

GÉNOME HUMAIN : ORDRE ET NÉCESSITÉ

Giorgio BERNARDI*

La multiplication des données obtenues ces dernières années sur les propriétés compositionnelles du génome, et plus encore sur la distribution des gènes, permettent d'éclairer le génome des vertébrés d'un jour nouveau. Les gènes, loin d'être dispersés au hasard sont concentrés dans certaines régions chromosomiques. Bien que la distribution des gènes soit largement conservée au cours de l'évolution, la composition de ces régions a évolué vers un enrichissement en guanine et cytosine chez les vertébrés à sang chaud. Ce changement est vraisemblablement le fruit d'avantages sélectifs.

Depuis l'apparition du terme génome, proposé il y a soixante-quinze ans par le botaniste allemand Hans Winkler pour désigner le jeu de chromosomes haploïdes des eucaryotes, beaucoup de données sont venues enrichir le concept de génome. L'époque où le génome était réduit à sa définition purement opérationnelle comme l'ensemble des gènes et des séquences intergéniques, a évolué. L'existence d'interactions structurelles et fonctionnelles entre la minorité de séquences codantes et la majorité des séquences non codantes longtemps soupçonnée, est aujourd'hui précisée grâce à la découverte des propriétés compositionnelles du génome. Établies principalement à partir du génome nucléaire des vertébrés (1, 2, 3), ces connaissances haussent le génome au niveau d'un système intégré répondant à des règles imposées par un véritable code génomique.

Isochores : les pièces d'une mosaïque

Le génome humain est une mosaïque d'isochores, c'est-à-dire de longs segments d'ADN homogènes en bases, et appartenant à un petit nombre de familles qui couvrent une gamme de composition extrêmement large (voir figure 1A). Les isochores peuvent être divisées en deux groupes : le premier,

représentant les deux tiers du génome et composé de deux familles pauvres en GC (L1 et L2) ; le second, couvrant le tiers restant, et comprenant trois familles riches en GC (H1, H2 et H3) (voir figure 1B). Leur taille a été définie grâce à la cartographie compositionnelle (2), qui consiste à déterminer, par analyse en base de YAC (yeast artificial chromosomes) contigus par exemple, la composition de longues régions du génome. Appliquée à trois régions du génome humain, cette technique a montré que la taille des isochores était comprise entre 0,2 et 1,3 Mb (mégabase). Ce n'est que très récemment (4) qu'a été décrite la nature d'une frontière entre isochores, en l'occurrence au niveau du locus majeur d'histocompatibilité. Cette frontière, très nette, est située entre une isochore de la famille H2 et une autre de la famille L. Elle est constituée par une série de séquences Alu I (Sines) du côté GC-riche, et d'une série de séquences L1 (Lines) du côté GC-pauvre (les Sines et les Lines sont respectivement des séquences répétées courtes riches en GC, et longues pauvres en GC). Il sera intéressant de voir si cette situation est spécifique de cette région du génome, ou si elle se retrouve au niveau d'autres frontières entre isochores.

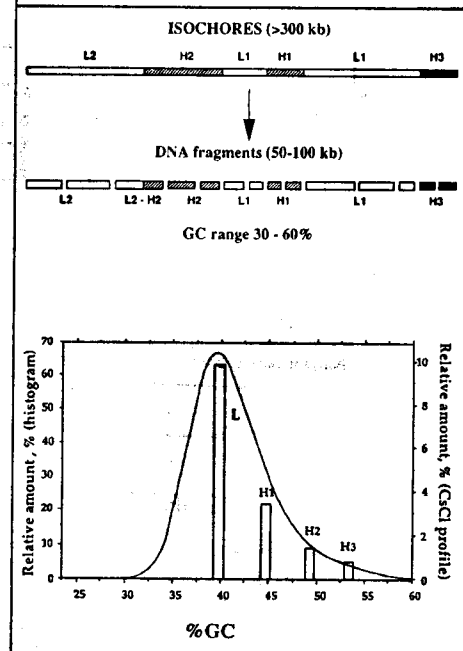
Le génome a ses phénotypes

Les distributions compositionnelles de longs fragments d'ADN (>100 kb, soit ceux qui constituent les préparations courantes d'ADN) des exons (et des bases en troisième position des codons) et des introns, forment des profils ou *patterns* (1, 2, 3) qui peuvent être décrits comme des phénotypes du génome. En effet, ils diffèrent non seulement entre

FIGURE 1 - LES ISOCHORES DU GÉNOME HUMAIN

A. Schéma de l'organisation en isochores du génome humain.

B. Histogramme des familles d'isochores du génome humain. Les quantités relatives des composants majeurs de l'ADN dérivées des familles d'isochores L (c'est-à-dire L1+L2), H1, H2, H3 sont superposées au profil en gradient de chlorure de césium de l'ADN humain.



les vertébrés à sang froid et ceux à sang chaud, mais également entre les mammifères et les oiseaux, et même entre les muridés et la plupart des autres mammifères (voir figure 2A et 2B). Ainsi, les fragments d'ADN des vertébrés à sang froid ont une composition en bases beaucoup plus restreinte que ceux des vertébrés à sang chaud. Parmi ces derniers, le génome des oiseaux atteint des niveaux de GC supérieurs à ceux des mammifères. Par ailleurs, la distribution compositionnelle du génome des muridés est plus serrée que celle des autres mammifères.

