

Communication de Monsieur François Le Tacon



Séance du 20 février 2009



Comment et pourquoi séquencer des génomes et en particulier le génome humain ?

Introduction

La découverte de l'ADN et de son rôle a bouleversé les connaissances que nous pouvions avoir du monde vivant. Nous débutons cet exposé par l'histoire de la découverte de l'ADN. Le second chapitre est consacré à la découverte des modes de transcription et de traduction de l'ADN en protéines. Le troisième traite de la définition du génome et des méthodes de séquençages de génomes entiers. Nous abordons ensuite la question de l'annotation des génomes en général, puis de celui du génome humain. Enfin nous terminerons par les utilisations potentielles de la connaissance des génomes : thérapie génique et épigénétique ; agriculture : sélection génétique assistée par marqueurs et OGM ; compréhension des mécanismes de l'évolution.

Histoire de la découverte de l'ADN

En 1869, le Suisse Friedrich Miescher découvre à l'Université de Tübingen en Allemagne la nucléine, une substance non lipidique et non protéique, riche en phosphore. Vingt ans plus tard, un allemand, Richard Altmann, caractérise l'acide nucléique de la nucléine. En 1882, Albrecht Kossel, un autre allemand, isole un des quatre constituants de l'ADN, la guanine ou lettre G, qui était déjà connue dans le guano. En 1886, il isole la lettre A ou adénine. En 1900, dans le groupe de Kossel, Hermann Steudel caractérise la lettre T ou thymine. La même année, Steudel et Kossel caractérisent la lettre C ou cytosine. En 1891,

Kossel avait isolé de l'acide nucléique un sucre, que le suédois Olof Hammarsten décrira un peu plus tard comme un pentose. En 1908, deux américains, Phoebus Levene, d'origine russe, et Walter Jacobs dénomment ce sucre ribose, pour Rockfeller Institute for Biochemistry. Kessel avait également découvert dans le thymus un autre sucre qui en 1929 sera appelé 2-désoxyribose. L'uracile de l'ARN, ou lettre U, est découvert, toujours dans le groupe de Kessel, par Albert Ascoli. Enfin, Kessel avait découvert en 1884, les histones, protéines associées aux acides nucléiques.

En 1909, Levene et Jacobs démontrent qu'une base purique, comme les lettres A, T, G, ou C, s'associe à un ose, le pentose, et à un phosphate pour former un mononucléotide. Ils démontrent ensuite que ces quatre mononucléotides s'enchaînent par l'intermédiaire d'une liaison phosphodiester entre la fonction alcool primaire en position 5' du désoxyribose et la fonction alcool secondaire en position 3' de ce même désoxyribose. La chaîne de l'ADN était ainsi découverte. Mais Levene pensait que les quatre bases puriques de l'ADN, les quatre lettres A, T, G et C, étaient en nombre sensiblement égal, ce qui est inexact.

A la fin du XIX^e siècle, plusieurs savants, comme l'Américain Edmond Wilson en 1886, pensaient déjà que ces substances nucléiques jouaient un rôle dans la transmission des caractères héréditaires.

En 1928, Franklin Griffith, un médecin britannique, découvre que de l'information génétique peut être transmise à des bactéries vivantes par des bactéries au préalable détruites par la chaleur.

Thomas Hunt Morgan obtient le prix Nobel en 1933 pour avoir démontré que les gènes sont situés sur les chromosomes. Morgan crée un département de biologie au Caltech. Sept de ses élèves recevront le prix Nobel. Une vingtaine de prix Nobel ont été attribués jusqu'à maintenant pour des travaux sur l'ADN.

En 1944, Oswald Avery, Maclyn McCarty et Colin MacLeod identifient l'agent transformant de Griffith comme étant de l'ADN. En 1947, Alfred Mirsky démontre que les chromosomes sont constitués d'histone et d'ADN. En 1949, Edwin Chargaff démontre que les rapports A/T et G/C sont identiques et ont une valeur de 1.

Divers chercheurs se passionnent pour la structure de l'ADN et à partir de 1951, des progrès décisifs vont être réalisés. En 1951, au *King's College* à Londres, Maurice Wilkins, Alexandre Stokes et Herbert Urson, travaillant sur un échantillon purifié par le Suisse Rudolph Signer, obtiennent des images de diffraction X montrant que l'ADN pouvait être une structure hélicoïdale. La même année, Rosalind Franklin travaille sur ce même échantillon, et en

mai 1952 obtient une image en croix de l'ADN indiquant une structure en double hélice. Francis Crick et James Watson à Cambridge, utilisant les clichés du King's collège, construisent en 1952-1953 avec du fil de fer et des cartons la structure en double hélice. En avril 1953, paraissent dans la revue *Nature* trois articles différents : *Molecular structure of nucleic acids. A structure for deoxyribonucleic acid* de Watson et Crick décrivant la double hélice et surtout la possibilité de réplication, un second article de Franklin et Gosling donnant les images en croix de l'ADN en diffraction X, et un troisième de Wilkins, Seeds, Stokes et Wilson décrivant la structure de l'ADN d'*Escherichia coli*. Dans leur article de 1953, Watson et Crick remercient Rosalind Franklin et Wilkins pour leur contribution.

