient les instructions codées; l'ARN, qui transfère les instructions à la cellule «fabrique»; et les protéines, qui sont construites selon les instructions et font fonctionner les cellules et les organismes.

transcriptomique et la protéomique sont suivies par la métabolomique et l'interactomique entre autres, et à un niveau encore plus élevé de complexité, la biologie des systèmes (Hood et al., 2004; cadre 71).

L'étude de la complexité biologique est une nouvelle frontière qui requiert une technologie moléculaire à haut débit, une vitesse et une mémoire d'ordinateur élevées, de nouvelles approches à l'analyse des données et l'intégration de compétences interdisciplinaires (cadre 72).

### Cadre 71 Les nouvelles disciplines scientifiques à suffixe «-omique»

La génomique s'occupe de la cartographie des gènes et des variations génétiques des individus et des groupes. Elle permet de comprendre la traduction de l'information génétique aux fonctions métaboliques et aux caractères phénotypiques. Elle dévoile les procédés biologiques et leurs interactions avec les facteurs environnementaux. La génomique comporte la combinaison d'un ensemble de technologies à haut débit, comme la protéomique et la métabolomique avec les techniques bio-informatiques qui facilitent la transformation, l'analyse et l'intégration d'un grand nombre de données.

## Cadre 72

### Evolutions récentes de la biologie moléculaire

A présent, les évolutions révolutionnaires de la recherche biologique moléculaire sur l'élevage et la conservation de la diversité génétique sont:

1. l'établissement d'une séquence complète du génome des races d'animaux d'élevage les plus importantes;
2. l'élaboration d'une technologie mesurant les polymorphismes aux loci éparpillés dans tout le génome (par ex. méthodes de détection des SNP); et
3. l'élaboration d'une technologie de puces à ADN pour mesurer la transcription des gènes à grande échelle.

Les informations obtenues par le séquençage du génome complet (atteint chez les poules et presque complet chez les porcs et les bovins), intégrées à la technologie SNP, accéléreront la recherche sur les gènes. La cartographie des loci à effets quantitatifs (QTL) pour identifier les régions du chromosome qui influencent un caractère cible, la présence de gènes candidats situés dans la même région et la recherche de leurs modèles d'expression (par ex. par les analyses à puces d'ADN et protéomiques) et de leur fonction entre les espèces s'assembleront pour identifier les gènes clés et expliquer la complexité de la régulation physiologique pour les caractères cibles.

Voir ci-dessous pour de plus amples détails sur ces évolutions.

## 2 Les fonctions des technologies moléculaires dans la caractérisation

L'information sur la diversité génétique est essentielle pour optimiser les stratégies de conservation et d'utilisation des ressources zoogénétiques. Les ressources nécessaires pour la conservation étant limitées, l'établissement

des priorités est souvent nécessaire. De nouveaux outils moléculaires font espérer d'atteindre l'identification des gènes impliqués dans un certain nombre de caractères, y compris les caractères adaptatifs, et des polymorphismes responsables de la variation génétique fonctionnelle (QTN – nucléotides de caractère quantitatif). Cependant, notre connaissance insuffisante nous empêche de donner la priorité aux choix de conservation sur la base de la diversité moléculaire fonctionnelle, et des mesures alternatives sont encore nécessaires. La caractérisation phénotypique fournit une estimation brute de la moyenne des variants fonctionnels des gènes d'un individu ou d'une population donnés. Cependant, la plupart des phénotypes de la majorité des espèces d'animaux d'élevage ne sont pas enregistrés.

**Première fonction.** En l'absence de données fiables sur les phénotypes et les QTN, ou à complément des données existantes, les mesures les plus rapides et rentables de la diversité génétique sont obtenues du dosage des polymorphismes en utilisant les marqueurs génétiques moléculaires anonymes. Les marqueurs anonymes fournissent probablement des informations indirectes sur les gènes fonctionnels pour les caractères importants, tout en supposant que les populations uniques avec une histoire évolutionnaire particulière aux marqueurs neutres (par ex. à cause d'un ancien isolement ou d'une domestication indépendante) portent des variants uniques des variations fonctionnelles. Les techniques moléculaires sont également utiles pour la recherche sur l'origine et la domestication des espèces d'animaux d'élevage et sur leurs migrations ultérieures, et pour obtenir l'information sur les relations évolutionnaires (arbres phylogénétiques) et l'identification des zones géographiques du mélange génétique entre des populations d'origines génétiques différentes. Le sous-chapitre 3.1 présente un résumé des techniques moléculaires utilisées pour l'évaluation de la diversité génétique intra et interraciale.

## PARTIE 4

**Deuxième fonction.** La taille effective de la population ( $N_e$ ) est un indice qui estime le nombre effectif d'animaux au sein d'une population qui se reproduit et apporte des gènes à la génération suivante. Ne est étroitement liée au niveau de consanguinité et de dérive génétique d'une population et, par conséquent, est un indicateur clé pour l'évaluation du niveau de danger des populations (voir sections A et F). Les approches traditionnelles visant à obtenir des estimations fiables de  $N_e$  pour les populations reproductrices sont fondées sur les données généalogiques ou sur les recensements. Les données nécessaires sur la variabilité de la reproduction et des intervalles de générations sont souvent disponibles de façon peu fiable pour les populations des pays en développement. Les approches moléculaires peuvent, par conséquent, être une alternative prometteuse (voir sous-chapitre 3.2 pour de plus amples détails).

**Troisième fonction.** Une des priorités les plus importantes de la gestion des ressources zoogénétiques est la conservation de races avec des caractères uniques, comme la capacité de vivre et produire dans des conditions difficiles et de résister aux maladies infectieuses, surtout pour les pays en développement. Les caractères complexes, comme l'adaptation et la résistance aux maladies ne sont pas visibles ou facilement mesurables. Ils peuvent être étudiés grâce à des expériences où les animaux sont soumis à des conditions environnementales spécifiques ou sont infectés par l'agent pertinent. Cependant, de telles expériences sont difficiles et coûteuses à mettre en place et se heurtent aux questions de bien-être des animaux. Pour ces raisons, les chercheurs sont extrêmement intéressés à identifier les gènes qui contrôlent les caractères complexes. Ces gènes peuvent se rechercher avec un certain nombre d'approches différentes. Les instruments élaborés pour cibler la variation fonctionnelle sont décrits au sous-chapitre 3.3.

### 3 Vue d'ensemble des techniques moléculaires

Cette section décrit les techniques moléculaires les plus importantes utilisées et élaborées à présent pour l'évaluation de la diversité génétique et pour le ciblage de la variation fonctionnelle. Le cadre 73 décrit les façons d'extraire l'ADN et l'ARN du matériel biologique et de les préparer pour l'analyse. Les attributs des marqueurs moléculaires habituellement utilisés sont indiqués au cadre 74 et l'échantillonnage (un aspect très important des études moléculaires) est présenté au cadre 75.

Les polyphormismes des protéines ont été les premiers marqueurs utilisés pour les études génétiques dans le secteur de l'élevage. Cependant, le nombre de loci polymorphiques pouvant se doser et le niveau des polymorphismes observés aux loci sont souvent faibles, ce qui limite beaucoup leur application aux études sur la diversité génétique. Avec le développement de nouvelles technologies, les polymorphismes d'ADN sont devenus les marqueurs préférés lors des analyses moléculaires de variation génétique (cadre 74).

#### 3.1 Techniques utilisant les marqueurs d'ADN pour évaluer la diversité génétique

##### **Marqueurs d'ADN nucléaire**

Un certain nombre de marqueurs sont à présent disponibles pour détecter les polymorphismes d'ADN nucléaire. Dans les études sur la diversité génétique, les marqueurs les plus fréquemment utilisés sont les microsatellites.

##### **Microsatellites**

Les microsatellites (cadre 74) sont à présent les marqueurs les plus utilisés dans les études de caractérisation génétique des animaux d'élevage (Sunnucks, 2001). Le taux de mutation élevé et la nature codominante favorisent l'estimation de la diversité intra et interraciale, et le mélange

### Cadre 73

#### Extraction et multiplication d'ADN et d'ARN

La première étape de l'analyse de l'ADN, de l'ARN et des protéines est l'extraction et la purification des spécimens biologiques. Plusieurs protocoles et kits commerciaux sont disponibles. Les stratégies appliquées dépendent du matériel d'origine et de la molécule ciblée. Par exemple, l'extraction d'ADN du sang complet ou des globules blancs est relativement facile, tandis que son extraction des aliments transformés est plutôt difficile. L'extraction d'ARN du tissu pancréatique est difficile à cause d'une dégradation post-mortem très rapide de cet organe. La pureté de l'ADN, de l'ARN et des protéines est souvent un facteur clé négligé pour avoir des résultats fiables.

Après avoir isolé l'ADN (ou l'ARN) des cellules, l'étape suivante est d'obtenir des milliers ou des millions de copies d'un gène ou d'une partie d'ADN spécifique. La multiplication de fragments d'ADN peut être déléguée aux micro-organismes, d'habitude les *E. coli*, ou effectuée *in vitro* en utilisant une réaction de polymérisation en chaîne (PCR). Cette technique, grâce à laquelle son inventeur, Cary Mullis, a eu le prix Nobel, amplifie de façon exponentielle tout segment d'ADN d'une séquence connue. La composante clé dans une réaction PCR est l'ADN polymérase isolé du *Thermus aquaticus*, un micro-organisme adapté à vivre et à se multiplier à une température très élevée. Ce polymérase thermostable *Taq*- (du nom *Thermus aquaticus*) favorise la réplication en chaîne à cycles et produit une croissance géométrique du nombre de copies d'ADN cible. Un cycle de PCR se compose de trois étapes: i) la dénaturation d'ADN à 90–95°C pour séparer l'ADN en deux simples brins à utiliser comme modèle; ii) l'hybridation d'une paire d'oligonucléotides à simple brin (amorces) complémentaires aux régions cibles flanquant le fragment d'intérêt, à 45–65°C; iii) l'extension ou l'elongation de brins d'ADN nouvellement synthétisés conduits par les amorces et facilités par le polymérase *Taq*-, à 72°C. Ce cycle peut se répéter, habituellement de 25 à 45 fois, pour permettre l'amplification d'un nombre suffisant d'amplicons (un fragment d'un gène ou d'ADN synthétisé en utilisant la PCR) à être détectés.

génétique entre les races, même si elles sont très proches.

