

Atelier N°7 « Génomique et évolution »

Bernard DUJON Professeur à l'Institut Pasteur et à Paris VI
Philippe LOPEZ Maître de conférences à Paris VI

Les deux diaporamas seront en ligne sur le site de l'INRP, ainsi que les échanges du deuxième atelier.

Exposé de B. DUJON

Le génome ne se limite pas aux gènes codants pour des protéines :

	Homme	Levures
Gènes codants (EXONS)	23000 (<2% génome)	5770 (3/4 génome)
Introns	100 000	280
Pseudogènes	25000	10
Transposons	1 100 000	50
Génome (en Mbases)	2900	12,5

Si les fonctions cellulaires sont bien ce que les protéines font, tout le reste du génome est impliqué dans la régulation et l'évolution : l'Homme évolue beaucoup, les Levures très peu.

Trois principes de la génomique permettent de comprendre l'évolution :

1. La divergence des séquences :

L'évolution se fait au cours de centaines, voire de milliers de générations successives à l'origine du polymorphisme génétique des populations. Au cours de ces générations un équilibre s'établit entre les erreurs de réplication qui augmentent le polymorphisme et la sélection naturelle ou la dérive génétique qui le diminuent.

Chez l'Homme, en moyenne trois mutations se produisent par mitose, ou exprimé différemment la fréquence des mutations est de 10^{-9} mutations/nucléotide/an, ce qui engendre un polymorphisme 15 fois plus faible que chez la Levure. 500 000 générations seulement nous séparent de la bifurcation Chimpanzé – Homme ce qui se traduit par une divergence de 1,4% des séquences nucléotidiques.

Mais tout cela n'explique que le polymorphisme des populations et pas réellement leur évolution.

2. Le réarrangement des cartes génétiques :

La duplication des gènes augmente leur redondance, tandis que la perte des gènes la diminue : la combinaison des deux avec les mutations est à l'origine de nouvelles fonctions.

Exemple de la Paramécie : elle possède 49000 exons mais la plupart sont en deux exemplaires, si on efface l'effet de la duplication du génome à l'origine de ces 49000 exons souvent redondants, on obtient 25000 exons antérieurs dont beaucoup sont redondants, il y a donc eu une deuxième duplication du génome et sur les 12500 exons d'origine un certain nombre étant encore redondants on peut penser que le génome de la Paramécie a subi trois réplifications successives depuis l'origine de cette espèce.

Exemple des souches de Levure : on peut rendre des souches de Levure (à 16 chromosomes) malades en leur enlevant une séquence nucléotidique ; si on les met en culture sur un milieu défavorable, elles se multiplient très peu et avec une fréquence de mutations de 10^{-7} par mitose. On voit alors se développer sur le milieu quelques colonies en forte croissance chez qui les mutations ont fait apparaître un 17^{ème} chromosome : peut-on encore dire que ce mutant à 7 chromosomes appartient toujours à *Saccharomyces cerevisiae* ? ou s'agit-il d'une nouvelle espèce ?

En conclusion, si une redondance par duplication est suivie d'une mutation, une nouvelle fonction peut apparaître mais ces variations sont compensées par les pertes de gène qui

sont la norme chez tous les êtres vivants au cours des divisions ou par le déplacement des gènes au sein des chromosomes qui sont également fréquents !

Ces deux phénomènes en équilibrant l'apparition de nouveaux gènes et de nouveaux allèles font qu'il existe un certain équilibre du génome original.

3. la création de nouveaux gènes :

4 mécanismes sont à l'origine de nouveaux gènes :

- la fusion – troncation
- les changements de fonction des exons codants
- les transferts de gènes
- l'EXON *shuffling* par lequel un exon est traduit en ARN, puis une copie de l'ARN s'insère dans un intron ailleurs dans le génome et suite à des mutations de type délétion, cette copie vient fusionner avec un exon voisin pour donner un nouvel exon de fonction différente (exemple du gène RPE2-1 chez l'homme dont le 3^{ème} des 6 exons codant pour cette protéine résulte de l'insertion d'une copie d'ARN dans l'intron initialement situé entre les 2^{ème} et 4^{ème} exons).

Si on applique ce mécanisme à l'origine de RPE2-1 à l'arbre phylogénétique des primates, on constate que l'insertion de la copie d'ARN dans l'intron s'est produite vers -35 millions d'années, que l'EXONISATION (= insertion-fusion de l'intron à copie d'ARN dans le 2^{ème} exon) s'est produite vers -15 MA et n'existe que chez les Gibbons et les hominoïdés puis qu'une réversion a fait disparaître cet exon chez le Gibbon, seul les hominoïdés le possédant encore parmi les primates.