En 1962, Watson, Crick et Wilkins reçoivent le prix Nobel de physiologie et de médecine. La contribution de Rosalind Franklin, décédée en 1958, d'un cancer est totalement oubliée.

En 1959, Arthur Kornberg, un médecin et biochimiste américain et Severo Ochoa de Albornoz, un chercheur espagnol, obtiennent ce même prix Nobel de physiologie et de médecine pour leur découverte du mécanisme biologique de la synthèse de l'ADN.

La découverte des modes de transcription et de traduction de l'ADN en protéines

Les travaux de Barbara McClintock publiés en 1951, puis plus tard ceux de trois Français, Jérôme Monod, François Jacob et André Lwoff, prix Nobel de médecine en 1965 pour leur découverte des ARN messagers et de leurs rôles ont joué un rôle déterminant dans la compréhension des mécanismes de transcription et de traduction de l'ADN en protéines. Ils démontrent à l'Institut Pasteur que chez une bactérie, *Escherichia coli*, que les gènes sont d'abord *transcrits* en une molécule intermédiaire, qu'ils appellent d'abord facteur X. Ils démontrent ensuite que ce facteur X est un acide ribonucléique qu'ils appellent messager (ARNm). Une enzyme, l'ARN polymérase, se fixe sur l'ADN et synthétise les ARN messagers. Jacob, Monod et Lwoff ont démontré que la transcription d'un gène dépend de la présence de plusieurs protéines appelées régulatrices. Ces protéines se fixent sur des régions de l'ADN situées en amont du gène à transcrire appelées promoteurs. Les ARN messagers sont sous forme de simple brin et diffèrent de l'ADN par la présence d'un ribose à la place du désoxyribose et de l'uracile à la place de l'adénine. Ils reproduisent le brin sens de l'ADN des gènes correspondants. Les ARN messagers, appelés parfois pré-messagers, subissent chez les eucaryotes une maturation consistant en l'excision des introns, c'est-à-dire des brins non codants pour des protéines

et un épissage des exons ou assemblage des brins codants. Les ARN messagers matures migrent du noyau vers le cytoplasme dans le réticulum endoplasmique. Là, dans les ribosomes, ils sont à leur tour traduits en une protéine dont la séquence reflète celle du gène correspondant.

La réalité est plus complexe. En effet, un gène est traduit en moyenne de deux ou trois façons différentes. C'est le phénomène d'épigénétisme qui consiste en des modifications de traduction par méthylation de l'ADN ou par des modifications de la structure des histones qui sont des protéines liées à l'ADN.

Définition d'un génome et méthodes de séquençages de génomes entiers

Un génome est constitué de milliards de paires de bases (ou nucléotides) organisées en séquences codantes (les gènes *sensu stricto* qui eux-mêmes possèdent des parties non codantes, les introns) qui peuvent être traduites en protéines fonctionnelles cellulaires, et des séquences non codantes (transposons, microsatellites et autres) qui ne sont pas traduites en protéines cellulaires fonctionnelles.

Les éléments mobiles ou transposons ou gènes sauteurs dont certains sont des rétrovirus sont en partie responsables de la variabilité génétique (allèles). Grâce à leur capacité à provoquer des mutations dans le génome, ils constituent des moteurs essentiels de l'évolution. Ces transposons ont été découverts par Barbara Mc Clintock sur le maïs en 1951 (Prix Nobel en 1983).

Le séquençage de l'ADN a été rendu possible par les travaux de Sanger, Nicklen et Coulson. Sanger a reçu deux prix Nobel l'un en 1958 et l'autre en 1980. En 1977, cette méthode de séquençage, qui porte maintenant le nom de Sanger, permet à son équipe de déchiffrer le premier génome complet d'un organisme vivant, le virus bactériophage ϕ X174. Le séquençage de génomes entiers a démarré dans les années 1990.

La première méthode d'assemblage utilisée a été la méthode BAC pour *Bacterial Artificial Chromosome*. Elle est en général précédée de l'établissement d'une carte génétique. L'établissement d'une carte génétique est basée sur deux phénomènes différents qui se passent au moment de la méiose ou formation des gamètes :

1) Ségrégation indépendante des gènes et loci portés par des chromosomes indépendants. Les gènes et loci ne sont pas liés. 50% des gènes et des loci de la mère et 50% des gènes et des loci du père se retrouvent chez l'enfant

2) *Crossing-over* ou recombinaison homologue par enjambement : au moment de la première phase de la méiose, les chromosomes peuvent se bri-

ser et se reconstituer en échangeant des fragments avec leurs homologues. La ségrégation des gènes ou des loci n'est plus indépendante. Leur proportion n'est plus de 50/50 dans la descendance.

