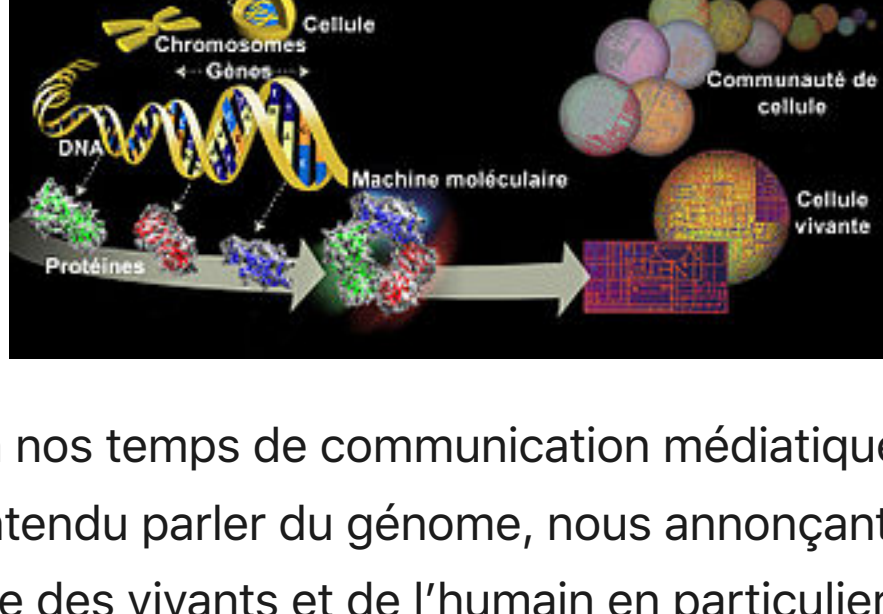


Diversité et génome

Professeur des Universités – génétique – Membre de l'Institut



Comme souvent en nos temps de communication médiatique déchaînée, nous avons beaucoup entendu parler du génome, nous annonçant l'arrivée à la connaissance ultime des vivants et de l'humain en particulier. En fait, on sait très peu de ces phénomènes fondamentaux. Plus les progrès scientifiques et techniques s'accumulent et plus on réalise tout ce que l'on ne connaît pas. Essayons de faire un retour en arrière pour comprendre comment les choses se sont développées dans les dernières décennies.

Vers le milieu du 20ème siècle, le génome était considéré selon un modèle stable, clair : un modèle simple de la structure génétique des individus. Le génome variait lentement, par l'accumulation de mutations redistribuées à chaque génération, par le jeu de la fécondation croisée. C'était une collection de gènes associés physiquement les uns aux autres et dont chacun portait l'information pour la structure d'une protéine. Il fallait 20 à 30000 gènes dans une cellule pour faire un génome. Cette notion résultait de très nombreux – et très brillants – travaux, depuis Mendel jusque dans les années 50 – 60.

Ce modèle a commencé à bouger vers 1950 parce que l'on a montré que de nouveaux gènes pouvaient s'intégrer à l'intérieur d'un génome. Notamment avec les premiers travaux de François Jacob et Elie Wollman sur les provirus des bactéries. Ce sont des ADN étrangers qui viennent s'insérer dans le génome. Cette notion a été étendue ensuite aux organismes supérieurs au point que toute une partie de notre génome est faite des restes de génomes de virus qui s'y sont intégrés. La notion fondamentale d'éléments transposables (des fragments du génome capables de se déplacer) établie par Barbara McLintock dans ces mêmes années a mis longtemps à être reconnue par le Prix Nobel en 1983, du fait de son caractère hétérodoxe.

Ce qui est venu compliquer le tableau a été le fait qu'un gène n'est pas un bloc unique, correspondant à une protéine. Un gène est en fait lu, dans beaucoup de cas, par morceaux, notamment chez les mammifères et l'homme.

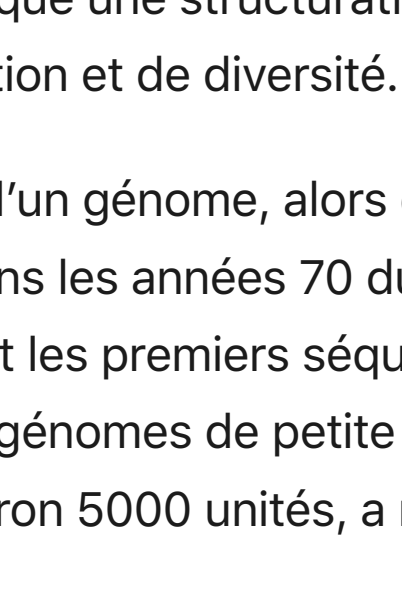
Le mécanisme de lecture va remettre ensemble ces morceaux, séparés par des régions au départ mal définies.