Quelques contestations ont entouré le choix d'un modèle de mutation – le modèle de mutation par allèles infinis ou progressif (Goldstein *et al.*, 1995) – pour l'analyse des données des microsatellites. Cependant, des études de simulation ont indiqué que le modèle de mutation par allèles infinis est généralement valable pour l'évaluation de la diversité intraraciale (Takezaki et Nei, 1996).

Le nombre faible d'allèles par population et l'hétérozygosité observée et prévue sont les paramètres les plus communément utilisés pour évaluer la diversité intraraciale. Les paramètres les plus simples pour évaluer la diversité interraciale sont la différenciation génétique ou les indices de fixation. Plusieurs estimateurs ont été proposés (par ex. FST – indice de fixation et GST – glutathion S transférase), et le plus utilisé est FST (Weir et Basten, 1990), qui mesure le degré de différenciation génétique des sous-populations par le calcul des variances standardisées des fréquences d'allèles des populations. La signification statistique se calcule pour les valeurs de FST entre paires de populations (Weir et Cockerham, 1984) pour tester l'hypothèse nulle d'un manque de différenciation génétique entre les populations et, par conséquent, la partition de la diversité génétique (par ex. Mburu *et al.*, 2003). L'analyse de variance moléculaire (AMOVA) (Excoffier *et al.*, 1992) s'effectue pour évaluer la distribution de la diversité dans ou entre ces groupes de races.

Les données des microsatellites sont aussi habituellement utilisées pour évaluer les relations génétiques entre populations et sujets par le biais de l'estimation des distances génétiques (par ex. Beja-Pereira *et al.*, 2003; Ibeagha-Awemu *et al.*, 2004; Joshi *et al.*, 2004; Sodhi *et al.*, 2005; Tapio *et al.*, 2005). La mesure des distances génétiques utilisée le plus souvent est la distance génétique standard de Nei (Nei, 1972). Cependant, pour les populations les plus proches, où la dérive génétique est le facteur principal de différenciation génétique, ce qui est souvent le cas pour les races d'animaux d'élevage, surtout

## PARTIE 4

### Cadre 74

#### Marqueurs d'ADN habituellement utilisés

Les polymorphismes de longueur des fragments de restriction (RFLP) sont identifiés en utilisant les enzymes de restriction qui coupent l'ADN uniquement sur des «sites de restriction» déterminés (par ex. EcoRI coupe au site défini par la séquence palindrome GAATTC). A présent, l'utilisation la plus courante des RFLP est en aval de PCR (PCR-RFLP) pour détecter les allèles qui diffèrent en séquence à un site de restriction donné. Un fragment de gène est d'abord amplifié en utilisant la PCR et ensuite exposé à une enzyme de restriction spécifique qui coupe uniquement une des formes alléliques. Les amplicons digérés sont généralement résolus par l'électrophorèse.

Les microsatellites ou SSR (répétitions de séquences simples) ou STR (séquences répétées en tandem) consistent en une séquence d'ADN longue de quelques nucléotides – 2 à 6 paires de base (pb) – répétée plusieurs fois en tandem (par ex. CACACACACACACA). Ils sont répandus sur un génome eucaryote. Les microsatellites ont une taille relativement petite et, par conséquent, sont facilement amplifiés utilisant les PCR d'ADN extraits de différentes sources comme le sang, les poils, la peau ou même les fèces. Les polymorphismes peuvent se visualiser sur un gel de séquençage et la disponibilité de séquenceurs automatiques d'ADN consent une analyse à haut débit d'un grand nombre d'échantillons (Goldstein et Schlötterer, 1999; Jarne et Lagoda, 1996). Les microsatellites sont hypervariables; sur un locus, ils montrent souvent des dizaines d'allèles différents l'un de l'autre dans le nombre de répétitions. Ils sont encore les marqueurs de choix pour les études sur la diversité, l'analyse de paternité et la cartographie des loci à effets quantitatifs (QTL) bien que ceci pourrait changer, dans un avenir proche, grâce à l'élaboration de méthodes peu coûteuses de dosage des SNP. La FAO a publié des recommandations sur les ensembles de loci des microsatellites à utiliser pour les études

sur la diversité des espèces principales, qui avaient été élaborées par le Groupe consultatif sur la diversité génétique des animaux ISAG-FAO (voir la bibliothèque de DAD-IS à l'adresse Internet <http://www.fao.org/dad-is/>).

Les minisatellites ont les mêmes caractéristiques que les microsatellites, mais les répétitions vont de dix à quelques centaines de pb. Les micro et les minisatellites sont également connus sous le nom de polymorphismes VNTR (nombre variable de séquences répétées en tandem).

Les polymorphismes de longueur de fragments amplifiés (AFLP) sont une technique d'empreintes d'ADN qui détecte les fragments de restriction d'ADN par l'amplification PCR.

Les séquences uniques détectées dans le génome (STS) sont des séquences d'ADN qui se produisent seulement une fois dans un génome, à une position connue. Elles ne sont pas nécessairement polymorphiques et sont utilisées pour créer des cartes physiques.

Les polymorphismes d'un seul nucléotide (SNP) sont des variations des nucléotides simples qui ne changent pas la longueur globale de la séquence d'ADN dans la région. Ils se produisent partout dans le génome. Ils sont très abondants et dans le génome humain se trouve un SNP chaque 1000 pb (Sachinandam et al., 2001). La plupart des SNP sont localisés dans les régions non codantes et n'ont aucun impact direct sur le phénotype d'un individu. Cependant, certains introduisent des mutations dans des séquences ou des régions d'expression influençant l'expression génique (promoteurs, amplificateurs) et peuvent donner lieu à des changements dans la structure ou la régulation des protéines. Ces SNP ont les potentialités de détecter la variation génétique fonctionnelle.

### Cadre 75 L'échantillonnage de matériel génétique

La collection d'échantillons est la première et la plus importante étape de toute étude sur la diversité. Idéalement, les échantillons devraient être isolés et représentatifs des populations que l'on étudie. Généralement, l'échantillonnage de 30 à 50 sujets par race, choisis avec attention, est considéré acceptable pour fournir une première réponse sur les caractères distinctifs de la race et la diversité intraraciale, si un nombre suffisant de marqueurs indépendants est dosé (par ex. 20–30 microsatellites; Nei et Roychoudhury, 1974; Nei, 1978). Cependant, les quantités réellement nécessaires peuvent varier selon le cas, elles peuvent être même plus faibles dans le cas d'une population locale hautement consanguine, et plus élevées dans le cas d'une population largement répandue, divisée en écotypes différents.

Le choix d'échantillons isolés est assez faisable pour une race clairement définie, pouvant se baser sur le herd-book ou le livre généalogique. Inversement, il peut être plutôt difficile pour une population semi-marronnisée pour laquelle aucune inscription écrite n'est disponible. Dans ce cas, l'utilisation d'un critère géographique est vivement recommandée, c.-à-d. collecter un seul animal ou très peu d'animaux (non apparentés) par troupeau provenant d'un certain nombre de troupeaux répandus sur une zone géographique étendue. L'enregistrement des coordonnées géographiques et la documentation photographique des sites d'échantillonnage, des animaux et des troupeaux sont très importants – pour contrôler les croisements en cas d'observations aberrantes non prévues ou l'identification de modèles géographiques intéressants de diversité génétique. Un ensemble bien choisi d'échantillons est une ressource de valeur et de longue durée qui peut donner des résultats significatifs même avec une technologie simple. Au contraire, un échantillon biaisé produira des résultats faussés ou difficiles à comprendre, même si l'on utilise les instruments les plus avancés.

dans les régions en développement, la distance modifiée Cavalli-Sforza est recommandée (Nei *et al.*, 1983). La relation génétique entre les races est souvent visualisée par la reconstruction d'une phylogénie, le plus souvent en utilisant la méthode «neighbour-joining» (Saitou et Nei, 1987). Cependant, le problème principal de la reconstruction de l'arbre phylogénétique est que l'évolution des lignées est présumée non réticulée, c'est-à-dire que les lignées peuvent s'écartez, mais ne peuvent jamais provenir des croisements entre lignées. Cette hypothèse est rarement valable pour les animaux d'élevage, car les nouvelles races sont souvent issues de croisements entre deux ou plusieurs races ancestrales. La visualisation de l'évolution des races par la reconstruction phylogénétique doit donc être interprétée avec beaucoup d'attention.

L'analyse multidimensionnelle et, plus récemment, les approches par groupements bayésiens ont été suggérées pour l'analyse du mélange des données microsatellites de populations différentes (Pritchard *et al.*, 2000). L'étude la plus complète dans le secteur de l'élevage est probablement une étude sur les bovins africains de tout le continent (Hanotte *et al.*, 2002), qui révèle les signatures génétiques des origines, des mouvements secondaires et de la différenciation de l'élevage pastoral des bovins africains.