En conséquence, les gènes ou loci portés par des chromosomes différents ne sont jamais liés dans la descendance. Au contraire, les gènes ou loci d'un même chromosome peuvent être liés. La liaison est d'autant plus forte que les gènes ou loci sont proches les uns des autres. Il suffit de calculer dans une descendance les fréquences de recombinaison entre tous les marqueurs utilisés qu'ils soient dans des séquences codantes ou non codantes. On obtient ainsi des groupes de liaisons qui correspondent aux chromosomes. A l'intérieur de ces groupes de liaisons, on peut replacer les marqueurs les uns par rapport aux autres par simple calcul des fréquences de recombinaison.

Après construction de la carte génétique, on construit une banque BAC. L'ADN de l'organisme à séquencer est d'abord extrait, puis fragmenter avec des enzymes de restriction découvertes par Hamilton Smith et Max Birnstiel. Smith a obtenu le prix Nobel en 1978. Ces fragments d'ADN sont multipliés par clonage dans des bactéries d'où le nom de *Bacterial Artificial Chromosome*. Ces fragments sont ensuite séquencés par la méthode Sanger. La banque BAC consiste en plusieurs milliers de clones d'environ 100 Kb couvrant 10 à 15 fois la taille du génome.

La dernière étape consiste à assembler les fragments ainsi obtenu entre eux. Ces séquences BAC sont ensuite ancrées sur la carte génétique, puis assemblées. On obtient ainsi une carte physique.

La méthode par *shotgun* a été découverte un peu plus tard. Après extraction de l'ADN et fragmentation, les séquences sont multipliée par PCR pour *Polymerase Chain Reaction*. Cette méthode de multiplication de l'ADN a été découverte par Kary Mullis, prix Nobel en 1993. Comme pour la méthode BAC, les fragments d'ADN couvrent plusieurs fois le génome. Ils sont d'abord assemblés informatiquement à l'aide d'algorithmes sans passer par une carte génétique par recherche des recouvrements des séquences. Mais l'assemblage peut aussi être amélioré par établissement d'une carte génétique et ancrage des séquences comme pour la méthode BAC. En effet, les assemblages par algorithmes sont souvent aléatoires en raison de la présence de nombreuses séquences répétées : microsatellites et transposons. S'il y a trop de séquences répétées, il faut repasser par la carte génétique.

Pour le décryptage du génome humain, la méthode BAC a été utilisée par le consortium public international (*Human Genome Project*) et la méthode *shotgun* par la société *Celera Genomics* créée par *Perkins Elmer corporation* et

un américain, Craig Venter. Craig Venter, cancre à l'école, vétéran de la guerre du Vietnam, scientifique du NIH et homme d'affaires américain, est un génie à qui l'on doit la méthode shotgun.

Craig Venter a fait beaucoup parler de lui lorsqu'il a séquencé tout seul avec sa petite société le génome humain en 2 ans au lieu de 13 pour l'énorme consortium international, puis lorsqu'il a voulu breveter les gènes et enfin en 2007 en synthétisant un chromosome artificiel avec comme objectif avoué de recréer la vie.

Actuellement, la méthode Sanger est de plus en plus remplacée par le pyroséquençage qui permet encore d'aller beaucoup plus rapidement. Dans cette méthode, les nucléotides sont incorporés les uns après les autres en fonction de l'ordre dans lequel ils se trouvent sur le brin mère. À chaque incorporation dans le brin synthétisé, le nucléotide ajouté libère un pyrophosphate, d'où le nom de la méthode. Cette libération de pyrophosphate induit un signal lumineux qui est capté. Le pyroséquençage associé au shotgun permet de décrypter un génome extrêmement rapidement.

Après le séquençage du bactériophage ϕ X174, divers organismes procaryotes ou eucaryotes ont été séquencés. Actuellement, il existe plus de 1000 génomes procaryotes séquencés. Les génomes procaryotes ou eucaryotes ont été choisis en fonction de leur intérêt (levure de bière, génome humain, riz) ou en fonction de leur petite taille comme l'arabette (*Arabidopsis thaliana*) qui constitue toujours un organisme végétal modèle. Le premier arbre, le peuplier, a été séquencé en 2008. En raison de l'accélération de la puissance de séquençage, il est devenu impossible d'établir une liste des organismes dont le génome a été décrypté. Leur nombre augmente de façon exponentielle.

La taille des génomes exprimée en Mbp (millions de paires de bases) est très variable : Rétrovirus de la grippe: 0,013 (8 fragments d'ARN codant pour 11 protéines)

Escherichia coli (bactérie du tube digestif) 4,64 (4243 gènes)

Laccaria bicolor (champignon) 65 (18 000 gènes)

Arabidopsis thaliana (Arabette) 115 (28 159 gènes)

Populus trichocarpa (peuplier) 485 (45 500 gènes)

Mus musculus (souris) 3 400 (30 000 gènes)

Homo sapiens 3500 (26 500 gènes)

Zea maïs 5000 (nombre de gènes inconnus)

Amoeba dubia (Amibe) 675 000 (nombre de gènes inconnus)

