

Formation " Génomique et élevage "

Rennes

21-23 novembre 2005

Glossaire de génétique moléculaire (animale)

Version 1.1

Français (anglais)

Les mots soulignés sont des entrées de ce glossaire

ADN (DNA) : acide désoxyribonucléique, support matériel de l'hérédité, un des constituants des chromosomes.

Macromolécule formée d'une succession linéaire (brin) de molécules élémentaires, les nucléotides. La succession des nucléotides, identifiés par la base (adénine, guanine, cytosine, thymine) contenue dans chacun d'eux, constitue la séquence de l'ADN. L'ADN est formé de deux brins complémentaires enroulés en hélice (double hélice), ce qui lui permet de se dupliquer en deux molécules identiques entre elles et identiques à la molécule mère lors du phénomène de réplication lors de la mitose.

La longueur d'une molécule d'ADN est proportionnelle au nombre de nucléotides ou de paires de bases (du fait de la double hélice) qui la constituent ; d'où sa mesure exprimée en nombre de paires de base (base pair ou bp) ou en multiples de ce nombre (kb ou Mb).

Dans les cellules eucaryotes (pourvues de noyaux), l'ADN peut être **nucléaire** (c'est-à-dire contenu dans le noyau, sous forme de chromosomes) ou **mitochondrial** (c'est-à-dire contenu dans les organites cellulaires que sont les mitochondries).

ADNc ou ADN complémentaire (cDNA, complementary DNA) : ADN synthétisé par une transcription inverse à partir d'ARN dont il est une copie.

Il offre l'avantage d'être plus stable que la molécule d'ARNm et de pouvoir être stocké, multiplié (voir clone et banque d'ADN) et séquencé.

Alignement (alignement) : processus par lequel deux séquences (succession de molécules élémentaires identifiées par des lettres) sont comparées afin d'identifier leur ressemblance quantifiée par le pourcentage d'identités ou la similarité.

Allèle (allele) : l'une des séquences alternatives possible (variant génétique) de l'ADN à une position donnée (locus). L'ensemble des allèles possibles en un locus traduit le polymorphisme de ce locus. Chez les organismes diploïdes, un allèle est hérité du père, l'autre de la mère.

Amorce (primer) : court fragment d'ADN (oligonucléotide) qui, hybridé avec une molécule simple brin d'acide nucléique, permet à une enzyme (polymérase) d'initier la synthèse du brin complémentaire. Utilisée en particulier dans des réactions de PCR et dans la synthèse d'ADNc.

Annotation (annotation) : l'annotation d'une séquence d'ADN ou d'un gène consiste à renseigner sa (ou ses) fonction(s) biologique(s) ou à la prédire à l'aide de diverses informations disponibles.

ARN ou acide ribonucléique (RNA, ribonucleic acid) : Macromolécule formée d'une succession linéaire (brin) de molécules élémentaires (nucléotides) qui permet de transférer et de traiter l'information génétique dans la cellule.
Selon leur fonction, on distingue différents ARN dont les ARN messagers.

ARN messenger ou ARNm (messenger RNA ou mRNA) : molécule d'ARN simple brin synthétisée à partir d'une séquence d'ADN (transcription) correspondant à un gène. Après différentes étapes de maturation elle permettra la synthèse d'une protéine (traduction) correspondant au gène de départ.

BAC ou chromosome artificiel bactérien (Bacterial Artificial Chromosome) : chromosome de bactérie modifié pour contenir un grand, (plusieurs centaines de **kb**) fragment d'ADN génomique exogène (de l'espèce à étudier), que l'on pourra alors stocker et multiplier sous forme d'un clone bactérien afin de l'étudier (cf. banque d'ADN).