Cela a été appelé épissage, par analogie avec le moyen inventé par les marins pour réunir deux cordages. Mais une variation est apparue du fait que plusieurs lectures sont possibles. C'est l'épissage alternatif.

L'ordre d'association des morceaux de génome n'est donc pas immuable et ceci est encore compliqué par le fait que le code génétique est à trois lettres et qu'un simple décalage de phase peut introduire une lecture complètement différente pour la même séquence.

La notion « un gène pour une protéine » est donc complètement à revoir si l'on définit le gène (vieux problème ...) comme un segment de la molécule d'ADN. Un tel gène peut donner des protéines différentes et parfois très différentes. Il peut même y avoir des structures plus compliquées où, entre deux éléments, va s'inclure un autre, ce qui à la fois donnera une structure plus composite où peuvent loger plus d'informations dans un génome donné, et beaucoup de possibilités de variations et même de nouvelles structures ou gènes.

Ce qui a encore compliqué les choses a été les progrès de notre connaissance de la molécule d'ADN. C'est une molécule polymérique extraordinaire pour un chimiste puisque constituée de quatre unités seulement répétées environ 3,5 milliards de fois dans un génome humain et disposées dans un ordre strict puisque c'est le support de l'information.



Cette molécule de 20 Angstroms (20 milliardièmes de millimètre) de diamètre fait plusieurs centimètres de long. L'ensemble de notre génome fait plus d'un mètre de longueur et il est logé dans le noyau de chacune de nos cellules (quelques 50 microns de diamètre) ! L'image classique est, avec les mêmes proportions, un fil de pêcheur d'une fraction de millimètre de diamètre et de ... plusieurs kilomètres de longueur ! Et, qui plus est, rangé en bon ordre, dans une tasse à café !

Ceci a comme conséquence que l'on ne peut pas isoler cette molécule intacte car elle est trop fragile et implique une structuration très particulière qui est en elle-même une cause de variation et de diversité.

Etablir la séquence complète d'un génome, alors qu'on ne peut en obtenir que des morceaux, était encore dans les années 70 du domaine du rêve. Beaucoup de laboratoires y travaillaient et les premiers séquençages obtenus autour de 1980 furent ceux de virus aux génomes de petite taille. Le premier génome ainsi séquencé, qui comportait environ 5000 unités, a représenté plus de deux années de travail (1978 – 80).

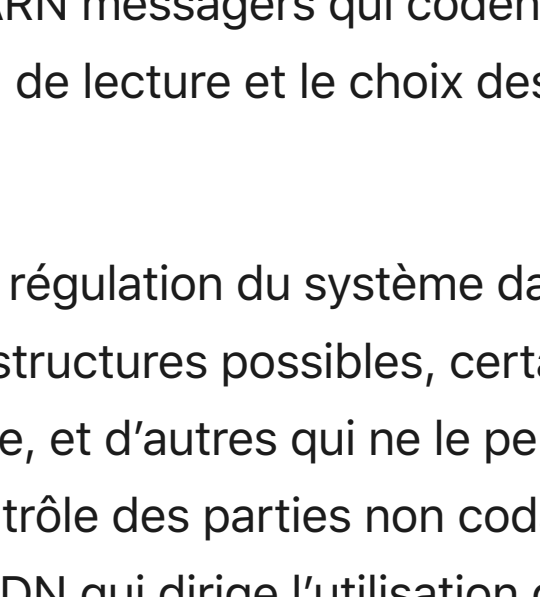
Mais aujourd'hui, le progrès technologique a été tel qu'à très court terme, établir le séquençage du génome d'un humain, avec des instruments entièrement robotisés, ne prendra que quelques heures !

Un autre progrès essentiel a été fait grâce à l'informatique qui permet de lire ces séquences de plusieurs milliards de lettres, d'y reconnaître des éléments du texte, trouver les motifs, ce qui a fait l'objet des informations dans les médias. La possibilité de lire la structure de nos protéines, avec les conséquences que l'on connaît, prédisposition à des maladies par exemple, a amené à l'idée que l'on pourrait « lire » les aptitudes de chacun, les maladies génétiques dont ils étaient porteurs, leur espérance de vie, etc.