Les données génétiques moléculaires, associées et complémentées par d'autres sources, comme les preuves archéologiques et les traces écrites, fournissent des informations utiles sur les origines, et les mouvements et les évolutions ultérieurs, de la diversité génétique des espèces d'animaux d'élevage. La cartographie des origines de la diversité génétique actuelle permet de déduire la localisation de la variation génétique fonctionnelle d'une espèce pour laquelle les données sur la variation phénotypique sont limitées.

L'analyse associée des données des microsatellites obtenues d'études distinctes est très souhaitable, mais elle a été rarement possible parce que la plupart des études génétiques des

## PARTIE 4

populations qui utilisent les marqueurs d'ADN sont limitées à un petit nombre de races, provenant souvent d'un seul pays (Baumung *et al.*, 2004). Les différents sous-ensembles de marqueurs recommandés par la FAO s'utilisent souvent et aucun échantillon standard n'est génotypé parmi les projets. L'application de différents systèmes de génotypage de microsatellites est la raison des variations entre les études sur la taille estimée des allèles aux mêmes loci. Pour promouvoir l'utilisation de marqueurs communs, la FAO propose une liste classifiée à jour<sup>3</sup> de microsatellites pour les principales races d'animaux d'élevage. La FAO recommande l'utilisation des marqueurs dans l'ordre de la classification pour optimiser le nombre de marqueurs qui se chevauchent entre les différentes études indépendantes. Pour certaines espèces, l'ADN des animaux types est disponible. Par exemple, des aliquotes d'ADN standard des moutons et des chèvres, utilisé dans le cadre du projet Econogene de l'Union européenne (UE), ont été distribuées à d'autres projets de grande échelle en Asie et en Afrique et peuvent se demander au site web d'Econogene (<http://www.econogene.eu>).

Quelques exemples seulement d'analyses à grande échelle de la diversité génétique des espèces d'animaux d'élevage sont disponibles. Hillel *et al.* (2003) et SanCristobal *et al.* (2006a) ont fait des études, respectivement sur la diversité des poules et des porcs en Europe; Hanotte *et al.* (2002) ont obtenu des données sur les bovins pour tout le continent africain; Tapiro *et al.* (2005) ont évalué la diversité des moutons à une grande échelle régionale dans les pays de l'Europe du Nord; et Cañon *et al.* (2006) ont étudié la diversité des chèvres en Europe et au Proche et Moyen-Orient. Cependant, pour la plupart des espèces, un examen complet n'est pas encore disponible. Une estimation globale de la diversité génétique pour certaines espèces, comme les moutons et les

chèvres, sera probablement disponible dans un avenir proche grâce à l'étroite coordination entre les différents projets de grande échelle. Entre-temps, de nouvelles méthodes d'analyse des données sont en voie d'élaboration pour favoriser la méta-analyse d'ensembles de données qui ont en commun seulement quelques races et aucun ou seulement quelques marqueurs (Freeman *et al.*, 2006). Cette perspective mondiale sur la diversité des animaux d'élevage aura une grande importance pour la reconstruction de l'origine et de l'histoire des populations d'animaux domestiques et, de façon indirecte, des populations humaines. Elle mettra également en lumière les points focaux régionaux et locaux de la diversité génétique que les activités de conservation pourront cibler.

### SNP

Les SNP (cadre 74) sont utilisés en alternative aux microsatellites dans les études sur la diversité génétique. Plusieurs technologies sont disponibles pour détecter et typer les marqueurs SNP (voir Syvänen, 2001, pour un examen approfondi). En tant que marqueurs bialléliques, les SNP ont des quantités d'informations relativement faibles et, pour atteindre le niveau d'information d'un panel standard de 30 loci de microsatellites, il faut en utiliser de plus grandes quantités. Cependant, les technologies moléculaires toujours en évolution accroissent l'automatisation et réduisent le coût du typage des SNP, ce qui permettra probablement, dans un avenir proche, l'analyse parallèle d'un grand nombre de marqueurs à un coût réduit. Dans cette perspective, des projets de grande envergure sont en œuvre pour plusieurs espèces d'animaux d'élevage afin d'identifier des millions de SNP (par ex. Wong *et al.*, 2004) et en valider plusieurs milliers et identifier les blocs d'haplotype dans le génome. De même que pour les informations sur les séquences, les SNP permettent une comparaison directe et une analyse conjointe des différentes expériences.

Les SNP seront probablement des marqueurs intéressants à appliquer à l'avenir dans les études sur la diversité génétique, parce qu'ils peuvent

<sup>3</sup> Les listes et les directives peuvent se trouver dans la bibliothèque de DAD-IS à l'adresse Internet <http://www.fao.org/dad-is>.

être facilement utilisés dans l'évaluation de la variation fonctionnelle ou neutre. Cependant, la phase préliminaire de la découverte des SNP ou de la sélection des SNP à partir des bases de données est critique. Les SNP peuvent être générés par différents protocoles expérimentaux comme le séquençage, le polyphormisme de conformation simple brin (SSCP – single-stranded conformational polymorphism) ou la dénaturation de la chromatographie liquide de haute performance (DHPLC - denaturing high-performance liquid chromatography) ou *in silico*, alignant et comparant des séquences multiples de la même région à partir des bases de données publiques sur les génomes et les étiquettes séquentielles d'expression (EST). Si les données ont été obtenues de façon aléatoire, les estimateurs standard des paramètres génétiques de la population ne peuvent pas s'appliquer. Un exemple fréquent est lorsque les SNP initialement identifiés dans un petit échantillon (panel) d'individus sont ensuite typés dans un échantillon plus large de chromosomes. Effectuant de préférence un échantillonnage de SNP aux fréquences intermédiaires, un tel protocole affectera la distribution des fréquences alléliques par rapport aux valeurs probables pour un échantillon aléatoire. Les SNP font réellement espérer une application future dans le cadre des analyses génétiques de la population; cependant, il est nécessaire d'élaborer des méthodes statistiques qui prendront explicitement en compte chaque méthode de découverte des SNP (Nielsen et Signorovitch, 2003; Clark et al., 2005).

#### AFLP

Les AFLP sont des marqueurs bialléliques dominants (Vos et al., 1995). Les variations sur de nombreux loci peuvent se ranger simultanément pour détecter les variations des nucléotides uniques des régions génomiques inconnues, où une mutation donnée peut être souvent présente dans les gènes fonctionnels non déterminés. L'inconvénient est qu'ils montrent un mode dominant d'hérédité, ce qui réduit leur pouvoir lors des analyses génétiques de la population sur la diversité intraraciale et la consanguinité.

Cependant, les profils des AFLP sont hautement informatifs dans le cadre de l'évaluation des relations entre les races (Ajmone-Marsan et al., 2002; Negrini et al., 2006; De Marchi et al., 2006; SanCristobal et al., 2006b) et les espèces apparentées (Buntjer et al., 2002).

#### Marqueurs d'ADN mitochondrial

Les polymorphismes d'ADN mitochondrial (ADNmt) ont été largement utilisés lors des analyses de la diversité phylogénétique et génétique. L'ADNmt haploïde transporté par les mitochondries du cytoplasme cellulaire, possède un mode maternel d'hérédité (les animaux héritent l'ADNmt de leurs mères et non de leurs pères) et un taux de mutation élevé; il ne se recombine pas. Ces caractéristiques consentent aux biologistes de reconstruire les relations évolutionnaires intra et interraciales par l'évaluation des modèles de mutation de l'ADNmt. Les marqueurs d'ADNmt peuvent également fournir un moyen rapide de détecter l'hybridation entre les espèces et les sous-espèces d'animaux d'élevage (par ex. Nijman et al., 2003).

Les polymorphismes dans la séquence de la région hypervariable de la boucle D ou de la région de contrôle de l'ADNmt ont largement contribué à l'identification des descendants sauvages des espèces domestiques, à l'établissement des modèles géographiques de la diversité génétique et à la compréhension de la domestication des animaux d'élevage (voir Bruford et al., 2003, pour un examen approfondi). Par exemple, l'origine imputable au Moyen-Orient des modernes bovins européens a été récemment démontrée par Troy et al. (2001). L'étude a identifié quatre lignées maternelles dans le *Bos taurus* et a également démontré la perte de variabilité génétique bovine lors de la migration humaine du Néolithique en provenance du croissant fertile. De la même façon, des origines maternelles multiples avec trois lignées d'ADNmt ont été mises en évidence chez les chèvres (Luikart et al., 2001), l'Asie et le croissant fertile étant les centres possibles d'origine. Récemment, une troisième lignée d'ADNmt a été découverte chez les moutons

## PARTIE 4

indigènes chinois (Guo *et al.*, 2005), une quatrième chez les chèvres indigènes chinoises (Chen *et al.*, 2005), et une cinquième chez les bovins indigènes chinois (Lai *et al.*, 2006). Chez les poules asiatiques, on a repéré neuf clades différents d'ADNmt (Liu *et al.*, 2006) qui suggèrent des origines multiples en Asie du Sud et du Sud-Est. Ces résultats indiquent que la connaissance de la domestication des animaux d'élevage et de la diversité génétique est encore largement incomplète. Pour de plus amples renseignements sur les origines des espèces domestiques d'animaux d'élevage voir partie 1 – section A.

### 3.2 Utilisation des marqueurs pour l'estimation de la taille effective de la population

Hill (1981) a proposé d'utiliser le déséquilibre gamétique des polymorphismes d'ADN pour estimer la taille effective de la population ( $N_e$ ). Cette estimation se base sur les génotypes pour les marqueurs liés (microsatellites ou SNP). La corrélation prévue des fréquences alléliques aux loci liés est une fonction de  $N_e$  et du taux de recombinaison.  $N_e$  peut donc se calculer à partir du déséquilibre observé. Hayes *et al.* (2003) ont suggéré une approche semblable basée sur l'homozygosité du segment chromosomique qui a, en outre, les potentialités d'estimer  $N_e$  des générations précédentes et permet ainsi de savoir si une population existante était par le passé de taille croissante ou décroissante. L'étude a démontré, avec des ensembles de données utilisés comme exemples, que la race bovine Holstein Frisonne a subi une réduction substantielle de  $N_e$  par le passé, tandis que la taille effective de la population humaine est à la hausse, ce qui est en accord avec les recensements et les études généalogiques.