Banque (d'ADN) ((DNA) library) : collection de fragments d'ADN, individuellement intégrés dans une molécule vecteur (plasmide, BAC,...) puis introduits artificiellement dans des cellules hôtes (bactéries par exemple) qui deviennent des lieux de stockage et de multiplication de ces fragments. Chaque cellule produira par multiplication asexuée un ensemble de cellules filles (clone) portant toutes le même fragment d'ADN exogène. L'ensemble de ces clones (chacun correspondant à l'un des fragments d'ADN de la collection) constitue la banque d'ADN.

Base (base) : un des constituants des nucléotides, eux-mêmes constituants des acides nucléiques (ADN et ARN). C'est sur cette molécule que repose l'identification du nucléotide et par suite celle de la succession des nucléotides d'un fragment d'acide nucléique (séquence).

Carte génétique (genetic map) :

Sens général : représentation graphique de la position de fragments d'ADN (gènes par exemple) les uns par rapport aux autres sur un génome.

Différentes méthodes de construction sont utilisées et donnent naissance à différentes cartes de résolutions différentes pour le même génome, exprimées en différentes unités.

Sens particulier (le plus généralement utilisé) ; on parle aussi de carte de liaison (linkage map) : elle prend en compte des petites régions d'ADN polymorphes (marqueurs) et repose sur la fréquence de transmission conjointe d'allèles déterminés aux marqueurs d'un parent et ses descendants. Cette fréquence est d'autant plus grande que la distance séparant les deux marqueurs est petite. L'aléatoire provient des remaniements chromosomiques possibles lors de la méiose, qui sont d'autant plus fréquents que la distance qui les sépare est grande. Les distances génétiques sont alors exprimées en centimorgan (cM).

Carte d'hybrides d'irradiation (ou irradiés) ou carte RH (radiation hybrid map (RH map)) : elle repose sur la fréquence de cassures chromosomiques induites artificiellement, observées chez des cellules hybrides d'irradiation (obtenues par la fusion de cellules irradiées de l'espèce étudiée avec des cellules de hamster). Deux petites régions d'ADN initialement localisés à proximité l'un de l'autre ont plus de chance d'être simultanément retrouvés dans les mêmes hybrides que deux marqueurs éloignés. Les distances se mesurent en centiray (cR)

(dose d'irradiation utilisée). L'avantage de la cartographie RH est que les régions cartographiées n'ont pas besoin d'être polymorphes.

Carte physique (physical map) : à sa plus grande résolution c'est la séquence complète d'un génome. Au préalable, c'est l'arrangement ordonné de grands fragments d'ADN identifiés (clone BAC), dont l'ensemble constitue l'ensemble du génome étudié. L'arrangement repose sur les similitudes partielles entre fragments (fragments chevauchants). Un ensemble de fragments chevauchants porte le nom de contig. Les distances sont exprimées en megabase (Mb).

Caryotype (caryotype) : représentation et ordonnancement (selon leur taille) des chromosomes du noyau d'une cellule, obtenus par microscopie ou microphotographie.

Clonage (cloning) : Méthode de multiplication cellulaire in vitro par reproduction asexuée aboutissant à la formation d'un ensemble de cellules identiques (clones) en particulier pour leur contenu en ADN.

Par extension, le clonage consiste en l'insertion d'un fragment d'ADN (au moyen d'un vecteur) dans une cellule hôte, généralement une bactérie. La culture de cette cellule et la purification ultérieure de l'ADN exogène permettent respectivement de stocker et de produire des quantités importantes du fragment d'ADN cloné qui peut alors être étudié.

Clonage positionnel (positional cloning) : méthode qui utilise les techniques de cartographie génique pour identifier un gène dont le polymorphisme est responsable de la variabilité d'un caractère (ou la mutation responsable d'une maladie): C'est une méthode progressive : elle permet d'abord l'identification de la région du génome qui contient le gène recherché, puis la réduction progressive de la taille de la région jusqu'à l'identification du gène lui-même.

Clone (clone) : ensemble de cellules provenant par divisions (reproduction asexuée) d'une cellule unique. Elles sont donc (considérées comme) identiques en particulier elles renferment les mêmes molécules d'ADN, originelles ou transférées artificiellement.