Les nouvelles techniques de séquençage :

La génomique a explosé depuis 15 ans grâce à la technique de Sanger qui permet d'identifier par tour de machine (« run ») à partir d'ADN purifié, des séquences de 70kb (70 000 nucléotides). Cette technique est à l'origine du séquençage complet du génome humain en 2001.

Actuellement chaque jour voit la publication d'un nouveau génome : dans le dernier *Nature*, trois génomes humains sont publiés : celui d'un chinois, celui d'un africain, celui d'un malade du cancer.

Cette accélération du décryptage des génomes est possible depuis 2007 grâce à la mise au point de deux techniques révolutionnaires :

- le PYROSEQUENCAGE qui permet d'obtenir de 100 Mb à 600 Mb (600 millions de nucléotides) par run sur de grosses séquences, ce qui facilite le réassemblage des parties du génome ;
- le SEQUENCAGE EN PHASE SOLIDE, encore plus performant (jusqu'à 2000 Mb/run) mais qui fournit de plus petites séquences dont le réassemblage au sein du génome est plus délicat.

Ces deux techniques ont divisé par un facteur 1000 le coût de séquençage des génomes et permet désormais explorer la diversité des êtres vivants et donc leur parenté sur l'intégralité des génomes. De ce fait, la biologie devient très dépendante de l'informatique et des mathématiques et les laboratoires de génomique recrutent maintenant des biomathématiciens, plutôt que des biologistes.

Exposé de P. LOPEZ

Problème : en quoi la connaissance des génomes est-elle utile à l'évolutionniste ?

En raison de la fluidité des génomes ...

La phylogénie des organismes en elle-même est peu développée en recherche, essentiellement en direction de la taxonomie et de la systématique, alors que la phylogénie moléculaire apporte de nombreux arguments à la phylogénie des organismes. Si l'on travaille sur la phylogénie d'un seul gène, le signal phylogénétique est insuffisant : si l'évolution est lente, c'est-à-dire à grande échelle, on ne parvient pas à distinguer les espèces proches car le gène étudié n'a pas muté ; si au contraire l'évolution est rapide, c'est-à-dire à petite échelle, on n'obtient plus de réponse pour les espèces éloignées, car les seuls gènes qu'il est possible d'étudier sont ceux qui ont très peu muté (protéines de fonction), les autres gènes n'étant plus apparentables car trop différents. Le facteur limitant de la génomique en évolution était donc la quantité de données.

Les techniques de séquençage des génomes entiers permettent de travailler de front pour les espèces dont on cherche à construire l'arbre phylogénétique, sur un nombre élevé de gènes (jusqu'à 125), ce qui permet de remonter loin dans le temps puisqu'on combine alors des gènes à évolution rapide et des gènes à évolution lente. On a de cette manière identifié six grands groupes d'eucaryotes établissant des liens de parenté parfois surprenants :

- les OPISTHOKONTES regroupent les animaux pluricellulaires, les protozoaires et les champignons qui seraient des protozoaires ayant perdu cils et flagelles
- les PLANTES regroupent les végétaux verts et les algues rouges, tous issus d'un ancêtre commun ayant développé l'unique endosymbiose chloroplastique de l'évolution de la vie.

Toutefois, peu de progrès ont été obtenus depuis 2003 car si le nombre de signaux phylogénétiques augmente, en revanche leur résolution s'améliore peu et on sait que la phylogénie moléculaire ne permettra pas de résoudre toutes les questions d'évolution.

L'évolution des génomes est connue actuellement par comparaison des chromosomes grâce au séquençage complet des génomes. Les gènes se dupliquent, se perdent, changent de région chromosomique : le génome n'est pas figé, il évolue, il est fluide et plastique tout comme le chromosome qui évolue lui aussi !

L'ADN non codant présente des régions conservées chez de nombreuses espèces, donc sélectionnées par la sélection naturelle ce qui démontre leur importance pour les organismes : ce sont entre autres les séquences régulatrices. Mais lorsque l'on regarde trois souches d'*Escherichia coli* non pathogène, pathogène pour les voies urinaires et pathogène pour les hématies, on constate qu'elles n'ont que 39,2% de gènes en commun (2996 gènes) et que la souche non pathogène possède 46% de ses gènes en propre, sans les partager avec les deux autres souches : cela pose la question de la définition de l'espèce chez les procaryotes.

