

LA Recherche

Dix-huit facettes d'un même concept

N°348 - 12/2001

Qu'est-ce qu'un gène ? A cette question en forme de quiz pour lycéens répondent dix-huit spécialistes français qui, dans leurs laboratoires, leurs amphithéâtres ou leurs cabinets d'avocats, ont quotidiennement à faire avec les « gènes ». Avec une telle diversité de définitions et d'aussi profondes contradictions, comment écrire aujourd'hui un manuel de génétique ?

Henri Atlan, biophysicien, directeur de recherche émérite à l'EHESS, auteur de *La Fin du tout génétique, vers de nouveaux paradigmes en biologie* (INRA Editions, 1999).

Il n'y a pas de réponse simple et univoque à cette question. Je ne peux donc que commenter une énumération, non exhaustive, de quelques définitions classiques :

On peut d'abord voir le gène comme une séquence d'ADN. Mais une même séquence peut participer au codage de plusieurs protéines...

Le gène peut alors être considéré comme un fragment d'ADN codant pour une protéine. La définition est ainsi projetée vers la protéine, selon une approche plus fonctionnelle. Mais que faire alors de l'ADN non codant ?

On peut aussi définir le gène, dans le contexte des théories de l'évolution, comme ce sur quoi agit la sélection naturelle. Le problème est que la sélection naturelle ne « voit » pas les fonctions des gènes, ne les trie pas selon leur plus ou moins grande valeur. La sélection naturelle voit en revanche les phénotypes*. Mais le phénotype ne se reproduit pas, puisqu'il est transmis par l'intermédiaire du gène, d'où la difficulté de sa définition comme unité de sélection.

Le gène serait-il une instruction de programme ? Voilà une définition contre laquelle je m'élève : le programme génétique n'est qu'une métaphore qui a pu avoir, en son temps, une certaine valeur heuristique, mais elle doit être aujourd'hui au minimum dépassée, sinon abandonnée car elle est source de confusion. On ne voit pas où se trouveraient ces

instructions de programme, ni dans quelle langue elles seraient écrites. Cette vision repose sur une confusion entre programme et codage.

Une définition encore pire serait d'appeler gène toute séquence d'ADN que l'on observe être associée statistiquement à un trait de caractère. C'est ce qu'on a pu lire dans les colonnes des plus grandes revues à propos du gène de l'homosexualité ou de l'alcoolisme. Ces séquences, le plus souvent non codantes, ne sont que des marqueurs. D'une simple corrélation statistique entre la présence d'un marqueur et un comportement plus ou moins bien défini on déduit une relation de causalité totalement hypothétique. On oublie trop souvent que les molécules d'ADN ne font rien par elles-mêmes, puisqu'elles sont inertes, contrairement aux protéines et aux ARN.

Pierre Chambon, biologiste moléculaire, directeur de l'institut de génétique et de biologie cellulaire et moléculaire de Strasbourg.

Personne ne sait plus vraiment ce que l'on entend par « gène », ni quelle définition adopter lorsqu'il s'agit de les compter. Il semblerait que les gènes correspondent aujourd'hui aux différentes régions du génome qui sont transcrites de façon indépendante (voir schéma) : un gène serait ainsi une « unité de transcription ». Mais savoir qu'une région est transcrite est en soi peu informatif. Tout au plus, cela nous apprend que telle région du génome contient des informations pour le codage de protéines, sans dire combien ni lesquelles. En effet, une unité de transcription peut posséder de multiples promoteurs*, ainsi que de multiples sites de terminaison permettant la synthèse de multiples transcrits ; de plus, chaque transcrit peut à son tour correspondre à plusieurs protéines par le jeu de l'épissage alternatif (voir schéma).