Par extension, on parle de clone d'un fragment d'ADN particulier pour nommer le clone issu de la cellule dans lequel cet ADN a été artificiellement introduit. On utilise classiquement des clones d'ADNc, comportant des ADNc d'une espèce donnée ; des clones de BAC, contenant de grands fragments d'ADN d'une espèce donnée.

Code génétique (genetic code) : le code génétique est le système de correspondance (code) permettant au message génétique porté par l'ADN (alphabet de 4 lettres correspondant aux 4 bases des nucléotides) et transcrit en ARN d'être traduit en protéine (alphabet de 20 lettres correspondant aux acides aminés) par une cellule. A chaque séquence de trois bases consécutives (codon) portées par l'ARN messenger, correspond un acide aminé donné et un seul. C'est le code génétique qui permet donc la traduction des messages codés dans le génome en protéines ayant des fonctions bien précises. Il y a 64 combinaisons de 3 lettres prises parmi 4. Un même acide aminé peut donc être codé par plusieurs codons différents (codons synonymes) : on parle de "code dégénéré". Certains codons ont des significations particulières pour la machinerie cellulaire de la synthèse protéique, en particulier 3 d'entre eux sont des codons « stop » et déterminent l'arrêt de la traduction.

Codon (codon) : séquence de 3 bases consécutives de l'ARN messenger auquel correspondra un acide aminé dans la protéine traduite de cet ARNm.

Centimorgan (cM) : unité de mesure de distance sur la carte génétique.

Contig (contig) : ensemble de fragments d'ADN contigus (en fait, plutôt chevauchants) et ordonnancés, utilisé pour reconstituer la carte physique d'un chromosome ou du génome. Ces fragments ont été clonés au préalable.

Délétion (deletion) : Perte d'une partie d'une molécule d'ADN.

Dénaturation (denaturation) : dissociation des deux brins complémentaires d'une molécule d'ADN. S'effectue expérimentalement par l'action de la chaleur ou d'agents chimiques, par exemple.

Déséquilibre de liaison (linkage disequilibrium) : situation dans laquelle deux allèles correspondant à deux sites chromosomiques différents (loci) sont transmis ensemble d'une génération à l'autre plus souvent que le hasard ne le laisse prévoir. On parle aussi de **déséquilibre d'association gamétique**.

Distance génétique (genetic distance) :

Entre loci (ou entre gènes) : mesure de la distance séparant deux positions (loci) sur une molécule d'ADN (voir cartes génétiques).

Elle s'exprime en centimorgan (cM) lorsqu'elle est évaluée à partir des la fréquences de transmission conjointe d'allèles à chacun des loci d'un parent à ses descendants (carte de liaison). Différentes métriques de distances sont utilisées (Kosambi et Haldane) et conduisent bien sûr à des valeurs différentes.

La distance est exprimée en centiray (cR) lorsqu'elle est évaluée à partir de cellules hybrides irradiées (carte d'hybrides d'irradiation).

Elle est exprimée en multiple du nombre de bases (kb, Mb) lorsqu'elle repose sur une distance physique sur la molécule d'ADN (carte physique).

Entre populations : mesure le degré de ressemblance entre les génomes de différentes populations. Différentes distances peuvent être associées à deux mêmes populations selon les critères pris en compte (nombre de loci, types de loci, ...) et la métrique de distance utilisée.

Dominance (dominance) : situation dans laquelle un allèle (variant génétique) à un locus (région du génome) masque la présence ou l'expression d'un autre allèle (au même locus). En d'autres termes, les génotypes homozygotes pour l'allèle dominant et hétérozygotes (avec l'allèle dominant) ne peuvent pas être distingués.

Par extension, toute situation dans laquelle les effets de deux allèles ne sont pas strictement additifs (la valeur des hétérozygotes est différente de la valeur moyenne des deux homozygotes).