Annotation des génomes

L'annotation des génomes consiste à décoder les informations contenues dans les séquences. Elle consiste à rechercher les gènes et par conséquent les régions codantes qui peuvent être exprimées soit en séquences nucléotidiques soit en séquences protéiques. Il faut ensuite tenter de déterminer la fonction de ces gènes par analogie avec des gènes connus d'autres organismes. L'annotation se fait d'abord informatiquement à l'aide de programmes complexes construits à cet effet et manuellement, puis par une approche plus intégrée faisant appel, en plus de l'aspect purement moléculaire, à d'autres disciplines comme la physiologie ou la génétique quantitative. Un locus (un emplacement précis sur le génome) à effets quantitatifs (QTL pour *Quantitative Trait Locus*) est un locus où la variation allélique est associée à la variation d'un caractère quantitatif. L'existence d'un QTL permet de placer avec plus ou moins de précision le QTL sur la carte génétique et si on a le génome assemblé sur les séquences correspondantes. Pour cela, il faut la carte génétique et encore mieux la carte physique des parents et une descendance de plusieurs centaines d'individus où l'on va mesurer le caractère d'intérêt. Dans la dernière décennie, la cartographie de QTL a permis d'identifier des centaines de régions chromosomiques contenant des gènes liés à l'asthme, l'artériosclérose, les différentes de diabète, l'hypertension, l'obésité, etc. L'objectif est d'identifier les gènes qui sont impliqués dans ces maladies à causes multigéniques et de mieux comprendre leurs implications. La détection des QTLs chez l'homme a d'abord été lente. La technique est maintenant prometteuse et la récolte ne fait que commencer.

On doit aussi à Craig Venter, la technique des ESTs pour *Express Sequence Tags*, qui a révolutionné l'étude du fonctionnement des génomes. Les ESTs sont des ARN messagers exprimés dans une situation particulière ou un organe précis. La transformation de ces ARN messagers en ADN complémentaire (ADNc) permet de retrouver les gènes correspondants et de préciser ainsi leur fonction dans un processus.

Le décryptage d'un génome autorise la construction de puces à ADN, ce qui permet de multiplier encore le décryptage par les ADNc. Des fragments d'ADN simple brin amplifiés par PCR et correspondant à l'ensemble de gènes d'un organisme sont déposés sur une lame de verre ou un autre support. Les ARN messagers de ce même organisme sont extraits dans une situation déterminée et transformés en ADN complémentaire simple brin. Cet ADNc est hybridé à l'ADN de la lame de verre, ce qui permet de mesurer l'expression des différents gènes dans une situation bien déterminée. On peut ensuite assembler les gènes ayant des niveaux d'expression voisins dans une même situation et ainsi reconstituer des réseaux de gènes intervenant dans un processus complexe.

Quel que soit l'organisme, l'annotation d'un génome n'est jamais terminée et progresse continuellement en fonction des nouvelles données acquises en biologie.

La CGH ou *Comparative Genomic Hybridization* ou Hybridation génomique comparative permet aussi de déterminer le fonctionnement de gènes par exemple dans une tumeur. Son principe est basé sur l'hybridation, en métaphase, de deux sondes marquées par des fluorochromes différents. L'une correspond à un ADN d'intérêt (ADN tumoral) et l'autre à un ADN normal (lymphocytes). On peut distinguer les pertes ou les gains de fragments d'ADN dans une tumeur. Mais la précision est faible (10-20 mégabases).

Le génome humain

Le consortium international qui a séquencé le génome humain comprenait 20 centres de recherches de six pays différents (Allemagne, Chine, Etats-Unis, France, Japon, Royaume-Uni). Le génome humain est en accès libre sur plusieurs sites dont celui du NCBI (*National Center for Biotechnology Information*) : http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606

La longueur totale du génome humain est d'environ 3.000 Mb (trois milliards de bases). Le nombre de gènes est d'environ 25 000. Les séquences étaient (et sont toujours) d'accès libre et immédiat, assemblées et orientées grâce aux données de cartographie génétique et physique, et de comparaisons de séquences. La montée en puissance de ce programme international est une conséquence des importants progrès technologiques et informatiques, qui ont par exemple permis d'atteindre une vitesse globale de séquençage de 1.000 nucléotides par seconde. Aujourd'hui 96 % du génome humain est totalement séquencé et plusieurs chromosomes sont terminés.

Thérapie génique et épigénétique

L'objectif du consortium international (*Human genome Project*) était, en dehors du séquençage des 3,5 milliards de bases du génome humain, de

1. Connaître la fonction des 25 000 gènes
2. Créer une base générale des données
3. Transférer les résultats dans la pratique médicale.

La connaissance de la fonction des 25 000 gènes humains progresse continuellement. La base de données du NCBI est un site exceptionnel de connaissances, mis à jour en permanence et accessible à tous.