Elle est bel et bien morte, l'ancienne définition fonctionnelle du gène, comme l'équation « un gène, une protéine ». Alors que signifie le fait que l'homme et la mouche drosophile posséderaient environ 30 000 gènes ? Tels qu'ils sont actuellement répertoriés, pas grand-chose, car nous ignorons si ces 30 000 gènes codent pour 100 000 protéines chez la drosophile, et quelques millions chez l'homme. Ce qui compte *in fine*, c'est la fonction, le nombre de protéines codées par un *génome* donné, car ce sont elles qui sont responsables du phénotype, donc de ce que nous sommes. Aussi ne peut-on, comme certains se sont empressés de le faire avec soulagement, et pour des raisons idéologiques évidentes, déduire du nombre de gènes chez l'homme et chez la mouche, que la part déterminée génétiquement chez l'homme est nécessairement accessoire.

Daniel Cohen, biologiste moléculaire, directeur scientifique de la société Genset.

Votre question tombe à pic : avec les travaux sur l'analyse du génome humain, on ne sait plus ce qu'est un gène. Classiquement, on disait qu'un gène est un segment d'ADN qui code pour une protéine. Mais cette définition apparaît fautive. D'une part, il y a des segments de génome qui peuvent être transcrits en plusieurs ARN, qui peuvent à leur tour donner plusieurs protéines. D'autre part, on se rend compte que les deux brins de l'ADN d'un même fragment peuvent tous les deux coder pour des protéines différentes. La transcription saute parfois même d'un brin à l'autre. Enfin, la transcription d'une région peut être sous le contrôle de régions régulatrices situées très loin sur le génome. Comment alors borner les limites du gène, en dresser la géométrie ?

La définition de l'ARN reste en revanche sans ambiguïté. Je pense donc qu'il va falloir redescendre la définition du gène de la génomique vers la transcriptomique, car l'individualité est dans les ARN. Une telle redéfinition centrée sur l'ARN entraînerait une complexification redoutable. Il faudrait en particulier revisiter toutes les données sur le polymorphisme* de l'ADN. Tout ce que nous en savons aujourd'hui porte sur l'ADN sans que l'on en connaisse les conséquences en termes d'expression génétique. De plus, on trouve de plus en plus souvent des régions de polymorphisme fonctionnellement importantes et situées hors des régions codantes, donc probablement dans des régions régulatrices. Mais on ne sait pas quels sont les gènes dont l'expression est régulée par ces régions !

Alain Fischer, spécialiste de thérapie génique, directeur de l'unité Inserm « Développement normal et pathologique du système immunitaire » (Paris).

J'en resterai à une définition très classique : un gène est une structure constituée d'ADN, qui code pour un ARN permettant à son tour la synthèse de protéines.

Philippe Froguel, spécialiste de génétique moléculaire humaine, directeur de l'unité CNRS « Génétique des maladies multifactorielles » à l'Institut Pasteur de Lille.

Un gène est une unité qui détermine un caractère phénotypique. Certes, voilà qui est un peu réducteur : aux 40 000 gènes humains ne correspondent pas 40 000 phénotypes. Par exemple, un gène peut être impliqué dans plusieurs maladies. Mais cette définition représente le « gène idéal » qu'on étudie en génétique humaine.

Cela dit, je pense que ce n'est pas l'étude du génome humain qui va nous faire progresser dans la compréhension des maladies, mais plutôt celle du *protéome*. On se rend compte, par exemple, du rôle physiologique extrêmement important de la maturation* des protéines et

de leur niveau d'expression. De tels phénomènes ne peuvent être étudiés que par une approche protéomique, et non génomique.

En fait, je crois même que l'avenir est dans l'approche du *métabolome*, c'est-à-dire l'étude systématique des molécules que l'on trouve dans un fluide, par exemple le sang. Le métabolome me paraît le niveau pertinent parce qu'il est la résultante de toutes les influences : celle du génome (la structure des protéines), celle de la cellule (la maturation des protéines) et enfin de l'environnement. On étudie ainsi un phénomène tardif par rapport à l'expression du gène, résultant de la somme des interactions existant dans un tissu. Et, plus on se situe tardivement, mieux on pourra cerner le rôle des gènes et de leur expression.

Jean-Christophe Galloux, professeur de droit à l'université Paris-II, ancien conseiller de l'Unesco pour la déclaration sur le génome humain.