Épissage (splicing) : processus de maturation de l'ARN après sa synthèse à partir de l'ADN (transcription) conduisant à l'ARN messager. Cette étape élimine certaines parties de la séquence du gène appelées introns et accole les parties restantes appelées exons.

e-QTL (QTL d'expression) (e-QTL, expression QTL) : région du génome dont le polymorphisme explique une part importante de la variabilité d'une quantité d'ARN messager. C'est un QTL pour un caractère d'expression génique.

EST ou étiquette de séquence exprimée (EST, Expressed Sequence Tag) : séquence partielle de l'extrémité de l'ADNc correspondant à un gène exprimé.

Exon (exon) : partie d'un gène qui sera représentée dans l'ARNm mature, comportant en particulier les régions qui seront traduites en acides aminés (protéines), mais pas uniquement. Les exons des extrémités d'un ARNm comportent des régions (séquences) non traduites, nommées UTR (de l'anglais *UnTranslated Region*).

Facteur de transcription (transcription factor) : protéine qui régule la transcription d'un gène en se fixant sur une région ADN appelée promoteur. Un facteur de transcription donné se fixe sur une courte séquence du promoteur qui lui est spécifique (site de liaison ou *binding site* en anglais).

Famille de gènes (gene family): ensemble de gènes ayant de grandes ressemblances fonctionnelles et structurelles.

Gène (gene) : segment d'ADN situé à un endroit précis (locus) sur un chromosome et porteur d'une information génétique.

Gènes orthologues (orthologs) : gènes d'espèces différentes dont les séquences dérivent d'un même gène ancestral et ont divergé à la suite d'un évènement de spéciation. Ces séquences se ressemblent (homologie). Ils ont généralement la même fonction.

Gènes paralogues (paralogs) : gènes d'une même espèce dont les séquences résultent de la duplication d'un même gène ancestral et donc se ressemblent.

Gène majeur (major gene) : gène ayant un fort effet (> 3 écarts types phénotypiques) sur la variabilité d'un caractère.

Génome (genome) : patrimoine génétique d'un organisme constitué de molécules d'ADN, transmis totalement (reproduction asexuée) ou partiellement (reproduction sexuée) à chacun de ses descendants.

Génomique (genomics) : discipline qui étudie les génomés dans leur ensemble. Elle se compose de deux volets complémentaires. Le premier correspond à l'analyse structurale (structure physique et organisation) du génome (molécules d'ADN). Le second nommé **génomique fonctionnelle** (functional genomics), concerne la fonction des gènes, la régulation de leur expression et leurs interactions ; il s'intéresse aux molécules d'ARNm et/ou aux protéines résultant de l'expression des gènes.

Le terme de **génomique structurale** est généralement utilisé pour la structure physique et l'organisation des protéines. La génomique fonctionnelle est parfois aussi appelée – à tort – « post génomique » pour signifier que cette discipline se développe après les grands travaux de séquençage des génomes. L'ensemble des domaines d'étude de la génomique est en fait concomitant pour la plupart des espèces.

Haplotype (haplotype) : combinaison particulière de variants génétiques (allèles) à des positions (loci) différentes.

Homologie (homology) : ressemblance entre molécules (en particulier l'ADN) parce qu'elles proviennent par copies (obtenues au cours des générations) d'une même molécule ancestrale. L'homologie est mise en évidence en recherchant des similitudes entre les séquences.

Hybridation moléculaire (molecular hybridization) : c'est l'appariement par complémentarité des bases de deux chaînes d'acides nucléiques simple brin pour former des doubles brins. Les brins de la double hélice d'ADN ont la capacité de reformer spontanément

cette double hélice dès qu'ils se retrouvent face au brin complémentaire : les quatre molécules élémentaires de l'ADN (nucléotides) ont la particularité de s'unir deux à deux par des liaisons hydrogènes : l'adénine (A) avec la thymine (T), la cytosine (C) avec la guanine (G). Expérimentalement, deux brins complémentaires peuvent être séparés l'un de l'autre (on parle de **dénaturation**) par l'action de la chaleur ou d'agents chimiques, par exemple, et s'hybrider à nouveau dans des conditions favorables (réversion donc possible entre états dénaturé et renaturé).