### 3.3 Outils moléculaires pour cibler la variation fonctionnelle

#### *Approches basées sur la position: cartographie des locis à effets quantitatifs*

Les marqueurs génétiques se comportent comme les caractères mendéliens; en d'autres termes, ils suivent les lois de la ségrégation et de l'assortiment indépendant décrites pour la première fois par Mendel. Deux gènes localisés sur le même chromosome sont physiquement liés et ont tendance à être hérités ensemble. Au cours de la méiose, la recombinaison entre les chromosomes homologues peut interrompre ce lien. La fréquence de recombinaison entre deux gènes localisés sur le même chromosome dépend de leur distance. Le taux de recombinaison entre les marqueurs est, par conséquent, une indication de leur degré de lien: plus le taux de recombinaison est faible, plus les marqueurs sont proches. La création des cartes génétiques utilise cette caractéristique pour déduire l'ordre possible des marqueurs et la distance entre eux.

Les activités de cartographie sont généralement accomplies en suivant la coségrégation des marqueurs polymorphiques chez les populations d'expérimentation structurée (par ex. F2 ou rétrocroisement) ou les populations existantes des programmes de sélection (familles de frères ou demi-frères). Les cartes génétiques, qui ont une densité moyenne à haute – de quelques centaines à quelques milliers de marqueurs, sont disponibles pour la plupart des espèces d'animaux d'élevage.

Pour identifier le QTL pour un caractère donné, la famille dissociée pour le caractère est génotypée avec un ensemble de marqueurs moléculaires cartographiés répandus de façon homogène sur le génome (cadre 76). Un certain nombre de méthodes statistiques sont disponibles pour déduire la présence d'un QTL significatif dans un intervalle de marqueurs donné, mais elles dépendent du fait que les familles possèdent un haut niveau de déséquilibre de liaison, c.-à-d. les grands segments des chromosomes sont transmis sans recombinaison des parents à la descendance.

Le résultat d'une expérience de cartographie des QTL est l'identification d'une région chromosomique, recouvrant souvent la moitié d'un chromosome dans lequel un effet significatif est détecté pour le caractère cible. La recherche moderne utilise activement la cartographie pour identifier les QTL influençant les caractères adaptatifs. Quelques exemples de tels caractères incluent, chez les poules, une plus grande résistance à la colonisation et à l'excrétion de *Salmonella* (Tilquin *et al.*, 2005), et la tendance à développer un syndrome d'hypertension pulmonaire (Rabie *et al.*, 2005); et chez les bovins, la trypanotolérance (Hanotte *et al.*, 2002).

La phase de cartographie des QTL est généralement suivie par l'affinement de la position des QTL (cartographie affinée du QTL). Pour accomplir cette tâche, on analyse des marqueurs additionnels et, surtout, des événements de recombinaison additionnels dans la zone ciblée. Une approche intelligente a récemment été conçue et appliquée à la cartographie affinée d'une région chromosomique sur BTA14 ayant un QTL significatif pour le pourcentage de matière grasse dans le lait et d'autres caractères (Farnir *et al.*, 2002). Cette approche utilise la recombinaison historique des générations passées pour limiter la position à une région relativement petite de 3,8 cM (centimorgan), une taille qui a permis le clonage positionnel du gène (DGAT1) (Grisart *et al.*, 2002).

Grâce à la cartographie affinée, les gènes déterminant le caractère de la performance peuvent se rechercher parmi les gènes situés dans les régions identifiées. Les gènes candidats peuvent se rechercher chez la même espèce (par ex. lorsqu'une carte riche d'EST est disponible ou lorsque le génome est complètement séquencé) ou dans les régions orthologues d'un organisme modèle pour lequel l'information complète sur le génome est disponible.

De temps en temps, les informations clés sur la fonction des gènes sont fournies de sources inattendues. Ceci s'est avéré avec le gène myostatine, dont la fonction a été découverte pour la première fois chez les souris et ensuite repérée

## Cadre 76

### Cartographie des QTL

Si un QTL pour un caractère cible existe, les allèles plus ou moins variants du gène responsable inconnu (Q et q) coségrèguent avec les allèles d'un marqueur M1 proche (M1 et m1) que nous pouvons génotyper au laboratoire. Si M1 coségrège avec Q et m1 avec q, cela signifie que M1 et Q sont proches sur le même chromosome et m1 et q sur le chromosome homologue (M1Q et m1q).

Imaginons également qu'une population F2, dérivée de l'accouplement d'individus hétérozygotes F1, est génotypée. En suivant le génotypage, les descendants F2 sont groupés sur la base de leur génotype marqueur (M1M1 et m1m1; M2M2; ... MnMn et mnmn) et ensuite le phénotype moyen des groupes est comparé. Si aucun QTL est lié à un marqueur donné (par ex. M2), alors aucune différence significative ne sera détectée entre la valeur phénotypique moyenne des descendances M2M2 et m2m2 pour le caractère cible. Au contraire, lorsque les descendances sont groupées par leur génotype au marqueur M1, alors le groupe M1M1 sera principalement QQ au QTL et le groupe m1m1 sera surtout qq. Dans ce cas, une différence significative est observée entre les moyennes de la descendance et, par conséquent, la présence d'un QTL est détectée. Dans les espèces comme les volailles et les porcs, où les lignées et les races sont habituellement croisées commercialement, cet exercice peut s'accomplir dans les populations expérimentales (F2, BC) tandis que chez les ruminants, deux (dispositif fille) ou trois (dispositif petite-fille) arbres généalogiques de génération sont généralement utilisés. Dans le dispositif fille, la ségrégation de marqueurs hétérozygotes chez le père (première génération) est suivie chez les filles (deuxième génération) desquelles les données phénotypiques sont collectées. Dans le dispositif petite-fille, la ségrégation des marqueurs hétérozygotes chez un grand-père (génération I) est suivie chez ses fils demi-frères (génération II), dont le phénotype est déduit des marqueurs des petites-filles (génération III).

## PARTIE 4

chez les bovins dans la région chromosomique où le gène de l'hypertrophie musculaire avait été précédemment cartographié (McPherron et Lee, 1997).

Il est clair que l'identification du gène responsable (gènes de caractère quantitatif – QTG) et la mutation fonctionnelle (QTN) d'un caractère complexe est encore une tâche importante et plusieurs approches sont nécessaires pour réduire le nombre de gènes candidats positionnels. Les informations sur la fonction des gènes sont fondamentales dans ce sens. Cependant, on ignore encore la ou les possibles fonctions de la plupart des gènes identifiés par le séquençage du génome et de l'ADNc (ADN complémentaire). Pour cette raison, les recherches sur les modèles d'expression des gènes peuvent fournir des informations utiles, en combinaison avec l'approche positionnelle décrite ci-dessus, pour identifier les gènes candidats pour les caractères complexes. Cette approche combinée s'appelle génomique génétique (Haley et de Koning, 2006). De nouvelles avancées dans la recherche sur les schémas d'expression des gènes sont décrites à la section suivante.

Des approches alternatives sont à présent étudiées pour détecter les gènes adaptatifs en utilisant les marqueurs génétiques (cadre 77). Elles sont actuellement à l'état expérimental et seulement des recherches plus approfondies permettront une évaluation de leur efficacité.

L'objectif final de la cartographie des QTL est l'identification des QTG, et éventuellement des QTN. Bien qu'à ce jour les exemples soient rares dans le secteur de l'élevage, ces exemples sont le genre de mutations qui pourraient avoir un impact direct sur la sélection assistée par marqueurs et sur la prise de décision relative à la conservation. Les modèles de conservation qui considèrent les caractères fonctionnels et la mutation doivent se développer car un nombre croissant de QTG et de QTN seront découverts dans un avenir proche.

### ***Etude des modèles d'expression génique***

Par le passé, l'expression de caractères spécifiques, comme l'adaptation et la résistance, pouvaient se

mesurer uniquement au niveau phénotypique. A présent, le transcriptome (l'ensemble de tous les transcrits d'une cellule ou d'un tissu) et le protéome (l'ensemble de toutes les protéines) peuvent s'étudier directement grâce à des techniques à haut débit, comme la représentation différentielle (Liang et Pardee, 1992), les ADNc-AFLP (Bachem *et al.*, 1996), l'analyse en série de l'expression génique (SAGE) (Velculescu *et al.*, 1995; 2000), la spectrométrie de masse et les puces à protéines et à ADN. Ces techniques représentent une grande avancée dans l'analyse de l'ARN et des protéines et facilitent l'analyse parallèle de pratiquement tous les gènes exprimés dans un tissu à un moment donné. Ainsi, ces techniques apportent une contribution au décodage de réseaux qui codent probablement pour de nombreux caractères complexes.