Chez les bactéries Thermotogales, deux bactéries considérées comme très proches (*Thermosiphon africanus* et *Thermotoga maritima*) ont montré qu'en reconstruisant la phylogénie de *Thermosiphon*, un grand nombre de gènes étaient différentes entre les deux espèces : cela pose la question de la validité de l'arbre des procaryotes.

Enfin on découvre de plus en plus de transferts horizontaux entre espèces contemporaines de nature très différentes : par exemple la bactérie *Wolbachia pipiens* peut intégrer la totalité de son génome dans celui de la drosophile qui ensuite exprime plusieurs des gènes bactériens.

Il paraît probable qu'il faudra étendre le cadre darwinien de l'évolution et ne plus se limiter à la transmission de gènes entre parents et descendants.

La métagénomique est en plein essor. Pour séquencer un génome, il faut travailler sur de l'ADN pur obtenu à partir d'une seule espèce. Or seul 1% des bactéries peuvent être cultivées. Les 99% d'autres espèces bactériennes vivent en communauté indissociables, la séparation de l'une d'elles ne permettant pas sa survie. La métagénomique consiste à séquencer directement à partir du milieu naturel les génomes des espèces vivant en communauté, sans les distinguer.

En conclusion, un génome pris seul n'explique rien et le taxon espèce devra sans doute être revu en ce qui concerne les bactéries.

Echanges au sein des deux ateliers

Première séance :

Les pertes de gènes sont-elles à l'origine de pseudogènes ou de pertes de séquences nucléotidiques ?

Ce sont le plus souvent des délétions nettes : un enfant par exemple possède de 30 à 50 fragments d'ADN de moins que ses parents soit environ 500 000 nucléotides ; mais du fait de l'hémizygotie, ces déficiences ne sont pas délétères. C'est la même chose chez les levures et même chez les haploïdes en général, les délétions n'étant pas systématiquement suivies d'effet mesurable sur les phénotypes.

Vous avez parlé de deux adaptabilités différentes pour l'Homme et la Levure : est-ce en lien avec la place des organismes dans les écosystèmes ?

La phylogénie des primates montre une forte dynamique des chromosomes, en particulier en ce qui concerne les parties non géniques qui sont très mobiles. Inversement, les levures sont des champignons qui ont perdu des fonctions essentielles dans le contrôle des ARN. Deux stratégies évolutives s'opposent dans les écosystèmes :

- la stratégie de la croissance la plus rapide avec peu d'évolution (les levures n'ont que peu changé depuis 500 millions d'années et changeront peu dans les 500 MA à venir)
- la stratégie de l'évolution rapide avec forte diversification, c'est le cas des primates qui eux ne seront plus là dans 500 MA.

Les transposons sont-ils importants dans l'évolution et donnent-ils un avantage dans le maintien des gènes ?

Les transposons de type ARN → ADN, les rétrogènes insérés ailleurs dans le génome, sont de très importantes sources de pseudogènes. La fusion du domaine nucléotidique d'un ancien gène avec un transposon est souvent à l'origine d'un nouveau gène. En plus, le changement de position du gène dans le génome peut se traduire par des séquences dispersées et répétées : toutes ces recombinaisons anormales sont à l'origine des aberrations chromosomiques.

Les transposons créent donc des instabilités génomiques et la recherche s'oriente de plus en plus vers l'étude du rassemblement des gènes, les gènes apparentés de type opéron étant très fréquents chez les procaryotes mais très rares chez les eucaryotes.

Quels sont les mécanismes de conservation des parties d'ADN non codantes ?

Ce la pose la question de la définition du gène. Le gène n'est pas seulement une séquence codante, c'est aussi une séquence régulatrice, un promoteur, etc.

Si une séquence non codante est conservée, c'est qu'elle sert : lorsque deux génomes sont proches, on étudie les codons situés autour des gènes homologues pour voir si ils sont bien conservés. Les non codants servent principalement à la régulation mais la sélection naturelle agit sur l'ensemble du génome et pas seulement sur les gènes codants. Par exemple,

certaines bactéries ont tous leurs gènes orientés dans le même sens à cause de la pression sélective de leur environnement.

Compte tenu du nombre de bactéries vivant dans l'organisme humain peut-on encore parler d'espèce à propos de l'Homme ?

La vraie notion d'espèce est différente de celle de génome. Chez les procaryotes en revanche elle est difficile à définir (comme les autres groupes taxonomiques) car les transferts horizontaux de génome entre espèces très différentes sont répandus, en dehors de toute sexualité.