Hybrides d'irradiation (radiation hybrid, **RH**) : cellules obtenues par la fusion de cellules d'une espèce fortement irradiée aux rayons X, de façon à induire des cassures dans les chromosomes, avec des cellules de hamster. Les cellules hybrides contiennent les chromosomes de hamster qui ont intégré aléatoirement des fragments d'ADN de l'autre espèce. Les cellules hybrides sont mises en culture et conduisent à des clones cellulaires dans lesquels toutes les cellules contiennent les mêmes molécules d'ADN (en particulier les mêmes fragments d'ADN de l'espèce de départ).

Identité (identity) : état de nucléotides ou d'acides aminés invariants (même molécule) entre deux séquences d'acides nucléiques ou de protéines.

Insert (insert) : séquence d'ADN étranger introduite dans une molécule d'ADN donnée, un plasmide ou un BAC par exemple.

Introgression (introgression) : action de transférer une partie de matériel génétique d'une population (donneuse) dans une autre (receveuse) par croisement entre ces deux populations puis une série de croisements en retour avec la population receveuse, en sélectionnant à chaque génération les individus porteurs de la partie de matériel génétique de la population donneuse.

Intron (intron) : Partie du gène éliminée au cours de la maturation de l'ARNm lors du processus appelé épissage. Elle est encadrée par les exons qui eux sont conservés dans la molécule d'ARN messager. Sa fonction reste généralement inconnue, bien que certains introns aient un rôle dans la régulation de l'expression du gène.

Liaison génétique (genetic linkage) : situation concernant deux positions sur le génome (loci) pour lesquelles les allèles sont préférentiellement transmis ensemble d'un parent à ses descendants parce qu'ils sont relativement proches sur une même molécule d'ADN (constituant un chromosome). La nature des allèles associés à chacun des loci constitue la phase de la liaison. Cette liaison physique peut toutefois être rompue par l'apparition de cassures chromosomiques conduisant à des recombinaisons entre fragments d'ADN, et une nouvelle phase de liaison. Ces phénomènes sont mis en évidence au moment de la méiose (division cellulaire conduisant à la formation des gamètes, avec passage d'un état diploïde (2n chromosomes) à l'état haploïde (n chromosomes)).

Groupe de liaison (linkage group) : ensemble des locus qui apparaissent liés (liaison génétique) par analyse de leur transmission héréditaire.

Locus (au pluriel, en toute rigueur, **loci**) (locus, loci) : région sur la molécule d'ADN (chromosome). La taille de cette région peut être quelconque (de 1 à des centaines de nucléotides).

Lod score (Lod score) : statistique de test utilisée dans les tests de liaison génétique, prenant en compte les probabilités d'occurrence des valeurs expérimentales obtenues sous différentes hypothèses. Ces probabilités sont appelées « vraisemblances »

kb ou kilobase (kb) : unité de longueur d'une molécule d'ADN correspondant à 10^3 bases (ou paires de bases)

Marquage (labelling) : introduction de nucléotides modifiés, ou modification chimique de certains nucléotides d'un acide nucléique afin de pouvoir le repérer. Les marquages classiques utilisent la radioactivité, la fluorescence ou la chimiluminescence.

Marqueur génétique ou marqueur (genetic marker): région polymorphe d'une molécule d'ADN. Les séquences de cette région pourront être différentes entre plusieurs individus.

marqueur anonyme (anonymous marker) : marqueurs correspondant à des séquences non traduites (les variations alléliques de ces marqueurs ne sont pas décelables qu'au niveau de la séquence).