Les technologies à suffixe -omique sont souvent comparées à une grande lumière allumée sur un tableau de Michel Ange qui fait tout voir, par rapport à une torche qui éclaire seulement des parties du tableau. La vue complète permet de comprendre le sens de la représentation et d'apprécier sa beauté. En fait, le pouvoir de ces techniques est accompagné par la difficulté et les coûts impliqués dans leur mise en œuvre et dans l'analyse des données produites. L'isolement d'échantillons de cellules homogènes est assez difficile et représente une condition préalable importante dans de nombreuses études sur les profils des expressions géniques. Le grand nombre de dosages parallèles produit un coût modéré par dosage, mais un coût élevé par expérience. Le matériel est coûteux et des compétences techniques élevées sont nécessaires dans toutes les phases de l'expérience, ce qui s'ajoute à la difficulté générale éprouvée lors de l'analyse de l'ARN, par rapport à celle de l'ADN. L'ARN est très sensible à la dégradation et il faut être particulièrement soigneux lors de son extraction des tissus ayant un métabolisme très actif. En fait, la conservation et la manipulation des échantillons sont fondamentales pour la réussite des expériences d'analyse de l'ARN. L'application des nanotechnologies à l'analyse des molécules

biologiques ouvre des perspectives prometteuses pour résoudre ces problèmes (Sauer *et al.*, 2005).

Le traitement des données représente un autre problème. Les ensembles de données moléculaires, comme les profils d'expressions géniques, peuvent être produits au cours d'une période de temps relativement limitée. Cependant, la standardisation des données entre les laboratoires est nécessaire si l'on veut obtenir une analyse cohérente des différents ensembles de données biologiques. Il est essentiel de trouver des accords sur la standardisation et sur la création de bases de données interconnectées pour obtenir une analyse efficace des réseaux moléculaires.

#### *Etablissement du profil des transcrits*

Cette section décrit brièvement la SAGE et les techniques des puces à ADN. Les descriptions d'autres techniques peuvent se trouver dans un certain nombre d'analyses récentes (par ex. Donson *et al.*, 2002). La SAGE produit des profils d'expression complets des tissus ou des lignées cellulaires. Elle implique la création de bibliothèques d'ARNm total pour entreprendre une analyse quantitative de tous les transcrits exprimés ou désactivés à des phases particulières de l'activation cellulaire. Elle est basée sur trois principes: i) un marqueur de séquence courte (9-14 pb) obtenu d'une région définie dans chaque transcrit d'ARNm qui contient les informations en quantité suffisante pour identifier de façon unique un transcrit spécifique; ii) les marqueurs de séquence peuvent être liés et former de longues molécules d'ADN (concatémères) qui peuvent être clonées et séquencées - le séquençage des clones de concatémères entraîne l'identification rapide de nombreux marqueurs individuels; iii) le niveau d'expression du transcrit est quantifié par le nombre de fois qu'un marqueur déterminé est observé.

Les puces à ADN peuvent être utilisées pour comparer, lors d'une seule expérience, les niveaux d'expression d'ARNm de plusieurs milliers de gènes entre deux systèmes biologiques, par exemple, entre les animaux vivant dans un

environnement normal et les animaux vivant dans un environnement difficile. La technologie des puces à ADN facilite également la compréhension des schémas temporeux et spatiaux de l'expression des gènes en réponse à une vaste gamme de facteurs auxquels l'organisme est exposé.

De très petits volumes de solution d'ADN sont imprimés sur toute surface de matériel non poreux, comme le verre, et créent des points d'un diamètre allant de 100 à 160 µm. A présent, environ 50 000 ADN complémentaires peuvent être automatiquement tachés sur une lamelle porte-objet du microscope. Les puces à ADN contiennent plusieurs centaines de gènes connus et quelques milliers de gènes inconnus. La puce à ADN est tachée avec des fragments d'ADNc ou avec des oligonucléotides préfabriqués. Cette dernière option a l'avantage d'une plus haute spécificité et reproductibilité, mais peut être conçue uniquement lorsque la séquence est connue. L'utilisation des puces à ADN se base sur le principe «d'hybridation», c'est-à-dire l'exposition de deux séquences d'ADN simple brin ou d'un ADN et un ARN, suivie par la mesure du montant de molécules double brin formées. L'expression de l'ARNm peut se mesurer de façon qualitative et quantitative. Elle indique l'activité génétique dans un tissu et, habituellement, est directement liée à la production de protéines induite par cet ARNm.

La caractérisation de l'expression génique contribue à la compréhension des mécanismes biologiques et, par conséquent, facilite l'identification des gènes candidats. Le pool de gènes engagé dans l'expression de la trypanotolérance chez les bovins, par exemple, a été caractérisé par la SAGE (Berthier *et al.*, 2003), et par l'analyse des puces à ADNc (Hill *et al.*, 2005). L'étude parallèle de l'expression de nombreux gènes favorise l'identification des gènes maîtres responsables des caractères phénotypiques qui restent non détectés par l'analyse de l'expression différentielle. Ces gènes maîtres peuvent, par exemple, posséder des allèles différents, mais tous exprimés au même niveau, ce qui facilite

## PARTIE 4

### Cadre 77

#### L'approche de la génomique des populations

Une approche alternative à l'identification des régions des génomes avec les gènes pertinents a été récemment proposée. Il s'agit de la détection des «signatures de sélection» par une approche de «génomique des populations» (Black *et al.*, 2001; Luikart *et al.*, 2003). Les trois principes fondamentaux de l'approche de génomique des populations à la cartographie des QTL sont:

- 1) les loci neutres à travers le génome seront affectés de façon semblable par la dérive génétique, par la démographie et l'histoire des évolutions des populations;
  - 2) les loci en sélection se comporteront souvent de façon différente et, par conséquent, montreront des modèles «aberrants» de variation, de perte de diversité (hausse de diversité si les loci sont dans une sélection équilibrée), de déséquilibre de liaison et d'indices de  $Gst/Fst$  accus/décrus; et
  - 3) par les effets «d'autostop», la sélection influencera également les marqueurs liés facilitant la détection d'une «signature de sélection» (effets aberrants) qui peut souvent se trouver par le génotypage d'un grand nombre de marqueurs dans un chromosome et par l'identification de groupements d'aberrants. Cette approche utilise les données phénotypiques au niveau de la race (ou des sous-populations intraraciales), plutôt qu'au niveau individuel et, par conséquent, est complémentaire aux approches classiques de cartographie des QTL de l'arbre généalogique.
- Cette approche peut également identifier les gènes sujets à une forte pression de sélection

et éventuellement fixés au sein des races et, en particulier, les gènes impliqués dans l'adaptation aux environnements extrêmement difficiles, la résistance aux maladies, etc. Plusieurs caractères, qui revêtent une grande importance pour la durabilité de la sélection animale, sont difficiles ou impossibles à étudier par la cartographie classique des QTL ou par des approches d'études associatives. Les potentialités de la génomique des populations ont été récemment étudiées d'un point de vue théorique (Beaumont et Balding, 2004; Bamshad et Wooding, 2003), et par un travail expérimental avec différents types de marqueurs chez les populations naturelles (AFLP: Campbell et Bernatchez, 2004; microsatellites: Kayser *et al.*, 2003; SNP: Akey *et al.*, 2002). L'approche a été récemment appliquée dans le cadre du projet Econogene (<http://lasig.epfl.ch/projets/econogene>). Lors d'une première analyse, trois SNP dans les gènes MYH1 (myosine 1), MEG3 (callipyge) et CTSB (cathepsine B) chez les moutons ont indiqué un comportement aberrant significatif (Pariset *et al.*, 2006).

Dans le cadre du même projet, une approche nouvelle, basée sur la méthode d'analyse spatiale, a été conçue pour détecter les signatures de la sélection naturelle dans le génome des animaux domestiques et sauvages (Joost, 2006). Les résultats préliminaires obtenus par cette méthode coïncident avec les résultats obtenus par l'application des modèles théoriques à la génétique des populations, comme ceux qui ont été élaborés par Beaumont et Balding (2004). La méthode d'analyse spatiale va plus loin des approches classiques, car elle est conçue pour identifier les paramètres environnementaux associés aux marqueurs sélectionnés.

l'expression des gènes en aval ayant une efficience différente. Dans ce cas, le gène maître peut se rechercher soit en exploitant la connaissance des voies métaboliques soit par l'approche des QTL d'expression (QTLe) (Lan *et al.*, 2006). Grâce à cette approche, le niveau d'expression des gènes

en aval se mesure dans une population sujette à la ségrégation. Le montant de transcrit de chaque gène est considéré un caractère phénotypique et le QTL qui influence l'expression génique peut se rechercher en utilisant les méthodologies décrites ci-dessus. Il faudrait toutefois noter que l'analyse

des données pour la détection du QTL est encore assez difficile à maîtriser. Ceci est également valable pour les techniques d'établissement du profil des transcrits à cause des nombreux faux signaux qui se produisent.

#### *Etablissement du profil des protéines*

L'étude systématique des structures des protéines, des modifications post-traductionnelles, des profils des protéines et des interactions protéine-protéine, protéine-acide nucléique et protéine-petite molécule, et l'expression spatiale et temporelle des protéines dans les cellules eucaryotes sont deux aspects fondamentaux si l'on veut comprendre les phénomènes biologiques complexes. Les protéines sont essentielles pour la structure des cellules vivantes et leurs fonctions.

La structure d'une protéine peut être révélée par la diffraction des rayons x ou par la spectroscopie de résonance magnétique nucléaire. La première méthode exige une quantité élevée de protéines cristallines et ce fait est souvent restrictif. Pour comprendre la fonction des protéines et les interactions protéine-protéine au niveau moléculaire, il serait utile de déterminer la structure de toutes les protéines présentes dans une cellule ou dans un organisme. Aujourd'hui toutefois, ceci n'est pas encore possible. Il est intéressant de noter que le nombre de variants de protéines différentes découlant de la synthèse protéique (épissage alternatif et/ou modifications post-traductionnelles) est beaucoup plus élevé que le nombre de gènes dans un génome.