Le droit est une science normative, pas cognitive. Il n'y a donc pas de définition juridique du « gène » et il n'y en aura jamais. En revanche, le juriste s'efforce de classer le « gène », quelle que soit la définition que les scientifiques en donnent, parmi les catégories du droit. Première question : le gène humain appartient-il à la catégorie des choses ou à la catégorie des personnes ? La plupart des éléments biologiques détachés du corps humain étant considérés comme des choses, les juristes considèrent en majorité qu'il en va de même pour les gènes. Deuxième question : le « gène » est-il une chose appropriable (comme un animal) ou une chose non appropriable (comme le corps humain ou ses éléments qui, selon la loi bioéthique de 1994 désormais intégrée au code civil, « ne peuvent faire l'objet d'un droit patrimonial ») ? Cette discussion n'est pas réglée. Pour ma part, je pense que le gène peut faire l'objet d'une appropriation. De même que pour le sang, il peut y avoir commerce de gène (au sens juridique d'échange), éventuellement rémunéré, mais encadré. A mon sens, la loi de 1994 n'a fait qu'exprimer maladroitement l'idée qu'une personne ne peut se vendre en morceaux.

Il ne faut pas oublier que le gène n'est pas un objet comme un autre, puisqu'il reste rattaché à sa personne source, puisque la connaissance du gène peut renseigner sur la personne dont il est issu. Il faut donc défendre les droits à l'intimité de la personne source tout en permettant l'appropriation du gène : c'est l'une des contradictions juridiques que posent les banques de données génétiques.

Patrick Gaudray, virologue, directeur scientifique adjoint du département Sciences de la vie du CNRS.

- Un gène comme unité de transcription ? Oui, mais un même gène peut être transcrit selon des cadres de lecture très différents et ainsi donner plusieurs ARN messagers (ARNm) différents.
- Un gène comme élément codant pour une protéine ? Oui, mais avec l'épissage alternatif, un gène peut donner plusieurs protéines.
- Un gène comme unité d'information, selon la définition génétique classique ? Oui, sans doute. Mais cette définition pose aussi problème : elle suppose un phénotype connu. Or, ce n'est pas le cas lorsque, et c'est de plus en plus fréquent, on dispose de la séquence d'un génome sans connaître les phénotypes associés. C'est pourquoi il faut passer par une étape de travail fonctionnel, pour comprendre le rôle de chaque région, avant de pouvoir définir les gènes en « rebouclant la boucle » qui mène du génotype au phénotype.

Les apports du séquençage ont à l'évidence considérablement complexifié la notion de gène : que dire des gènes eucaryotes situés à l'intérieur d'un intron* ? Que dire des gènes qui se lisent en sautant d'un brin d'ADN à l'autre ? Chacun a, dans sa spécialité, une compréhension instinctive de ce qu'est un gène. Aujourd'hui, on ne peut que souhaiter une définition un peu plus consensuelle et un peu plus objective.

Jean Générmont, généticien des populations, professeur émérite à l'université Paris Sud, ancien président de la Société française de génétique.

A brûle-pourpoint, je dirai qu'un gène est un segment d'ADN (en oubliant les virus à ARN qui viennent compliquer les choses) dont la fonction est de spécifier un ARN : celui-ci peut être un ARN messager*, un ARN ribosomique*, un ARN de transfert*, ou l'un des autres petits ARN.

En tant que généticien des populations, j'aimerais commenter cette définition en y ajoutant les notions d'allélisme* et de ségrégation mendélienne, mais cela nous emmènerait trop loin. Savez-vous qu'à la Société française de génétique nous avons eu le projet de rédiger un lexique de la génétique ? Il n'a pas abouti, mais je me souviens que l'entrée « gène » nous avait posé de très grandes difficultés.

Pierre-Henri Gouyon, biologiste des populations, directeur du laboratoire CNRS d'écologie systématique et d'évolution (Orsay).