Méthode du maximum de vraisemblance (Maximum of likelihood method) : méthode d'estimation statistique de paramètres qui repose sur la probabilité d'obtenir les valeurs observées dans une expérimentation en fonction de ces paramètres (vraisemblance des observations).

Métabolome (metabolome) : ensemble des métabolites produits par les cellules dans des conditions spécifiques.

Mb ou megabase (Mb) : unité de longueur d'une molécule d'ADN correspondant à 10^6 bases (ou paires de bases)

Microsatellite (microsatellite): petite portion d'ADN constituée d'une courte séquence d'ADN (1 à 4 nucléotides généralement) répétée un nombre variable de fois (généralement de 10 à 40). Chaque microsatellite correspond à un endroit (locus) unique dans le génome, parfaitement défini par les séquences uniques qui encadrent la répétition. Le polymorphisme très élevé (nombreux allèles) de ces loci est dû à des variations dans la longueur de la répétition : cette longueur identifie les allèles de chacun de ces loci. Les microsatellites sont répartis uniformément sur l'ensemble du génome et sont très utilisés comme marqueurs.

Mutation (mutation): modification de la séquence d'ADN par changement d'une (ou plusieurs) base(s) en une (ou plusieurs) autre(s) base(s).

Mutation ponctuelle (point mutation) : mutation portant sur une seule base (substitution, addition ou délétion). Voir snp, microsatellite.

Mutation silencieuse (silent mutation) : mutation qui ne modifie pas la traduction protéique d'un gène.

Nucléotide (nucleotide) : élément de base d'une molécule d'acide nucléique (ADN ou ARN), constitué d'un sucre (ribose pour l'ARN et désoxyribose pour l'ADN), d'un groupement phosphate et d'une base azotée. Il existe quatre nucléotides différents pour l'ADN, définis selon la base azotée qu'ils contiennent : adénine (A), thymine (T), guanine (G), cytosine (C) et quatre nucléotides différents pour l'ARN : uracile (U), guanine (G), cytosine (C), adénine (A). La séquence d'une molécule d'acide nucléique est définie par la succession des bases résultant de l'enchaînement des nucléotides de la molécule.

Oligonucléotide (oligonucleotide) : court fragment d'ADN, constitué de quelques (une ou plusieurs dizaines) molécules élémentaires (nucléotides)

Paire de bases (pb) (base pair, bp) : unité de mesure de la distance sur la carte physique de l'ADN (double brin), correspondant à un nucléotide (ou une base) pour le simple brin.

1 kb = 1 kilobase = 1 000 bases

1 Mb = 1 megabase = 1 000 000 de bases

PCR (réaction en chaîne de la polymérase) (PCR ou polymerase chain reaction) : réaction enzymatique in vitro consistant à dupliquer des fragments d'ADN et donc de générer un grand nombre de copies de ces fragments à partir d'une faible quantité initiale d'acide nucléique. On dit qu'on amplifie cette molécule initiale.

Pénétrance (penetrance) : c'est la probabilité d'avoir un phénotype particulier, connaissant le génotype au locus contrôlant majoritairement le caractère. Lorsqu'elle est différente de 1, elle traduit le fait que le génotype au locus considéré (la combinaison des deux allèles) ne s'exprime pas en un seul phénotype.

Phase ou phase de liaison génétique (phase) : caractérisation des allèles associés à chacun de deux loci génétiquement liés.

Phénotype (phenotype) : manifestation apparente de la constitution du génome sous la forme d'un trait morphologique, d'un syndrome clinique, d'une variation qualitative ou quantitative du produit final d'un gène (protéine). Le phénotype correspond à l'expression du génome modulée par les conditions environnementales.

Plasmide (plasmid): chez les bactéries, molécule d'ADN non chromosomique capable de se multiplier (répliquer) même sans division cellulaire et pouvant pénétrer dans une bactérie hôte dans certaines conditions. Cette molécule et ces propriétés en font un outil commun en génétique moléculaire ; c'est une des bases du clonage moléculaire.