Un code pour le génome

La composition en bases des exons (ou de chacune des trois différentes positions des codons) et des isochores qui contiennent les gènes correspondants n'est pas aléatoire (1). Il existe, en effet, des corrélations

* Institut Jacques-Monod, laboratoire de génétique moléculaire, 2, place Jussieu, 75005 Paris.

(1) G Bernardi *et al* (1985) *Science* 228, 953-958.

(2) G Bernardi (1989) *Ann Rev Genet* 23, 637-661.

(3) G Bernardi (1993) *Gene* 135, 57-66.

(4) T Fukagawa *et al* (1995) *Genomics* 25, 184-91.

FIGURE 2 - LES PROFILS OU PATTERNS COMPOSITIONNELS DES GÉNOMES DES VERTÉBRÉS

A. Histogrammes montrant les quantités relatives, les densités de flottaison modales et les niveaux de GC des composants majeurs de l'ADN chez le xénope, le poulet, la souris et l'homme, estimés après fractionnement d'ADN par centrifugation préparative en gradient de densité en présence de ligands séquence-spécifique. Les composants majeurs de l'ADN sont les familles des grands fragments d'ADN (voir figure 1a) dérivés des différentes familles d'isochores. Les composants satellites et mineurs de l'ADN (tels que l'ADN ribosomique) ne sont pas montrés dans ces histogrammes.

B. Distribution compositionnelle des troisièmes positions des codons des gènes des vertébrés. Le nombre de gènes pris en compte est indiqué. Une fenêtre de 2,5% de GC est utilisée. La ligne en pointillé à 60% de GC est montrée pour référence.

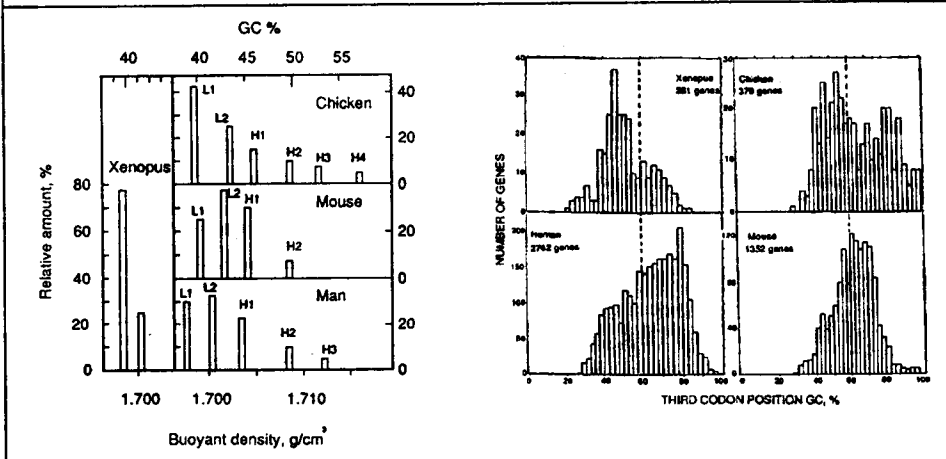
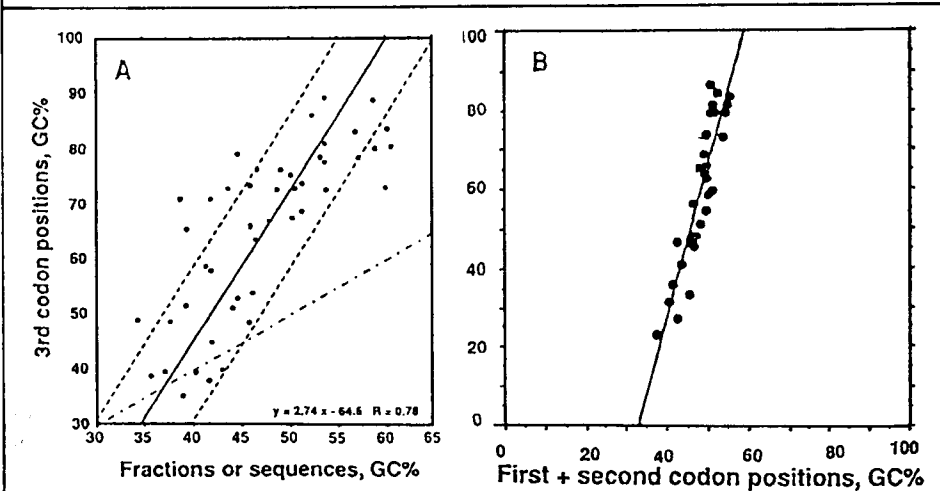


FIGURE 3 - LE CODE GÉNOMIQUE

A. Diagramme des niveaux de GC des troisièmes positions des codons des gènes humains en fonction des niveaux de GC des fractions de l'ADN ou des séquences étendues dans lesquels les gènes sont localisés. Le coefficient de corrélation et la pente sont indiqués. La ligne en pointillé-tiret est la ligne diagonale (pente = 1). Les niveaux de GC des troisièmes positions des codons tomberaient sur cette ligne s'ils étaient identiques aux niveaux de GC de l'ADN qui entoure les gènes correspondants. La ligne en pointillé indique un domaine de GC de $\pm 5\%$ de GC autour de la droite.