La majorité des organismes procaryotes ou eucaryotes fongiques n'ont pas de reproduction sexuée : ils n'en ont pas besoin. Chez ces organismes, l'expansion clonale est dominante et des « espèces très différentes » peuvent se croiser, la notion d'espèce s'atténue, exactement comme chez les pucerons où l'expansion clonale (parthénogenèse) s'accompagne d'échanges.

Où placer les virus au sein de l'évolution ?

Il existe beaucoup de sortes de virus différents : en milieu marin, les bactéries et les bactériophages constituent la part la plus importante de la biomasse. En dehors d'une cellule, le virus n'a pas d'activité, mais il n'a pas non plus de durée de vie : le virus est « éternel ». Certains virus ont beaucoup de gènes, d'autres très peu ; certains ont un spectre d'hôtes très large d'autres sont monospécifiques. Ils ne font pas partie du vivant car ils ne se reproduisent pas seuls et ne métabolisent pas seuls. Mais ils transfèrent un grand nombre de gènes entre les organismes qu'ils parasitent, en particulier dans le cadre des transferts horizontaux. Le génome des virus évolue peu (sauf in vitro) mais beaucoup lors des réplifications, c'est encore plus complexe quand le génome viral est intégré dans le génome de l'hôte.

A quel niveau agissent les mécanismes de l'évolution : de l'espèce ou du génome ?

L'évolution porte sur le génome, mais l'impact sur les phénotypes.

Si le génome ne varie pas, il n'y a pas d'évolution : les mécanismes qui créent ces variations se déroulent à l'intérieur des populations, qui n'évoluent pas de manière coordonnée. Le problème reste « Comment définir une nouvelle espèce ? » : est-ce que le chimpanzé et l'homme sont deux variants ou deux espèces ? Il est pratique d'en faire deux espèces mais ce n'est pas tout à fait la réalité scientifique.

Toutes les espèces de champignons sont en réalité mosaïques par croisement de lignées parentales différentes, c'est la même chose pour le blé dont le génome spécifique est inférieur à la somme des génomes des parents. Ce sont en fait des OGM naturels.

Quelle place donner aux avancées de la génomique dans l'enseignement secondaire ?

On enseigne trop la science comme un savoir révélé, ce qui n'est pas le cas. La science bouge sans cesse, il faut faire entrer cette idée dans les programmes, aller vers le tâtonnement expérimental. Le discours médiatique actuel est très orienté vers la médecine et c'est ce qui tue la science chez les élèves : il faut développer leur curiosité scientifique plutôt que de faire de la fausse médecine. La science doit satisfaire la curiosité des élèves et non pas les besoins de société : il faut se centrer sur les phénomènes fondamentaux.

Les programmes n'ont pas à être à la pointe de la recherche : il ne faut pas perdre les élèves en leur montrant que tout ce qui concerne l'évolution est actuellement d'une énorme confusion.

Et pour les étudiants ?

Les cours universitaires sont là pour poser des questions non résolues, ce qui est résolu est dans les manuels. Les étudiants devraient avoir un stock initial de connaissances et ils sont ignorants de l'histoire des sciences. Certains étudiants ne veulent pas être stimulés, ils préfèrent les certitudes, c'est pour cela que l'histoire des sciences est enseignée à bac + 5 en génomique.

Deuxième séance :

Qu'en est-il exactement des transferts de gènes des procaryotes vers les eucaryotes ?

Au départ on a cru qu'il n'y avait que des transferts entre procaryotes ; on sait qu'ils existent entre les bactéries et les rotifères, que le cnidocyte des Cnidaires est d'origine bactérienne, que ces transferts concernent tout particulièrement le génome non codant autre que les promoteurs.

Ces gènes transférés ne sont pas uniquement de l'ADN poubelle : c'est pour cela que l'on cherche les séquences d'ADN non codant qui n'évoluent pas. Le codant strict passe par le code génétique (moins de 2% chez l'Homme) : le projet de recherche actuel sur le génome humain porte sur les gènes enchevêtrés dont les parties fonctionnelles sont séparées au sein du génome.

On sait aussi que beaucoup d'ARN est à l'origine d'ADN alors que très peu d'ARN code pour des protéines.

Quel est le potentiel évolutif du génome non codant ?

Il faut raisonner en terme de dynamique : renouvellement, transposition, transcription, qui tous trois vont faire évoluer le génome. Un des gènes impliqués dans la maladie d'Alzheimer est ainsi un rétroélément intégré au génome à partir d'ARN et il n'existe pas chez le chimpanzé.

Quelle distinction faire entre transferts horizontaux et hybridation de génomes ?