La thérapie génique est une méthode permettant de corriger ou de remplacer un gène défectueux, responsable d'une maladie. Mais il s'agit d'une

thérapie somatique, c'est-à-dire d'une thérapie n'intéressant que des cellules non sexuelles. De notre point de vue, la thérapie génique germinale portant sur des cellules sexuelles (spermatozoïdes ou ovules) ou sur les premiers stades de division cellulaires d'un embryon n'est pas envisageable chez l'homme. Pour la thérapie génique somatique, il faut tout d'abord isoler puis cloner le gène ayant un intérêt thérapeutique, ce qui ne pose pas de problèmes particuliers en raison des connaissances actuellement acquises sur le génome humain et qui, rappelons le, progressent continuellement.

Pour arriver à la cellule somatique à traiter, il faut un vecteur qui peut être un rétrovirus, un adénovirus ou un virus associé, ou encore le virus de l'herpès. Les virus utilisés comme vecteur proviennent de souches atténuées. Il faut aussi orienter le vecteur viral vers un type déterminé de cellule cible et donc modifier le système de reconnaissance entre le virus et la surface cellulaire de la cellule cible. Il est possible aussi d'insérer directement l'ADN dans la cellule, ou d'utiliser des liposomes, c'est-à-dire de petites particules lipidiques contenant le gène désiré. On peut également envisager d'introduire un 47^{ème} chromosome artificiel dans la cellule cible.

Le gène d'intérêt avec son vecteur est soit inséré *ex vivo* par voie sanguine, soit *in vivo* par insertion dans des cellules cibles prélevées sur le patient puis réinsérée dans le tissu malade.

La thérapie génique est encore uniquement au stade expérimental. En 1999, la mort de Jesse Gelsinger lors d'un essai thérapeutique a donné un brutal coup d'arrêt aux recherches. Un nouvel arrêt s'est produit en 2003 à la suite d'une leucémie contractée par un enfant bulle lors d'un essai français effectué en 2002.

Il existe encore de nombreux problèmes non résolus. Les cellules se divisent dans l'organe traité et l'effet initial peut être très temporaire. Le système immunitaire peut répondre à l'introduction du gène additionnel et induire des troubles graves. Les vecteurs viraux ne sont pas sans danger, même s'il s'agit de souches atténuées. Il peut enfin apparaître des désordres multigéniques.

Cependant, quelques progrès récents ont pu être réalisés. En 2008 en Angleterre, une amélioration de la vue dans un cas de cécité innée a pu être observée.

En 2007 aux USA, un succès dans le traitement du cancer du poumon par deux suppresseurs d'activité de gène chez la souris a été obtenu. Notons encore en 2006 une amélioration de l'activité des lymphocytes contre les métastases à myéloïdes (sites consultables : www.medisite.fr - www.inapg.inra.fr - www.genethique.org).

La thérapie épigénétique procède d'une autre approche. Il s'agit d'utiliser des substances qui sont des inhibiteurs spécifiques de certains effecteurs épigénétiques intervenant dans les modifications post traductionnelles de l'expression des gènes. La caractérisation d'inhibiteurs spécifiques de certains effecteurs épigénétiques a ouvert une nouvelle voie thérapeutique, la thérapie épigénétique, qui semble très prometteuse. L'article de Ken Karner paru en 2002, aurait été considéré comme une hérésie avant 2002. Maintenant, la majorité des scientifiques acceptent l'idée que le remodelage de la chromatine est essentiel dans le traitement des cancers. La méthylation et la déacétylation des histones peut empêcher la transcription et la traduction de gènes pouvant inhiber le développement des tumeurs. D'une manière schématique la méthylation réduit les gènes au silence alors que l'acétylation les active. On dispose maintenant de toute une panoplie de traitements potentiels. On peut ainsi orienter le développement des cellules tumorales et les amener au suicide ou à la sénescence. Les chercheurs travaillent sur cette nouvelle théorie en recherchant des inhibiteurs de méthylation ou des histone deacetylases (HDAC). En 2009, l'azacitidine est devenue le traitement de référence des myélodysplasies (MDS) à haut risque. Bientôt, d'autres molécules vont apparaître et on assistera probablement à la naissance de traitements à base d'associations d'agents épigénétiques.

L'ultime objectif du Projet sur le génome humain est d'utiliser toutes les informations acquises ou en cours d'acquisition non seulement pour traiter de nombreuses maladies, mais pour prévenir les milliers de maladies qui affectent l'homme. Dans les toutes prochaines années, il sera possible de séquencer le génome de chaque individu. L'objectif est d'arriver à un coût de 1000 \$ pour le séquençage d'un génome individuel.

Il sera alors possible de prédire les risques d'apparition de beaucoup de maladies et de préconiser en conséquence des traitements préventifs ou des modes de vie adaptés.

Métagénomique

La métagénomique a pour objectif de déterminer à l'aide d'outils moléculaires l'ensemble des espèces constituant un écosystème, ce qui est possible grâce au développement rapide des méthodes de séquençage à haut débit.