En fait, on en rabat actuellement beaucoup car lorsqu'on analyse tout ce que l'on a obtenu, on se rend compte qu'on ne sait très peu de choses. Mais, cependant, s'ouvrent des voies nouvelles auxquelles on n'avait pas pensé mais qui sont très prometteuses pour la progression des connaissances.

On sait maintenant que tous les gènes qui codent pour les protéines ne représentent que 3 % seulement de l'ensemble de l'ADN d'une cellule. On y trouve effectivement des choses très intéressantes, comme par exemple des gènes dont les protéines produites sont altérées pour induire les maladies génétiques.

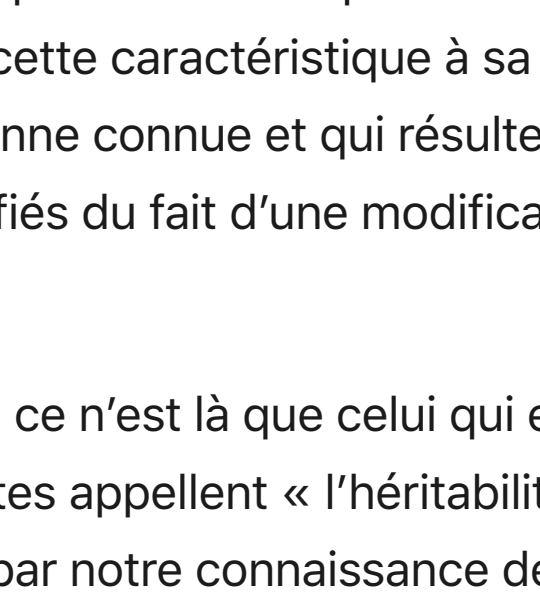
Notre connaissance ainsi brusquement explosée de la structure de nos protéines a fait faire des progrès considérables à la biochimie et à la physiologie. Nous avons également appris à modifier, principalement chez la souris, la structure d'un gène, et donc d'une protéine pour en analyser les conséquences.



Mais le gros problème, ce sont les 97 % du génome qui ne servent pas à maintenir l'information génétique à la structure des protéines. On a considéré un temps que la plus grande partie était un ADN poubelle (junk DNA en anglais) qui ne servait à rien et qui était constitué de résidus accumulés au cours de l'évolution. On a pensé que certains éléments pouvaient servir éventuellement pour l'évolution.

Mais ceci ne tient pas depuis que l'on dispose d'un certain nombre de génomes complets et que l'on constate que les séquences fonctionnelles sont très rigoureusement conservées. Elles sont le siège d'un certain taux de mutation qui est rigoureusement maintenu à chaque génération. Ce système est donc calibré pour ajuster ce nombre d'erreurs dans l'ADN. Si ce taux de mutation dépasse une certaine limite, par exemple à la suite d'une irradiation radioactive, on a un système de correction qui se met en action, ne remettant jamais le taux à zéro, mais à un taux que l'évolution a admis comme étant optimal.

S'il y a une conservation complète, cela veut dire qu'il y a une sélection.



On commence à avoir quelques idées sur ces séquences. Il y en a effectivement un petit nombre qui ne servent à rien. Ce sont les séquences hypervariables qui sont l'ADN des polices scientifiques. Elles permettent d'identifier chacun de nous d'une manière irréfutable car elles sont propres à chaque individu. Elles mutent au cours des générations mais sont inutiles pour le génome fonctionnel.

Mais cela n'est qu'une petite partie de l'ADN génomique. La plus grande partie, ce sont des séquences transcrites. Elles servent de matrice pour fabriquer non pas des protéines, mais essentiellement des molécules d'ARN. Il y en a des quantités et on en découvre de nouvelles classes tous les jours. Ces « ARN non codants », à l'inverse des ARN messagers qui codent pour des protéines, ont pour rôle d'ajuster le niveau de lecture et le choix des gènes qui sont lus dans la partie codante de l'ADN.

C'est donc un ensemble de régulation du système dans le noyau. L'ADN y est structuré. Il y a différentes structures possibles, certaines qui permettent la lecture du gène par exemple, et d'autres qui ne le permettent pas. Ce phénomène est sous le contrôle des parties non codantes pour les protéines. C'est donc une partie de l'ADN qui dirige l'utilisation de l'information génétique. On a comparé tout cela à un orgue dont les tuyaux servent à émettre chacun une note, mais qui ne sont mis en action au niveau de leur intensité ou de leur expression et pour que l'ensemble soit correct, par une machinerie commandée par le musicien.