Pléiotrope (pleiotope) : se dit d'un gène qui influence divers caractères.

Promoteur (promoter) : séquence d'ADN nécessaire à l'initiation de la transcription (copie d'une molécule d'ADN en ARN) d'un gène, située en amont de la partie qui sera transcrite.

Protéome (proteome) : ensemble des protéines exprimées par le génome d'une espèce donnée. Il peut aussi se comprendre comme l'ensemble de protéines présentes dans une cellule donnée à un instant donné.

Puce à ADN ou micro-réseau d'ADN (DNA chips ou DNA arrays-DNA microarrays) : support (membrane de nylon ou lame de verre) sur lequel on peut déposer des milliers de molécules d'ADN (banque d'ADN).

Ces molécules permettront de reconnaître et d'identifier des molécules semblables à elles-mêmes dans une population d'ADN ou d'ARN par hybridation moléculaire.

Les molécules immobilisées sur le support constituent les sondes ; la population de molécules à identifier constitue la cible. Dans le cas des études de génomique fonctionnelle, cette cible est constituée d'ARNm extraits de cellules, tissus ou organismes entiers, convertis en ADNc et marquées (cf. marquage) par un traceur (radioactif ou fluorescent).

QTL ou locus de caractère quantitatif (QTL, Quantitative Trait Locus) : région du génome dont le polymorphisme explique une part importante de la variabilité d'un caractère mesuré.

Recombinaison génétique (genetic recombination) : tout processus permettant d'obtenir un nouvel assemblage physique de fragments d'ADN. La recombinaison peut être naturelle, au moment de la méiose, et conduit à une nouvelle phase de liaison entre allèles aux loci concernés. Elle peut être expérimentale lorsque l'on construit une nouvelle molécule d'ADN à partir de molécules d'origine différentes, comme dans le cas du clonage moléculaire.

Réplication (replication) : mécanisme de copie d'une molécule d'ADN par synthèse de 2 molécules « filles » à partir d'une molécule « mère ». C'est ce qui permet de transmettre l'information génétique d'une cellule ou d'un organisme à sa descendance.

RFLP ou polymorphisme de fragments de restriction (RFLP, Restriction Fragment Length Polymorphism) : polymorphismes d'ADN détecté après morcellement de la molécule par l'action d'enzymes de restrictions (qui créent des coupures sur la molécule en des sites bien définis) et reconnaissance des fragments obtenus par hybridation moléculaire avec un fragment déterminé d'ADN marqué (on parle de sonde).

Région codante (coding sequence, CDS) : portions de la séquence d'ADN correspondant à un gène qui seront traduites en protéines.

SAM ou sélection assistée par marqueur (MAS, marker assisted selection) : utilisation des génotypes à certains marqueurs dans l'évaluation génétique d'un individu candidat à la sélection.

SAG ou sélection assistée par gène (GAS, gene assisted selection) : utilisation de génotypes à des gènes identifiés comme responsables de la variabilité d'un caractère observé dans l'évaluation génétique d'un individu candidat à la sélection.

SAGE (SAGE, Serial Analysis of Gene Expression) : méthode d'analyse quantitative de l'expression d'un grand nombre de gènes qui repose sur une technique de comptage des molécules d'ARN messenger présentes dans un échantillon de cellules, tissus ou organes.

Séquence (sequence) : succession de monomères dans un polymère. L'orientation de la séquence est définie par la synthèse du polymère. C'est aussi l'identification des monomères successifs.

Les séquences nucléiques (ADN ou ARN) sont des polymères de nucléotides. La séquence d'une molécule d'acide nucléique est définie par la succession des bases résultant de l'enchaînement des nucléotides de la molécule. Les molécules linéaires présentent une extrémité dite 5'P (en raison du nucléotide dont une fonction alcool, en position 5' du sucre, est estérifiée par un acide phosphorique) et une extrémité 3'OH. Orientation de 5' en 3'. Les séquences protéiques (= protéines) sont des polymères d'acides aminés. Orientation NH₂ terminal à COOH terminal.