Par exemple, le microbiote intestinal, défini comme l'ensemble des communautés microbiennes hébergées par le tractus digestif (principalement des bactéries, environ 600 espèces différentes) est aujourd'hui reconnu comme partenaire essentiel de la santé de l'hôte. Les premiers décryptages de métagénomies intestinales de volontaires sains et de patients porteurs de maladies spécifiques, ont débuté il y a 5 ans.

L'objectif d'un des projets de l'Unité INRA Ecologie et Physiologie du système digestif est d'explorer le métaprotéome du microbiote intestinal dans diverses situations physiopathologiques afin de découvrir de nouvelles thérapies pour de grandes pathologies (maladies inflammatoires chroniques de l'intestin, diabète, obésité, cancers du tube digestif).

Agriculture

Sélection génétique assistée par marqueurs

La découverte des marqueurs moléculaires de l'ADN nucléaire ouvre une nouvelle ère pour la sélection végétale ou animale. En rendant possible l'étiquetage de certains gènes d'intérêt, il est maintenant possible de mieux tirer parti de la variabilité génétique naturelle pour construire des génotypes ayant de plus en plus d'allèles ou d'associations d'allèles favorables.

Nous ne donnerons qu'un seul exemple. En 2009, des chercheurs californiens ont découvert les gènes responsables de la tolérance aux basses températures du blé d'hiver. Cette découverte ouvre la voie à la création de nouvelles variétés pouvant avoir une production significative dans des zones plus septentrionales qu'actuellement.

OGM

Les OGM actuels sont encore à un stade embryonnaire. Avec le décryptage du génome de la plupart des plantes cultivées, les possibilités de transformation deviennent infinies. Malgré les polémiques actuelles en Europe, les OGM devraient dominer sous peu l'agriculture mondiale.

Comprendre l'évolution

L'évolution des gènes

Le séquençage de nombreux génomes permet de comparer les gènes entre eux et de mieux comprendre l'évolution des gènes. Un des phénomènes majeurs d'évolution des gènes, découvert ces vingt dernières années est la duplication. Celle-ci peut affecter une partie d'un gène, un gène entier, un groupe de gènes, un chromosome (aneuploïdie) ou le génome dans son entier dans le cas de la polyploïdie (Sankoff, 2001).

Le génome d'*Arabidopsis thaliana* a-t-il subi plusieurs duplications plus ou moins importantes ayant abouti à la présence de nombreux gènes dupliqués. La duplication d'un gène unique donne naissance à un tandem. Des duplications successives en tandem donnent naissance à des groupes de gènes situés sur un même locus.

Les éléments transposables peuvent jouer un rôle dans la duplication des gènes ou le déplacement de fragments importants. Les éléments transposables sont des séquences d'ADN mobiles dans le génome. Certains sont proches des rétrovirus. Ils ont été découverts par Barbara McClintock chez le maïs, ce qui lui a valu le prix Nobel de médecine en 1983. Les transposons se répandent dans le génome par « coupé collé » ou par « copié collé » (rétro transposons). L'intégration d'un transposon dans un gène ou à proximité d'un gène altère son expression. Les éléments transposables jouent un rôle important non seulement dans la duplication de gènes, mais aussi dans le taux de mutation dans leur voisinage.

Les duplications de gènes sont suivies de divergence modifiant les séquences par accumulation aléatoire de mutations. Les protéines correspondantes sont modifiées en conséquence. Il se forme ainsi des familles de gènes, appelés paralogues, par opposition aux gènes orthologues ayant évolué dans des organismes différents après spéciation. Par comparaison de séquence, il est possible d'établir la phylogénie des gènes, c'est-à-dire d'établir comment ils ont dérivé les uns des autres soit à l'intérieur d'une même espèce, soit entre espèces différentes.

L'évolution des génomes et des espèces

Le concept de base est que les espèces étudiées ont un ancêtre commun et que plus leur divergence est grande, plus leur séparation est éloignée dans le temps. La comparaison des génomes entre eux permet de mieux comprendre leur évolution et donc l'évolution des espèces. Mais la plupart du temps, pour comprendre l'évolution des espèces on n'utilise qu'un ou quelques gènes. Tous les gènes n'évoluent pas à la même vitesse. Pour comparer des organismes différents ou très différents, on utilise des gènes qui se sont conservés au cours de l'évolution en raison de leur importance dans des processus essentiels, ce qui est le cas des gènes codant pour les ARNs ribosomiques et qui sont très utilisés en phylogénie. En général on utilise des séquences codantes pour les analyses remontant loin dans le temps et des séquences non codantes, où les mutations s'accumulent sans effets délétères, pour des analyses de divergence récente. L'ADN mitochondrial, qui se transmet essentiellement par la mère, est souvent mis à contribution pour les phylogénies récentes et les marqueurs microsatellites pour les analyses très récentes d'évolution au sein d'une population.

L'évolution de l'homme

Les études de l'ADN mitochondrial, de l'ADN du chromosome Y et de l'ensemble du génome humain ont fait récemment progresser nos connaissances sur l'origine de l'homme moderne. Dans la communauté scientifique, il existe un large consensus, mais qui ne fait pas l'unanimité complète, pour considérer cette origine comme africaine.