Tout d'abord, un gène est une information, et non de la matière. Cette distinction a des conséquences juridiques : lorsqu'on dépose un brevet sur une plante transgénique, on brevète l'information introduite, pas la séquence d'ADN. Cela dit, je définirai un gène comme une information portée par un acide nucléique, qui peut se recopier et coder pour une fonction.

Pour définir formellement ces termes, on peut définir deux relations d'équivalence* parmi l'ensemble des gènes. La première est la relation d'homologie. Des gènes homologues vont s'apparier en face lors de la méiose. Une classe d'équivalence de gènes pour la relation d'homologie s'appelle un *locus*. La seconde est la relation d'isoaction. Parmi les gènes homologues, c'est-à-dire à l'intérieur d'un locus, l'ensemble des gènes qui donne le même phénotype est dit isoactif. Une classe d'équivalence pour la relation d'isoaction est *un allèle*. La notion d'allèle est donc définie par le phénotype et non par la séquence, qui n'est qu'un phénotype parmi d'autres.

Philippe Jarne, évolutionniste, directeur de recherche au Centre d'écologie fonctionnelle et évolutive du CNRS (Montpellier).

Je proposerai deux définitions :

La première est celle d'une unité d'information, transmise entre les générations. Le point important est que cette unité - ou son expression - est soumise à la sélection naturelle. Tout en sachant qu'une partie de la sélection ne se fait pas sur le génotype, mais sur le phénotype... Il faut aussi nuancer l'idée du gène comme « unité d'information » parce qu'on sait bien que les gènes ne fonctionnent pas indépendamment. On peut, par exemple, supposer qu'un phénotype dépend d'un grand nombre de gènes et que la relation phénotype-génotype doit être approchée à l'aide de lois statistiques même si la formalisation mathématique est extrêmement ardue.

La deuxième définition est une définition de généticien des populations : le gène peut être défini en termes statistiques comme un ensemble de variables aléatoires. Le gène présente en effet de nombreux états, les allèles, et la transmission de ces derniers au travers des générations peut être analysée avec des méthodes statistiques.

Dans ces deux définitions, on ne cherche pas à faire du gène un objet aux frontières bien délimitées. Cela ne veut pas dire que la structure moléculaire du gène soit sans importance. De fait, les évolutionnistes tirent partie de toutes les connaissances au niveau moléculaire pour mieux comprendre quelles régions de la séquence sont importantes dans l'évolution des gènes.

Bertrand Jordan, biologiste moléculaire, président de la Génopole de Marseille, auteur des *1 mposteurs de la génétique* (Le Seuil, 2000).

Je donnerai une définition très opérationnelle du point de vue d'un biologiste moléculaire. Un gène est une région d'ADN qui code pour une (ou des) protéine(s), *via* l'épissage alternatif et la maturation. Cette région comprend des séquences codantes et des séquences non

codantes, dont des séquences de régulation, l'ensemble s'étendant sur des milliers de nucléotides.

Axel Kahn, biologiste moléculaire, directeur de l'Institut Cochin de génétique moléculaire (Paris).

Un gène est un polymère chimique formé d'un enchaînement de nucléotides* qui a la propriété de coder une protéine.

Fondamentalement, le gène est l'un des codéterminants de l'apparition de la vie, entendue comme apparition de la cellule vivante. La seule possibilité pour marquer le début de la vie est de marquer l'avènement d'un ensemble fini, limité par une membrane, et doté d'un métabolisme dont la reproductibilité comme la variabilité sont liées à un programme génétique. En d'autres termes, le gène est l'un des éléments, et un élément important en ce qu'il est le siège des mutations, qui permet à l'évolution de débiter.

André Langaney, généticien, professeur à l'université de Genève et au Musée de l'Homme (Paris).