Séquençage (sequencing) : détermination de l'ordre linéaire des composants d'une macromolécule. Pour l'ADN, identification des bases des nucléotides successifs de la molécule. Pour les protéines, identification de acides aminés successifs.

Séquence codante (coding sequence): Partie d'un gène qui définit directement la séquence en acides aminés de la protéine correspondante.

Similarité (similarity) : mesure de la ressemblance entre séquences protéiques ou nucléiques. Le degré de similarité entre deux séquences, quantifiée par un score, est fonction du pourcentage d'identités entre les séquences (même base à une position donnée). On parle aussi de **similitude**.

Site (site) : séquence d'ADN reconnue par une protéine qui pourra s'y fixer et avoir une action sur la molécule d'ADN.

Site de restriction (restriction site) : site de reconnaissance d'enzyme de restriction, capable de couper la molécule d'ADN.

SNP ou polymorphisme d'un seul nucléotide (SNP ou Single Nucleotide Polymorphism) : mutation ponctuelle isolée, se retrouvant le long de l'ensemble de l'ADN du génome toutes les 100 à 300 bases environ, et affectant au moins quelques pour cent de la population concernée. Très utilisé comme marqueur.

Sonde (probe) : fragment d'ADN identifié utilisé pour reconnaître une molécule similaire dans une population de molécules d'ADN ou d'ARN. La reconnaissance repose sur l'hybridation entre les molécules, rendue possible par la similitude des séquences. La sonde joue le rôle d'«hameçon» moléculaire.

Synténie (synteny) : groupe de gènes liés (groupe de liaison) dont l'ordre est conservé entre les génomes de deux espèces différentes.

Traduction (translation) : synthèse d'une chaîne polypeptidique (protéine) à partir d'une molécule d'ARN messenger.

Transcrit (transcript) : copie ARN d'un ADN, résultant de la réaction de transcription. S'emploie aussi pour désigner les ARN messagers.

Transcription (transcription) : synthèse d'une molécule d'ARN à partir d'une molécule d'ADN.

Transcriptome (transcriptome) : ensemble des ARN messagers transcrits (exprimés) à partir du génome d'une espèce donnée. Il peut aussi se comprendre comme l'ensemble de ARN messagers présents dans une cellule donnée à un instant donné.

Vecteur (vector) : molécule d'acide nucléique dans laquelle il est possible d'insérer des fragments d'acide nucléique étranger, pour ensuite les introduire et les maintenir dans une cellule hôte (utilisé dans les opérations de clonage moléculaire). Les vecteurs couramment utilisés sont les plasmides et les BAC dans lesquels sont insérés des fragments d'ADN de tailles de l'ordre de 1kb et de 100 kb, respectivement.

Vraisemblance (likelihood) : dans le domaine de l'estimation et des tests statistiques, la vraisemblance des valeurs expérimentales observées traduit la probabilité de les obtenir sous certaines hypothèses.

Sources de documentation, et pour en savoir plus

Infobiogen : Centre national de ressources informatiques appliquées à la génomique

http://www.infobiogen.fr/glossaire/glossaire_idx.php

INRIA Rhône-Alpes

<http://www-helix.inrialpes.fr/IMG/pdf/termesbiomol.pdf>

UMR Diversité et génomes des plantes cultivées – Agro.M, CIRAD, INRA, IRD - Montpellier

Glossaire de Biologie Moléculaire (B. Marin et C. Franche)

http://www.dgpc.org/ressources_globales/glossaire_biol_mol.html

Lycée Montesquieu - 95220 Herblay

<http://www.ac-versailles.fr/etabliss/herblay/GENETIQU/FICHES/sommaire.htm>

Mais date de 1998

L'encyclopédie en ligne (en anglais)

<http://en.wikipedia.org/wiki/Category:Genetics>