Selon des données moléculaires et paléontologiques qui se recoupent convenablement, l'homme moderne serait apparu il y a environ 200 000 ans en Afrique. Il aurait quitté ce continent il y a 60-70 000 ans pour coloniser le reste du monde. Selon les études de l'ADN mitochondrial de Doron M. Behar et al. (*The Dawn of Human Matrilineal Diversity*, 2008 *The American Society of Human Genetics*), il y a environ 200 000 ans une première branche dite LO se serait différencié d'un ensemble appelé L1 et comprenant six branches dénommées L1' à L6'. De la branche L3, il y a environ 70 000 ans, se seraient différenciés deux sous groupes dits M et N qui auraient traversé la Mer rouge. On ne sait pas pour l'instant si les deux sous groupes M et N se sont différenciés avant la traversée de la Mer rouge, ce qui indiquerait deux traversées différentes, ou après, ce qui n'impliquerait qu'une seule traversée. Il semble qu'à partir d'une population africaine comprenant entre 2 000 et 5 000 femmes, seules 150 auraient traversé la Mer rouge pour se répandre dans le monde entier.

L'analyse des haplotypes du chromosome Y donne des résultats plus complexes. Des événements relativement récents impliquent une dispersion de l'haplotype E-M78 de l'Afrique de l'Est en Afrique et hors d'Afrique. Des événements plus anciens (environ 45 000 ans) impliqueraient des migrations de faible amplitude à l'intérieur de l'Afrique (haplotype E-M78 et E-V6) ainsi que d'Afrique du Nord vers l'Europe (haplotypes E-M81 et E-M78), ce qui indiquerait une autre migration hors d'Afrique.

Bien que l'analyse de l'ADN mitochondrial et de celui du chromosome Y soit particulièrement utile pour tenter de déchiffrer l'histoire humaine, d'autres approches peuvent être mises en œuvre comme l'étude du polymorphisme de l'ADN de populations appartenant à des groupes différents. En Juin 2009, une analyse de SNPs (polymorphisme n'affectant qu'un seul nucléotide) a été réalisée dans le cadre du projet international *HapMap* et le centre d'étude du polymorphisme humain sur 1 138 individus appartenant à 53 populations différentes. Contrairement à ce qu'attendaient les généticiens, seuls trois groupes sont ressortis : les Africains, les Eurasiens (Europe, Moyen-Orient, Inde et Pakistan) et les Asiatiques orientaux (Asie, Japon, Amérique et Océanie). Cette étude semble montrer que le génome humain évolue plus lentement que prévu.

Cependant, une autre étude publiée en 2008 dans *Nature*, montre une évolution rapide des populations européennes. L'analyse de 500 000 SNPs sur un échantillon de plus de 3 000 individus a montré que les groupes ainsi constitués recoupaient parfaitement les frontières et en particulier les frontières linguistiques, ce qui suppose une évolution très rapide.

Conclusions

La liste des organismes séquencés augmente chaque jour. Une des grandes surprises a été la découverte de l'importance des séquences non codantes et en particulier celle des éléments transposables qui jouent un rôle essentiel dans l'évolution. Les transposons constituent 42 % du génome humain et les séquences codantes seulement 5 %. L'expression de ces séquences codantes est sous le contrôle de différents facteurs dont les phénomènes d'épigénétisme encore mal connus. La connaissance du rôle des gènes avance rapidement par l'utilisation de puces à ADN ou par co-localisation sur les cartes génétiques avec des caractères quantitatifs. Cependant, plus nous avançons dans la compréhension des mécanismes de fonctionnement des génomes, plus la complexité de la vie paraît inouïe. Néanmoins, la connaissance du génome humain permet d'envisager de nouveaux traitements par thérapie génique ou épigénétique. La connaissance des génomes individuels devrait aussi bientôt permettre d'améliorer la prévention de nombreuses maladies. D'une manière plus large, la connaissance des génomes permet de mieux comprendre l'évolution des gènes, des génomes et des espèces. L'analyse du génome humain a récemment permis de faire de nouvelles hypothèses sur le mystère de nos origines. Nul doute que de nouveaux résultats viendront bientôt confirmer ou infirmer ces hypothèses.



Discussion

Le président félicite et applaudit F. Le Tacon pour la splendide leçon qu'il nous a donnée, allant même jusqu'à des observations qui relèvent de la linguistique.

S'ensuit alors une question de Michel Vicq qui demande si les vrais jumeaux, c'est-à-dire les monozygotes, ont le même ADN ? «*oui*, répond le conférencier» M. Vicq fait encore observer que si leur ADN est identique, leurs empreintes digitales sont différentes - François Le Tacon, très étonné, pense qu'il s'agit là d'un cas d'épigénétisme.