Toutes ces fonctions de régulation à grande échelle constituent l'épigénétique et on sait que, dans certains cas, il peut y avoir des modifications dans cette organisation. Ces structures épigénétiques qui gouvernent l'expression du reste, peuvent être modifiées soit de façon stable, voire même instables. C'est ainsi que l'homme possède deux chromosomes différents : X et Y, alors que les femmes ont deux chromosomes X. Le chromosome X est très gros et contient donc beaucoup de gènes pour la synthèse des protéines. Les femmes le possèdent en double exemplaire, ce qui devrait produire deux fois plus des protéines concernées. Il y a, en fait, un système épigénétique qui, très tôt dans le développement de l'embryon, fait que l'un des chromosomes X devient au hasard totalement inactif chez la femme.

On revient donc au même niveau que chez l'homme.

L'autre exemple est que certaines variations épigénétiques peuvent être héritées comme notre laboratoire l'a récemment montré chez la souris. C'est une hérédité non mendélienne puisqu'elle ne passe pas par les systèmes de distribution des chromosomes à la suite des mutations. Le texte génétique n'est pas modifié mais l'utilisation que la cellule fait de ce texte modifié est différente et même parfois très différente, au point que l'on peut avoir un organisme nouveau – parfois très différent, au point que l'on peut avoir la taille des souris ordinaires, et qui transmet cette caractéristique à sa descendance en dehors de toute ségrégation mendélienne connue et qui résulte du fait qu'un petit nombre de gènes clés ont été modifiés du fait d'une modification des ARN qui contrôlent le système.

En touchant à ce problème, ce n'est là que celui qui est très à la mode qui est celui que les épidémiologistes appellent « l'héritabilité manquante ». C'est en fait une approche permise par notre connaissance des génomes puisque, maintenant, on peut comparer la séquence des gènes de nombreux individus. On peut regarder si certaines variations de séquence peuvent être corrélées à des caractères donnés.

Un exemple est celui de la taille, en partie héritée, mais si on cherche dans les génomes s'il y a quoi que ce soit que l'on puisse corréliser avec l'hérédité de la taille, on ne trouve pas. Une hypothèse possible est que ce n'est pas un gène qui serait en cause, mais un très grand nombre de gènes qui contribueraient chacun pour une part minime.

L'autre hypothèse, c'est qu'il s'agirait en fait d'une modification au niveau des structures des génomes qui gouvernent l'ensemble. On touche là à un problème que les épidémiologistes rencontrent de plus en plus, qui est le caractère familial de maladies pour lesquelles on ne peut définir un déterminisme génétique dans les termes de la génétique mendélienne. On sort ici de la génétique pour aller vers l'épigénétique.

Un des progrès majeurs apportés par notre connaissance des structures de génomes a été de préciser les mécanismes de l'évolution.

Un premier point fondamental a été d'asseoir définitivement la loi darwinienne de l'unité des génomes, tous dérivés d'un ancêtre commun (« common descent »). Il est clair que le mode de variation sera la mutation au sens large, c'est-à-dire le changement d'un monomère par un autre ou un petit fragment manquant dans la séquence.

Mais il y a aussi un phénomène très intéressant qui est de plus en plus mis en évidence, ce sont les grandes duplications. Au cours de l'évolution, il y a eu de façon discontinue, sporadique, des duplications importantes de génomes. A l'origine des mammifères, on pense qu'il y a eu ainsi une duplication très importante. Ceci ouvre évidemment des possibilités à l'évolution qui sont considérables dans la mesure où une mutation va habituellement empêcher la synthèse d'une protéine mais, d'un autre côté, il faut qu'il y ait des mutations pour créer des structures. De plus en plus, ces duplications semblent la règle. Dans les blocs qui codent pour les protéines, on va souvent retrouver la même carte très précise, les mêmes blocs, mais quelquefois, l'un aura été dupliqué et quelquefois, un autre aura été dupliqué en sens inverse.

Tout ça donne l'image d'une plasticité beaucoup plus grande par rapport au point de départ qui était le génome, la carte génétique clairement définie pour une espèce donnée. On a maintenant quelque chose qui est plus difficile à appréhender sur le plan de la construction d'hypothèses et le développement des technologies, une structure dynamique dont nous commençons seulement à connaître les modes de fonctionnement.