Le physicien Richard Feynman disait de la « force » et d'autres concepts qu'ils sont d'autant plus utiles qu'ils sont flous. Cette réflexion s'applique très bien au concept de gène. Dans les années 1970, quand je commençais à enseigner la génétique, on recensait pas moins de onze concepts de gène. A l'époque, on avait même essayé de nommer l'objet correspondant à chacune de ces définitions : l'unité de transcription (le cistron), l'unité de recombinaison, de mutation... Mais c'était une tâche désespérée. Il n'y avait pas de raisons de s'arrêter à onze définitions, tant on pouvait en trouver de nouvelles. Je ne dis pas cela pour le regretter. Comme le notait Feynman : « *Je suis bien décidé à continuer à explorer la nature sans en donner de définitions .* » Par contre, je retiens ce qu'il y a d'unificateur dans tout cela : le concept d'unité d'information, transmis de génération en génération.

Daniel Louvard, biologiste cellulaire, directeur scientifique de l'institut Curie (Paris).

Je donnerai deux définitions. La première correspond à l'idée qu'on se faisait du gène avant la biologie moléculaire, c'est-à-dire avant qu'on en connaisse la structure. Un gène, c'est quelque chose qui contrôle un phénotype, c'est-à-dire une caractéristique d'un être vivant : la couleur des yeux des drosophiles de Morgan, l'aspérité des petits pois de Mendel... Pour les généticiens de l'époque, je ne suis pas certain qu'il y ait eu une notion informative. Un gène était surtout quelque chose qui définissait une caractéristique transmissible de manière héréditaire.

La deuxième définition est apparue après la découverte que l'ADN est le support de l'hérédité. On a alors intégré à la première définition des

éléments de la théorie de l'information, en déclarant que l'ADN contient de l'information. Comme tous les caractères sont codés dans l'ADN, on a pu lire des séquences, des fragments d'information, qui peuvent être non seulement transmis (ce qui a permis de faire le lien avec la première approche) mais aussi altérés, ce qui donne les mutations. La notion a encore évolué avec la découverte que la structure des gènes était éclatée, avec des introns et des exons, grâce auxquels on peut générer une extraordinaire diversité de protéines à partir d'un petit nombre de gènes, *via* le phénomène d'épissage. Les différents types cellulaires composant un organe peuvent ainsi produire des protéines différentes en utilisant seulement une petite partie des gènes. Quand une cellule se différencie, ou quand elle est malade, elle synthétisera des combinaisons de gènes tous différents.

On est donc passé d'un concept simple, tiré des bactéries et des êtres unicellulaires, où la panoplie de gènes est limitée à quelques milliers de gènes, à un concept plus compliqué, où les combinaisons permettent de produire de la diversité, ce qui nécessite de penser en termes de réseaux de gènes.

François Rechenmann, bio-informaticien, directeur de recherche à l'Inria (Grenoble).

La notion de gène varie parce que les problématiques biologiques sont différentes. Un biologiste choisit en quelque sorte le modèle qui l'arrange. En tant que bio-informaticien, je ne peux que reproduire cette variété de problématiques. Chaque programme informatique qui dit trouver les gènes dans une séquence brute doit ainsi avoir un modèle de gène différent(II). Par exemple, mon laboratoire a développé un programme pour la détection des gènes dans les séquences de génomes bactériens. Une de nos stratégies est de rechercher les codons START le plus en amont possible d'un codon STOP*, puis d'appeler « gène » l'ensemble borné par ces codons, si l'on trouve en amont une séquence de fixation du ribosome et que l'ensemble fait plus de 300 nucléotides.

Que le terme gène ait des acceptions différentes, cela ne me choque guère. Ce qui me perturbe davantage, en tant que bio-informaticien, c'est de retrouver cette pluralité dans les bases de données de séquences d'ADN. Quand vous discutez avec quelqu'un, vous pouvez toujours vous mettre d'accord en quelques minutes sur ce qu'il entend par gène. Mais comment faire lorsque vous consultez une base de données ? Cette question devient cruciale lorsque l'on veut fusionner ou unifier des données entre des bases qui n'ont pas la même définition du gène. Il existe certes des travaux de consensus, qui tentent de fonder une « ontologie » en identifiant les concepts et les relations qui fondent la

biologie moléculaire. Ces chercheurs ont l'ambition, à mon avis une illusion, de se mettre d'accord sur ce qu'est un gène. Je parle d'illusion parce que je ne suis pas certain que l'on ait avantage à standardiser les définitions du gène, dans un domaine où l'effervescence est plutôt signe de vitalité.