La spectrométrie de masse (une technique analytique utile pour la détermination de la masse moléculaire) en combinaison avec les techniques de séparation chromatographique ou électrophorétique, est à présent la méthode préférée pour l'identification des protéines endogènes des cellules, pour la caractérisation des modifications post-traductionnelles et la détermination de la densité des protéines (Zhu et al., 2003). L'électrophorèse sur gel bidimensionnelle est unique pour ce qui concerne le nombre élevé de protéines ( $>10\,000$ ) qui peuvent être séparées et visualisées au cours d'un seul essai. Les taches

de protéines sont séparées du gel, à la suite de la digestion protéolytique, et les protéines sont identifiées en utilisant la spectrométrie de masse (Aebersold et Mann, 2003). Cependant, la standardisation et l'automatisation de l'électrophorèse sur gel bidimensionnelle ont été difficiles à mettre en place et l'utilisation des modèles de protéines résultants en tant que cartes protéomiques n'a réussi que dans quelques cas. Une technique complémentaire, la chromatographie liquide, est plus facile à automatiser et peut être directement connectée à la spectrométrie de masse. Les méthodes protéomiques basées sur l'affinité et fondées sur les puces sont une approche alternative à l'établissement du profil des protéines (Lueking et al., 2003), et peuvent également s'utiliser pour détecter les interactions protéine-protéine. Ces informations sont essentielles pour la modélisation algorithmique des cycles biologiques. Cependant, la spécificité liante reste un problème dans l'application des puces à protéines parce que la réactivité croisée ne peut pas se prévoir de façon précise. Une des approches alternatives pour la détection des interactions protéine-protéine est le système à deux hybrides (Fields et Song, 1989). Cependant, aucune des méthodes utilisées ne permet la détection quantitative des protéines liantes et il n'est pas encore clair jusqu'à quel degré les interactions observées représentent les interactions physiologiques protéine-protéine.

Ces méthodes basées sur les arrangements ont été également élaborées pour détecter l'interaction ADN-protéine *in vitro* et *in vivo* (voir Sauer et al., 2005, pour un examen plus approfondi) et l'identification des protéines inconnues liant les séquences régulatrices des gènes. Les puces à ADN sont employées avec efficacité pour l'examen des extraits nucléaires des complexes liant l'ADN, tandis que les puces à protéines sont principalement utilisées pour l'identification des protéines liant l'ADN inconnues au niveau du protéome. A l'avenir, ces deux techniques révéleront des informations détaillées dans les réseaux régulateurs transcriptionnels.

De nombreuses méthodes de prévision de la fonction d'une protéine se basent sur l'homologie

## PARTIE 4

avec d'autres protéines et sur sa localisation à l'intérieur de la cellule. Les prévisions des fonctions protéiques sont assez compliquées et demandent des techniques pour détecter les interactions protéine-protéine et les liaisons des protéines avec d'autres molécules, car les protéines remplissent leurs fonctions dans ces procédés de liaison.

### 4 Le rôle de la bio-informatique

L'élaboration de technologies à haut débit serait inutile si l'on ne disposait pas des capacités d'analyse des données biologiques en croissance exponentielle. Ces données doivent se stocker dans des bases de données électroniques (cadre 78) associées à des logiciels spécialement conçus pour la mise à jour, l'interrogation et l'extraction. Les informations doivent être facilement accessibles et flexibles aux interrogations pour faciliter l'extraction des informations pouvant être analysées pour éclaircir les voies métaboliques et le rôle des protéines et des gènes impliqués.

La bio-informatique est fondamentale pour associer les informations provenant de sources différentes et générer une connaissance nouvelle à partir de données existantes. Elle dispose également des potentialités pour simuler la structure, la fonction et la dynamique des systèmes moléculaires et est donc utile pour la formulation des hypothèses et pour la conduite du travail expérimental.

### 5 Conclusions

La caractérisation moléculaire peut jouer un rôle important dans la découverte de l'histoire et dans l'estimation de la diversité, du caractère distinctif et de la structure de la population de ressources zoogénétiques. Elle peut également apporter une aide à la gestion génétique des petites populations pour éviter la consanguinité excessive. Un certain nombre de recherches ont décrit la diversité

#### Cadre 78

#### Bases de données sur les molécules biologiques

Un certain nombre de bases de données existent et collectent les informations sur les molécules biologiques:

##### Bases de données sur les séquences de l'ADN:

- European Molecular Biology Lab (EMBL): <http://www.ebi.ac.uk/embl/index.html>
- GenBank: <http://www.ncbi.nlm.nih.gov/>
- DNA Data Bank of Japan (DDBJ): <http://www.ddbj.nig.ac.jp>

##### Bases de données sur les protéines:

- SWISS-PROT: <http://www.expasy.ch/sprot/sprot-top.html>
- Protein Information Resource (PIR): <http://pir.georgetown.edu/pirwww/>
- Protein Data Bank (PDB): <http://www.rcsb.org/pdb/>

##### Bio-portail sites utilitaires pour l'identification des gènes:

- GenomeWeb: <http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-geneid.html>
- BCM Search Launcher: <http://searchlauncher.bcm.tmc.edu/>
- MOLBIOL: <http://www.molbiol.net/>
- Pedro's BioMolecular Research tools: [http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research\\_tools.html](http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research_tools.html)
- ExPASy Molecular Biology Server: <http://www.expasy.ch/>

##### Bases de données d'intérêt particulier pour les animaux domestiques:

- <http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl>
- <http://www.cgd.csiro.au/cgd.html>
- <http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/>
- <http://www.marc.usda.gov/genome/genome.html>
- <http://www.ncbi.nlm.nih.gov/genome/guide/pig/>
- <http://www.ensembl.org/index.html>
- <http://www.tigr.org/>
- <http://omia.angis.org.au/>
- <http://www.livestockgenomics.csiro.au/ibiss/>
- <http://www.thearkdb.org/>
- <http://www.hgsc.bcm.tmc.edu/projects/bovine/>

intra et interraciale – certaines plutôt à grande échelle. Cependant, ces études sont fragmentées et difficiles à comparer et à intégrer. De plus, une enquête globale au plan mondial sur les espèces pertinentes n'a pas été mise en place. Ainsi, il est stratégiquement important d'élaborer des méthodes pour la combinaison des ensembles de données existants et partiellement chevauchés, et d'assurer la fourniture d'échantillons et de marqueurs standards pour une utilisation future en tant que références mondiales. Un réseau d'installations qui collectent des échantillons de matériel génétique autochtone pour la communauté scientifique, dans le cadre d'une réglementation appropriée, faciliterait la mise en œuvre d'une enquête mondiale.

Les technologies associées aux marqueurs sont en évolution et il est probable que les microsatellites soient de plus en plus complémentés par les SNP. Ces marqueurs sont prometteurs car ils sont très nombreux dans le génome et appropriés pour l'automatisation dans la production et l'évaluation. Cependant, l'efficacité des SNP dans les recherches sur la diversité des espèces animales reste à explorer de façon approfondie. Le sujet devrait être abordé avec un détachement critique suffisant à éviter la production de résultats entachés d'erreurs systématiques.

Les méthodes d'analyse des données sont également en évolution. De nouvelles méthodes facilitent l'étude de la diversité sans les hypothèses *a priori* sur la structure des populations enquêtées; l'exploration de la diversité pour identifier les gènes adaptatifs (par ex. en utilisant la génomique des populations, voir cadre 77); et l'intégration de l'information provenant de sources différentes, y compris les paramètres socio-économiques et environnementaux, pour établir les priorités relatives à la conservation (voir section F). L'adoption d'une stratégie correcte d'échantillonnage et la collecte systématique de données phénotypiques et environnementales restent parmi les exigences clés si l'on veut exploiter tout le potentiel des nouvelles technologies et des nouvelles approches.

En plus de la variation neutre, la recherche se penche activement sur les gènes qui influencent les caractères clés. La résistance aux maladies, le rendement de la production et la qualité des produits sont parmi les caractères hautement prioritaires. Un certain nombre de stratégies et de nouvelles technologies à haut débit à suffixe –omique sont utilisées à ces fins. L'identification des QTN offre de nouvelles possibilités et de nouveaux défis en matière de gestion des ressources zoogénétiques. Les informations sur la diversité adaptative complètent les renseignements sur la diversité génétique phénotypique et neutre et peuvent s'intégrer aux instruments décisionnels sur la gestion et la conservation des ressources zoogénétiques. L'identification d'allèles uniques ou de combinaisons d'allèles pour les caractères adaptatifs de populations spécifiques peut renforcer la justification de leur conservation et utilisation ciblée. La sélection assistée par les gènes peut également réduire les différences d'efficacité dans la sélection entre les grandes populations des systèmes de production industrielle et les petites populations locales, où les systèmes d'évaluation génétique et les programmes de sélection ne peuvent pas s'appliquer de manière efficace. La sélection assistée par marqueurs et par les gènes n'est pas toujours la meilleure solution. Ces options doivent être évaluées et optimisées au cas par cas, en considérant les effets à court et à long terme sur la structure et les taux de consanguinité de la population et les coûts et les avantages en termes environnementaux et socio-économiques – surtout les impacts sur les moyens d'existence des populations humaines.

De même que dans d'autres technologies avancées, il est hautement souhaitable que les avantages des avancées scientifiques dans le domaine de la caractérisation moléculaire soient partagés de par le monde, contribuant ainsi à une meilleure compréhension, utilisation et conservation des ressources zoogénétiques dans le monde pour le bien des générations présentes et futures.

## PARTIE 4

## Cadre 79

### Glossaire: marqueurs moléculaires

Aux fins de cette section, les définitions suivantes ont été utilisées.