Dominique Flon fait observer qu'il n'est pas assuré que les locuteurs de l'indo-européen ont constitué un phylum, car certaines populations ont pu adopter cette langue par imprégnation, comme plus tard les Gaulois ont adopté le latin et délaissé leur propre langage. Cette réflexion génome-population renvoie aussi aux recherches des linguistes du XIX^{ème} siècle, tel qu'Adolphe Pictet, qui étaient persuadés de la supériorité de la race blanche, et à celles de Gobineau qui s'intéressait à l'interaction entre milieux naturels et développe-

ment technique. Les phénomènes de migrations donnent à penser que sur le très long terme, le mélange des génomes bloquera la séparation des hommes entre plusieurs espèces.

Le Docteur Paul Vert indique que les gènes peuvent s'exprimer différemment selon des influences épigénétiques. Ceci explique qu'une même anomalie génique puisse donner, selon le moment où elle s'exprime, des malformations congénitales différentes chez deux frères. Une substance toxique, comme l'alcool par exemple, peut, ou non, modifier la morphogénèse selon le stade du développement de l'embryon. Ceci est plus vrai encore quand plusieurs gènes sont en cause simultanément.



Bibliographie sommaire

- Avery O.T., McLeod C.M. & McCarthy M., 1944. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III J. Exp. Med. 79, 137-158.
- Cavalli-Sforza L.L., Menozzi, P. and Piazza A., 1994. The History and Geography of Human. Genes, 292 p., Princeton Univ. Press.
- Cavalli-Sforza L.L., Piazza A., Menozzi P., and Mountain J., 1988. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data (origin of modern humans/phylogenetic trees/paleoanthropology), *Evolution*, 85, 6002-6006.
- Smith Hamilton O. and Birnstiel Max L., 1976. A simple method for DNA restriction site mapping, *Nucleic Acids Research*, 3, 9, 2387-2398 .
- Chargaff E., 1950. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, 6, 201-209.
- Dahm Ralf, 2005. Friedrich Miescher and the discovery of DNA. *Development Biology*, 278, 2, 274-288.
- David J. Samuelson, Jill D. Haag, Hong Lan, Dinelli M. Monson, Millicent A. Shultz, Bradley D. Kolman and Michael N. Gould, 2003. Physical evidence of Mcs5, a QTL controlling mammary carcinoma susceptibility, in congenic rats. *Carcinogenesis*, 24, 9, 1455-1460.
- Franklin R.E. and Gosling R.G., 1953. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356), 740-741.
- Franklin R.E. and Gosling R.G., 1953. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, 172 (4369), 156-157.
- Glass Bentley, 1965. A century of biochemical genetics. Symposium Commemorating the Publication of Gregor Mendel's Pioneer Experiments in genetics. *Proceedings of the American Society of Philosophy*, 109, 4, August.

- Greenstein Jesse P. , 1943. Friedrich Miescher, 1844-1895. *Scientific Monthly*, 57, 523-532.
- Jones Mary Ellen, 1953. Albrecht Kossel, A Biographical Sketch *Yale J Biol Med.*, 26 (1), 80-97.
- Karner Ken, 2002. *Breaking the Silence: The Rise of Epigenetic Therapy. Journal of the National Cancer Institute*, 94, 12, 874-875.
- McClintock B., 1950. The origin and behavior of mutable loci in maize. *P.N.A.S.*, 36, 344-355.
- McClintock B., 1951. Chromosom organization and genetic expression. Cold Spring Harbor Symposium. *Quantitative Biology*, 16, 13-47.
- McClintock B., 1952. Mutable loci in maize. Carnegie Institute, Washington, Year Book, 51, 212-219.
- Mendel, Johann, «Versuche über Pflanzenhybriden», *Verhandlungen der Naturforschenden Verein*, 4, 3-47, 1866, Brünn, traduit et repris dans *Journal of Heredity*, 42 : 3-47, 1951.
- Monod J. and Jacob F., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3, 3, 318-356.
- Morgan T.H., 1917. The Theory of the Gene. *The American Naturalist*, LI, 609.
- Mullis K.B., Faloona F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain réaction, 155, 335-50.
- Novembre John et al., 2008. Genes mirror geography within Europe. doi:10.1038/nature 07331.
- Sanger F., Nicklen S. and Coulson A.R., 1977. DNA sequencing with chain-terminating inhibitors. *P.N.A.S.*, 74, 12, 5463-5467.
- Swartz M.N., Trautner T.A. and Kornberg A., 1962. Enzymatic Synthesis of Deoxyribonucleic acid. *The Journal of Biological Chemistry*, 237, 1961-1967.
- Venter J. Craig et al., 2001. The Sequence of the Human Genome, *Science*, 291, 5507, 1304-1351.
- Watson J.D. and Crick F.H.C., 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 25, 171 (4356), 737-738.
- Wilkins M. H., Seeds W. E., Stokes A. R., Wilson H. R., 1953. Helical structure of crystalline deoxypentose nucleic acid. *Nature*, 172 (4382),759-762.
- Wilkins M. H., Gosling R.G., Seeds W.E., 1951. Physical studies of nucleic acid. *Nature*, 167(4254) 759-760.
- Wilson Edmond,1886. *The cell in Development and Inheritance*.