Pierre Sonigo, virologue à l'institut Cochin de génétique moléculaire (Paris), co-auteur (avec Jean-Jacques Kupiec) de *Ni Dieu ni gène, pour une autre théorie de l'hérédité* (Le Seuil, 2000).

Selon la définition initiale, issue des travaux de Mendel, le gène est une abstraction représentant le caractère héréditaire. Cette représentation est non matérielle et arbitraire. De même que le nombre est une représentation conventionnelle qui permet d'effectuer des opérations mathématiques, le gène, ainsi défini et associé aux lois de Mendel, permet de faire des calculs sur la transmission héréditaire d'un caractère. Plus tard, une seconde définition est apparue, qui visait à donner une réalité matérielle et objective au gène. C'est le gène de la biologie moléculaire moderne : un segment d'ADN qui code pour la synthèse d'une protéine. Dans cette nouvelle définition, le gène devient ainsi le créateur, l'explication du caractère héréditaire.

Ces deux définitions sont, à mon avis, incompatibles, ce qui explique les nombreuses incohérences théoriques de la biologie contemporaine. Pour que la génétique mendélienne ait un sens, le gène et le caractère doivent être indissociables. Pour que cette relation essentielle soit maintenue dans la définition de la biologie moléculaire, il faut que le chemin menant de l'ADN au caractère - *via* la protéine - soit sans ambiguïté. Or, ce n'est vrai que dans les cas, très rares, où la protéine elle-même, voire son absence, est le caractère. Pour les caractères plus complexes, on ne retrouve qu'une corrélation statistique avec un segment d'ADN. Cette corrélation est souvent faible, elle n'atteint en tout cas presque jamais 100 %. De plus, une corrélation, même forte, n'est pas une relation de cause à effet. De fait, un même segment d'ADN peut être associé à des caractères différents. La gravité d'une maladie héréditaire peut, par exemple, varier pour une même séquence d'ADN. A l'inverse, avec des ADN différents, on peut avoir le même caractère. Par exemple, le déterminisme du sexe peut être altéré par de multiples mutations dans des endroits différents. Dès lors, il est impossible de parler du « gène du caractère » et de concilier la définition de la biologie moléculaire avec la définition mendélienne. Autrement dit, l'ADN ne porte pas les gènes.

Louis Thaler, professeur émérite à l'université de Montpellier-II, président du Bureau des ressources génétiques.

Je suis très vieux jeu : j'appelle gène un déterminant d'un caractère. Je reviens donc à la définition de Mendel, qui ignorait le mot mais avait parfaitement perçu ce concept. Comme évolutionniste, je m'intéresse plus aux fonctions des gènes, donc à l'information biologique qu'ils sont à mes yeux, qu'à leur support matériel qu'est l'ADN. C'est aussi une façon de dire que tout l'ADN n'est pas formé de gènes.

Jacques Warcoin, conseiller en propriété industrielle, spécialisé en biotechnologie.

Pour moi, un gène, c'est tout simplement la séquence d'ADN qui code pour une protéine. C'est une définition utilitaire. Quand on dépose un brevet, on étend évidemment la définition le plus possible pour avoir une protection maximale, y compris parfois à des introns qui peuvent servir pour certaines applications biotechnologiques.

Un problème peut être de savoir si la définition du gène incorpore également les séquences de régulation, en amont et en aval de la séquence codante proprement dite. On vient de le constater lors du procès qui a opposé la firme californienne Amgen à l'allemande TKT à propos du brevet sur la production d'érythropoïétine (EPO). Chaque partie a fait défiler les plus éminents experts pour qu'ils définissent leurs notions de gènes, en particulier si les gènes devaient inclure les séquences régulatrices. Les controverses scientifiques sur la définition du gène ont clairement des répercussions sur les batailles de brevets en biotechnologie.

Propos recueillis par Nicolas Chevassus-au-Louis