**Gène candidat:** tout gène qui pourrait plausiblement déterminer des différences dans les caractéristiques observables d'un animal (par ex. la résistance aux maladies, la production de protéines de lait ou la croissance). Le gène peut être candidat parce qu'il est localisé dans une région particulière du chromosome dont la fonction déduite suggère qu'il peut être impliqué dans le contrôle du caractère (par ex. les gènes de la protéine du lait dans la production de protéines lactiques).

**ADN:** l'information génétique dans un génome est codée dans l'acide désoxyribonucléique (ADN), qui est stocké à l'intérieur du nucléus d'une cellule. L'ADN a deux brins structurés en double hélice, constituée de sucre (déoxyribose), de phosphate et de quatre bases chimiques – les nucléotides: adénine (A), guanine (G), cytosine (C) et thymine (T). Une A sur un brin s'accouple toujours avec une T sur l'autre brin par deux liens d'hydrogène, tandis qu'une C s'accouple toujours avec une G par trois liens d'hydrogène. Les deux brins sont donc complémentaires.

**ADN complémentaire (ADNc):** séquences d'ADN générées par la transcriptase inverse de séquences d'ARNm. Ce type d'ADN inclut les exons et les régions non traduites aux fins 5' et 3' des gènes, mais n'inclut pas les introns.

**Marqueur génétique:** un polymorphisme d'ADN pouvant être facilement détecté par l'analyse moléculaire ou phénotypique. Le marqueur peut se trouver dans un gène ou dans l'ADN sans aucune fonction connue. Les segments d'ADN, qui se trouvent proches l'un de l'autre sur un chromosome, ayant tendance à être hérités ensemble, les marqueurs sont souvent utilisés en tant que moyens indirects pour tracer le modèle d'hérédité d'un gène qui n'a pas encore été identifié, mais dont l'emplacement approximatif est connu.

**Haplotype:** une contraction de la phrase «génotype haploïde». Il s'agit de la constitution

génétique d'un chromosome individuel. Dans le cas des organismes diploïdes, l'haplotype contiendra un membre de la paire d'allèle pour chaque site. Il peut se référer à un ensemble de marqueurs (par ex. les polymorphismes d'un seul nucléotide – SNP) qui sont statistiquement associés sur un chromosome simple. Grâce à cette connaissance, l'identification de quelques allèles d'un bloc d'haplotypes peut identifier sans aucune ambiguïté tous les autres sites polymorphiques de cette région. Cette information est de très grande valeur pour l'étude de la génétique à la base des caractères complexes.

**Groupe de liaison:** l'association de gènes et/ou de marqueurs qui sont les uns à côté des autres sur un chromosome. Les gènes et les marqueurs liés tendent à être hérités ensemble.

**Déséquilibre de liaison:** il s'agit d'un terme utilisé dans l'étude de la génétique des populations pour l'association non aléatoire d'allèles sur deux ou plus loci, pas nécessairement sur le même chromosome. Ce n'est pas la même chose qu'un groupe de liaison qui décrit l'association de deux ou plus loci sur un chromosome avec une recombinaison limitée entre eux. Le déséquilibre de liaison décrit une situation dans laquelle certaines combinaisons d'allèles ou de marqueurs génétiques se produisent plus ou moins fréquemment dans une population que l'on s'attendrait d'une formation aléatoire d'haplotypes des allèles basés sur leurs fréquences. Le déséquilibre de liaison est déterminé par les interactions de la valeur adaptative entre les gènes ou par de tels processus non adaptatifs, comme la structure de la population, la consanguinité et les effets stochastiques. Pour ce qui concerne la génétique de la population, le déséquilibre de liaison semble caractériser la distribution des haplotypes sur deux ou plus loci.

**Technologie des puces à ADN:** une nouvelle façon d'étudier les manières dont de nombreux gènes interagissent entre eux et dont un réseau réglementaire de cellules contrôle des vastes batteries

**Cadre 79 suite****Glossaire: marqueurs moléculaires**

de gènes simultanément. La méthode utilise un robot pour appliquer de façon très précise de minuscules gouttelettes contenant l'ADN fonctionnel sur des lamelles de verre. Les chercheurs attachent ensuite des étiquettes fluorescentes à l'ARNm ou à l'ADNc de la cellule qu'ils étudient. On laisse que les sondes étiquetées se lient aux brins d'ADNc sur les lamelles. Les lamelles sont posées dans un microscope de scannage qui peut mesurer la luminosité de chaque point fluorescent; la luminosité révèle la quantité

d'ARNm spécifique présente, qui est un indicateur de l'ampleur de son activité.

**Amorce:** une séquence courte d'oligonucléotide (simple brin) utilisée dans la réaction de polymérisation en chaîne (PCR).

**ARN:** l'acide ribonucléique est un acide nucléique simple brin qui est composé de trois bases sur les quatre présentes dans l'ADN (A, C et G). T est toutefois remplacée par l'uracile (U).

**Références**

- Aebersold, R. et Mann, M. 2003. Mass spectrometry-based proteomics. *Nature*, 422 (6928): 198–207. Review.
- Ajmone-Marsan, P., Negrini, R., Milanesi, E., Bozzi, R., Nijman, I.J., Buntjer, J.B., Valentini, A. et Lenstra, J.A. 2002. Genetic distances within and across cattle breeds as indicated by biallelic AFLP markers. *Animal Genetics*, 33: 280–286.
- Akey, J.M., Zhang, G., Zhang, K., Jin, L. et Shriver, M.D. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, 12(12): 1805–14.
- Aravin, A. et Tuschl, T. 2005. Identification and characterization of small RNAs involved in RNA silencing. *Febs Letters*, 579(26): 5830–40.
- Bachem, C.W.B., Van der Hoeven, R.S., De Bruijn, S.M., Vreugdenhil, D., Zabeau, M. et Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analyses of gene expression during potato tuber development. *The Plant Journal*, 9: 745–753.
- Bamshad, M. et Wooding, S.P. 2003. Signatures of natural selection in the human genome. *Nature Reviews Genetics*, 4(2): 99–111. Review.
- Baumung, R., Simianer, H. et Hoffmann, I. 2004. Genetic diversity studies in farm animals – a survey, *Journal of Animal Breeding and Genetics*, 121: 361–373.
- Beaumont, M.A. et Balding, D.J. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13(4): 969–80.
- Beja-Pereira, A., Alexandrino, P., Bessa, I., Carretero, Y., Dunner, S., Ferrand, N., Jordana, J., Laloe, D., Moazami-Goudarzi, K., Sanchez, A. et Cañon, J. 2003. Genetic characterization of southwestern European bovine breeds: a historical and biogeographical reassessment with a set of 16 microsatellites. *Journal of Heredity*, 94: 243–50.
- Berthier, D., Quere, R., Thevenon, S., Belemsaga, D., Piquemal, D., Marti, J. et Maillard, J.C. 2003. Serial analysis of gene expression (SAGE) in bovine trypanotolerance: preliminary results. *Genetics Selection Evolution*, 35 (Suppl. 1): S35–47.

## PARTIE 4

- Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M. et Snyder, M. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science*, 306: 2242–2246.
- Black, W.C., Baer, C.F., Antolin, M.F. et DuTeau, N.M. 2001. Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, 46: 441–469.
- Bruford, M.W., Bradley, D.G. et Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4: 900–910.
- Buntjer, J.B., Otsen, M., Nijman, I.J., Kuiper, M.T. et Lenstra, J.A. 2002. Phylogeny of bovine species based on AFLP fingerprinting. *Heredity*, 88: 46–51.
- Campbell, D. et Bernatchez, L. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, 21(5): 945–56.
- Cañon, J., García, D., García-Atance, M.A., Obexer-Ruff, G., Lenstra, J.A., Ajmone-Marsan, P., Dunner, S. et le consortium ECONOGENE. 2006. Geographical partitioning of goat diversity in Europe and the Middle East. *Animal Genetics*, 37: 327–334.
- Chen, S.Y., Su, Y.H., Wu, S.F., Sha, T. et Zhang, Y.P. 2005. Mitochondrial diversity and phylogeographic structure of Chinese domestic goats. *Molecular Phylogenetics and Evolution*, 37: 804–814.
- Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. et Nielsen, R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15: 1496–1502.
- Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibe, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. et Georges, M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, 38: 813–818.
- De Marchi, M., Dalvit, C., Targhetta, C. et Cassandro, M. 2006. Assessing genetic diversity in indigenous Veneto chicken breeds using AFLP markers. *Animal Genetics*, 37: 101–105.
- Donson, J., Fang, Y., Espiritu-Santo, G., Xing, W., Salazar, A., Miyamoto, S., Armendarez, V. et Volkmuth, W. 2002. Comprehensive gene expression analysis by transcript profiling. *Plant Molecular Biology*, 48: 75–97.
- Excoffier, L., Smouse, P.E. et Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131: 479–491.
- Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisio, S., Simon, P., Wagenaar, D., Vilki, J. et Georges, M. 2002. Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics*, 161: 275–287.
- Fields, S. et Song, O. 1989. A novel genetic system to detect protein–protein interactions. *Nature*, 340: 245–246.
- Freeman, A.R., Bradley, D.G., Nagda, S., Gibson, J.P. et Hanotte, O. 2006. Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Animal Genetics*, 37: 1–9.

- Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L. et Feldman, M.W. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139: 463–471.
- Goldstein, D.B. et Schlötterer, C. 1999. *Microsatellites: evolution and applications*. New York, Etats-Unis d'Amérique. Oxford University Press.
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. et Snell, R. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, 12: 222–231.
- Guo, J., Du, L.X., Ma, Y.H., Guan, W.J., Li, H.B., Zhao, Q.J., Li, X. et Rao, S.Q. 2005. A novel maternal lineage revealed in sheep (*Ovis aries*). *Animal Genetics*, 36: 331–336.
- Haley, C. et de Koning, D.J. 2006. Genetical genomics in livestock: potentials and pitfalls. *Animal Genetics*, 37(Suppl 1): 10–12.
- Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y. et Hill, E.W. 2002. African pastoralism: genetic imprints of origins and migrations. *Science*, 296: 336–339.
- Hayes, B.J., Visscher, P.M., McPartlan, H.C. et Goddard, M.E. 2003. A novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13: 635–643.
- Hill, E.W., O'Gorman, G.M., Agaba, M., Gibson, J.P., Hanotte, O., Kemp, S.J., Naessens, J., Coussens, P.M. et MacHugh, D.E. 2005. Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics. *Veterinary Immunology and Immunopathology*, 105: 247–258.
- Hill, W.G. 1981. Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, 38: 209–216.
- Hillel, J., Groenen, M.A., Tixier-Boichard, M., Korol, A.B., David, L., Kirzhner, V.M., Burke, T., Barre-Dirie, A., Crooijmans, R.P., Elo, K., Feldman, M.W., Freidlin, P.J., Maki-Tanila, A., Oortwijn, M., Thomson, P., Vignal, A., Wimmers, K. et Weigend, S. 2003. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics Selection Evolution*, 35: 533–557.
- Hood, L., Heath, J.R., Phelps, M.E. et Lin, B. 2004. Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306: 640–643.
- Ibeagha-Awemu, E.M., Jann, O.C., Weimann, C. et Erhardt, G. 2004. Genetic diversity, introgression and relationships among West/Central African cattle breeds. *Genetics Selection Evolution*, 36: 673–690.
- Jarne, P. et Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. *Trees*, 11: 424–429.
- Joshi, M.B., Rout, P.K., Mandal, A.K., Tyler-Smith, C., Singh, L. et Thangaraj, K. 2004. Phylogeography and origin of Indian domestic goats. *Molecular Biology and Evolution*, 21: 454–462.
- Joost, S. 2006. *The geographical dimension of genetic diversity. A GIScience contribution for the conservation of animal genetic resources*. École Polytechnique Fédérale de Lausanne, Suisse. (thèse de PhD)
- Kayser, M., Brauer, S. et Stoneking, M. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular Biology and Evolution*, 20: 893–900.
- Lai, S.J., Liu, Y.P., Liu, Y.X., Li, X.W. et Yao, Y.G. 2006. Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. *Molecular Phylogenetics and Evolution*, 38: 146–154.

## PARTIE 4

- Lan, L., Chen, M., Flowers, J.B., Yandell, B.S., Stapleton, D.S., Mata, C.M., Ton-Keen Mui, E., Flowers, M.T., Schueler, K.L., Manly, K.F., Williams, R.W., Kendzierski, C. et Attie, A.D. 2006. Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genetics*, 2: 51–61.
- Liang, P. et Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257: 967–997.
- Liu, Y.P., Wu, G.S., Yao, Y.G., Miao, Y.W., Luikart, G., Baig, M., Beja-Pereira, A., Ding, Z.L., Palanichamy, M.G. et Zhang, Y.P. 2006. Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution*, 38: 12–19.
- Lueking, A., Possling, A., Huber, O., Beveridge, A., Horn, M., Eickhoff, H., Schuchardt, J., Lehrach, H. et Cahill, D.J. 2003. A nonredundant human protein chip for antibody screening and serum profiling. *Molecular and Cellular Proteomics*, 2: 1342–1349.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S. et Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4: 981–994.
- Luikart, G., Gielly, L., Excoffier, L., Vigne, J.D., Bouvet, J. et Taberlet, P. 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Science USA*, 98: 5927–5932.
- Mburu, D.N., Ochieng, J.W., Kuria, S.G., Jianlin, H. et Kaufmann, B. 2003. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Animal Genetics*, 34(1): 26–32.
- McPherron, A.C. et Lee, S.J. 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Science USA*, 94: 12457–12461.
- Negrini, R., Milanesi, E., Bozzi, R., Pellecchia, M. et Ajmone-Marsan, P. 2006. Tuscany autochthonous cattle breeds: an original genetic resource investigated by AFLP markers. *Journal of Animal Breeding and Genetics*, 123: 10–16.
- Nei, M. 1972. Genetic distance between populations. *The American Naturalist*, 106: 283–292.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89: 583–590.
- Nei, M. et Roychoudhury, A.K. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics*, 76: 379–390.
- Nei, M., Tajima, F. et Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, 19: 153–170.
- Nielsen, R. et Signorovitch, J. 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, 63: 245–55.
- Nijman, I.J., Otsen, M., Verkaar, E.L., de Ruijter, C. et Hanekamp, E. 2003. Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity*, 90: 10–16.
- Pariset, L., Cappuccio, I., Joost, S., D'Andrea, M.S., Marletta, D., Ajmone Marsan, P., Valentini, A. et le consortium ECONOGENE. 2006. Characterization of single nucleotide polymorphisms in sheep and their variation as an evidence of selection. *Animal Genetics*, 37: 290–292.

- Pritchard, J.K., Stephens, M. et Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.
- Rabie, T.S., Crooijmans, R.P., Bovenhuis, H., Vereijken, A.L., Veenendaal, T., van der Poel, J.J., Van Arendonk, J.A., Pakdel, A. et Groenen, M.A. 2005. Genetic mapping of quantitative trait loci affecting susceptibility in chicken to develop pulmonary hypertension syndrome. *Animal Genetics*, 36: 468–476.
- Saitou, N. et Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406–425.
- Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S. et Altshuler, D.; International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409: 928–933.
- SanCristobal, M., Chevalet, C., Peleman, J., Heuven, H., Brugmans, B., van Schriek, M., Joosten, R., Rattink, A.P., Harlizius, B., Groenen, M.A., Amigues, Y., Boscher, M.Y., Russell, G., Law, A., Davoli, R., Russo, V., Desautes, C., Alderson, L., Fimland, E., Bagga, M., Delgado, J.V., Vega-Pla, J.L., Martinez, A.M., Ramos, M., Glodek, P., Meyer, J.N., Gandini, G., Matassino, D., Siggins, K., Laval, G., Archibald, A., Milan, D., Hammond, K., Cardellino, R., Haley, C. et Plastow, G. 2006b. Genetic diversity in European pigs utilizing amplified fragment length polymorphism markers. *Animal Genetics*, 37: 232–238.
- Sauer, S., Lange, B.M.H., Gobom, J., Nyarsik, L., Seitz, H. et Lehrach, H. 2005. Miniaturization in functional genomics and proteomics. *Nature Reviews Genetics*, 6: 465–476.
- Sodhi, M., Mukesh, M., Mishra, B.P., Mitkari, K.R., Prakash, B. et Ahlawat, S.P. 2005. Evaluation of genetic differentiation in *Bos indicus* cattle breeds from Marathwada region of India using microsatellite polymorphism. *Animal Biotechnology*, 16: 127–137.
- Storz, G., Altuvia, S. et Wassarman, K.M. 2005. An abundance of RNA regulators. *Annual Review of Biochemistry*, 74: 199–217.
- Sunnucks, P. 2001. Efficient genetic markers for population biology. *Tree*, 15: 199–203.
- Syvänen, A.C. 2001. Accessing genetic variation by genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2: 930–941.
- Takezaki, N. et Nei, M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144: 389–399.
- Tapio, M., Tapio, I., Grislis, Z., Holm, L.E., Jeppsson, S., Kantanen, J., Miceikiene, I., Olsaker, I., Viinalass, H. et Eythorsdottir, E. 2005. Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Molecular Ecology*, 14: 3951–3963.

## PARTIE 4

- Tilquin, P., Barrow, P.A., Marly, J., Pitel, F., Plisson-Petit, F., Velge, P., Vignal, A., Baret, P.V., Bumstead, N. et Beaumont, C. 2005. A genome scan for quantitative trait loci affecting the *Salmonella* carrier-state in the chicken. *Genetics Selection Evolution*, 37: 539–61.
- Troy, C.S., MacHugh, D., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykesk, B.C. et Bradley D.G. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410: 1088–1091.
- Velculescu, V.E., Vogelstein, B. et Kinzler, K.W. 2000. Analyzing uncharted transcriptomes with SAGE. *Trends in Genetics*, 16: 423–425.
- Velculescu, V.E., Zhang, L., Vogelstein, B. et Kinzler, K.W. 1995. Serial analysis of gene expression. *Science*, 270: 484–487.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. et Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23: 4407–1444.
- Weir, B.S. et Basten, C.J. 1990. Sampling strategies for distances between DNA sequences. *Biometrics*, 46: 551–582.
- Weir, B.S. et Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38: 1358–1370.
- Wienholds, E. et Plasterk, R.H. 2005. MicroRNA function in animal development. *FEBS Letters*, 579: 5911–5922.
- Wong, G.K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruan, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L.W., Yang, S.P., Warren, W.C., Wilson, R.K., Brandstrom, M., Ellegren, H., Crooijmans, R.P., van der Poel, J.J., Bovenhuis, H., Groenen, M.A., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J.R., Rogers, S., Walker, B.A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S.J., Zhou, H., Hocking, P.M., Morrice, D., de Koning, D.J., Law, A., Bartley, N., Burt, D.W., Hunt, H., Cheng, H.H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M.T., Andersson, B., Webber, C., Ponting, C.P., Overton, I.M., Boardman, P.E., Tang, H., Hubbard, S.J., Wilson, S.A., Yu, J., Wang, J., Yang, H.; International Chicken Polymorphism Map Consortium. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, 432: 717–722.
- Zhu, H., Bilgin, M. et Snyder, M. 2003. Proteomics. *Annual Review of Biochemistry*, 72: 783–812.