L'hybridation entre génomes se fait surtout entre plantes voisines alors que les transferts horizontaux sont développés entre organismes phylogénétiquement éloignés : c'est le cas des gènes mosaïques des champignons ou des polyploïdies chez les plantes.

Pourquoi s'est-on limité à la comparaison des gènes codants dans la lignée humaine ?

Pour des raisons techniques car on ne pouvait pas travailler sur tout le génome. Les copy number variants montrent de plus grandes différences que les allèles entre deux individus ; en outre la perte de gène est fréquente et n'est pas forcément associée à un génotype particulier.

La partie contrainte du génome est surtout composée de gènes codants : de ce fait on retrouve peu d'homologies dans les parties non codantes, et cela rend impossible toute construction de phylogénie.

Vous avez comparé levures et homme : qu'en est-il des génomes des autres organismes eucaryotes ?

Beaucoup sont intermédiaires entre les génomes de ces deux groupes d'espèces (ex : éponges, insectes) mais le maïs est une espèce ayant plus de 90% de transposons dans son génome.

Les primates se sont diversifiés car leur génome comprend beaucoup de transposons : génome très dynamique → évolution très rapide.

Les levures ont perdu beaucoup de fonctions ancestrales (contrôle des ARN, ARN interférents, etc.) : elles sont spécialisées dans la croissance pour avoir un succès évolutif et dans ce sens elles sont plus évoluées que l'Homme : les plus complexes sur le plan morphologique ne sont pas les plus évolués, car les pertes de gènes sont irréversibles (ex : la synthèse de la vitamine C chez l'Homme par rapport aux autres primates).

Comment expliquer la diversité du nombre de complexes Ox chez les vertébrés ?

Amphioxus a un seul complexe Ox. Par duplication du génome à deux reprises, les vertébrés ont pour la plupart 4 complexes Ox sauf ...

... les Téléostéens qui ont subi trois duplications et ont ensuite perdu un complexe : ils ont 7 complexes Ox.

Ceci traduit l'explosion d'espèces à la faveur de pertes aléatoires, en accord avec la théorie de l'incompatibilité Dobzhansky-Müller.

L'explosion cambrienne des phylums est-elle explicable par l'analyse génomique ?

500 millions d'années, ce n'est pas long mais la phylogénie des métazoaires ne permet pas de placer les éponges et les cnidaires. L'ancêtre commun à l'Homme et aux éponges est très ancien à tel point que son signal phylogénétique a été perdu : les phylogénies moléculaires perdent tout signal au delà de 400 MA, elles ne permettent pas de remonter au dernier ancêtre commun.

Quel lien existe entre un événement individuel et la descendance de cet individu ?

S'il n'a pas de reproduction sexuée, il n'y a aucun problème.

Si la reproduction sexuée est obligatoire pour cet individu, on obtient des hybrides bizarres à fertilité très faible : l'évolution vers une nouvelle espèce semble passer par une phase difficile de très faible fertilité. Le succès est rare mais comme les duplications totales sont fréquentes, on peut penser que comme chez les végétaux ou chez les champignons, l'hybridation d'alloploïdes est fréquente. Au contraire les autopolloïdes produiront des gamètes (2n) et des œufs (3n).

Quel est l'effet du stress sur les réarrangements génomiques ?

Beaucoup de transposons sont différents et régulables : par exemple chez la Levure, beaucoup sont thermosensibles. En absence de reproduction sexuée, les transposons ont donc tendance à disparaître.

La reproduction sexuée apporte-t-elle plus de diversité ?

C'est compliqué.

Si l'organisme a le choix, il oublie la reproduction sexuée et s'oriente vers une croissance rapide. Les recombinaisons méiotiques ne réarrangent pas tous les allèles, car chez de nombreuses espèces les blocs haplotypiques sont répandus (c'est le cas de l'Homme) parfois jusqu'à 1 Mb du génome monoparental : dans la réalité, ces réarrangements se limitent à un ou au maximum deux crossing-over par bras de chromosome. Il y a donc peu de réarrangement allélique et chez les champignons de nombreux individus font les recombinaisons méiotiques avec eux-mêmes.

Que pouvez-vous dire des transferts horizontaux et des OGM ?

Les OGM semblent utiles dans la bioremédiation, par exemple en fabriquant des bactéries mangeuses de pétrole : d'ailleurs les OGMP naturels sont partout mais cela n'exclut pas un risque chez l'Homme.

En conclusion, le message à transmettre aux lycéens ?

Donner une bonne image, dynamique, du génome pour comprendre l'évolution : toutes les populations sont en variance interne continue, il n'y a rien qui ne change pas en permanence !