B. Diagramme des niveaux de GC des troisièmes positions des codons des gènes procaryotique et eucaryotique en fonction des niveaux de GC des premières et deuxième positions. Toutes les valeurs sont des moyennes par génome, ou par compartiment de génome dans le cas des génomes compartimentés compositionnellement.



lations entre ces différents éléments (voir figure 3A) : les séquences codantes ne représentent que 3% du génome, les 97% restantes correspondent à des séquences non codantes. De même, il y a une corrélation universelle (6) entre les niveaux de GC des différentes positions des codons. Par exemple, entre le niveau de GC en troisième position (GC_3), et celui de la première et deuxième positions (voir figure 3B). L'ensemble de ces corrélations représente un code génomique (3, 5) apparemment dû à des contraintes compositionnelles opérant dans la même direction (vers une augmentation de GC ou d'AT), mais dont les amplitudes ont des degrés variables selon qu'il s'agit de séquences codantes ou non codantes, ou selon la position de la base dans le codon.

Des gènes à leur place

Les corrélations compositionnelles entre le niveau de GC des troisièmes positions des codons et celui des isochores enfermant ces gènes ont permis de découvrir la surprenante distribution des gènes dans le génome humain (3, 7). La concentration en gènes est faible et constante dans les isochores L1 et L2, augmente avec l'accroissement de GC dans les familles H1 et H2, et atteint un maximum dans la famille H3, la plus riche en GC. Dans cette famille, la concentration en gènes est au moins 20 fois supérieure à celle des isochores pauvres en GC (voir figure 4A). Si l'on porte les pourcentages cumulés d'ADN ou de gènes en fonction des niveaux de GC ou de GC_3 , on obtient une courbe sigmoïde dans le cas de l'ADN (ce qui était prévisible du fait de la forme en cloche du profil en CsCl) et une droite dans le cas des gènes (voir figure 4B).

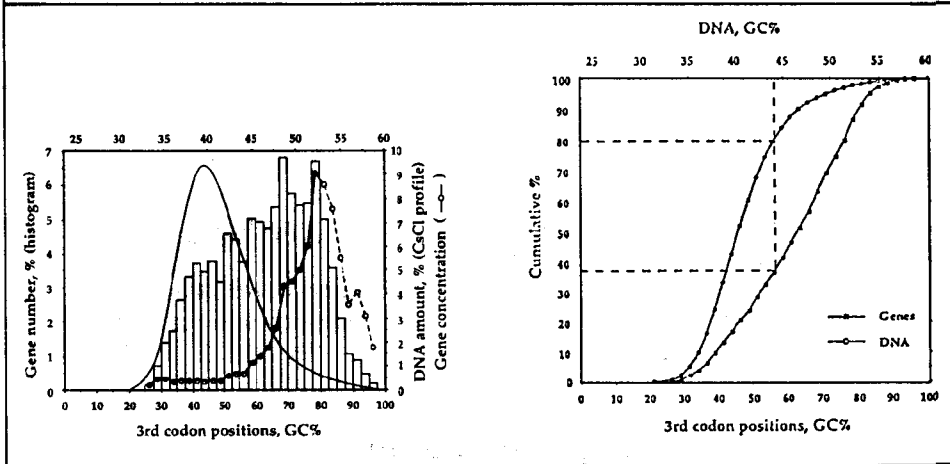
Cette distribution au sein du génome va de pair avec des structures spécifiques de la chromatine. Ainsi, les isochores les plus riches en gènes présentent une structure chromatiniennne « ouverte », comme le montre l'accessibilité aux DNases, la faible quantité d'histone H1, l'acétylation des histones H3 et H4 et l'important espacement entre nucléosomes (8). En outre, les taux de transcription et de recombinaison sont les plus élevés, et la réplication a lieu à un stade précoce du cycle cellulaire.

- (5) G Bernardi (1990) *Biofutur* 94, 43-46.
- (6) G D'Onofrio & G Bernardi (1992) *Gene* 110, 81-88.
- (7) D Mouchiroud *et al* (1991) *Gene* 100, 181-187.
- (8) J Tazi & A Bird (1991) *Cell* 60, 909-920.

FIGURE 4 - LA DISTRIBUTION DES GÈNES

A. Profil de la concentration en gènes du génome humain obtenu en divisant les quantités relatives des gènes de chaque intervalle de 2,5% de GC de l'histogramme, par les quantités relatives correspondantes d'ADN déduites du profil de gradient en chlorure de césium. Les dernières valeurs de concentration sont incertaines car elles correspondent à des valeurs très basses en GC (G Bernardi [1993]).

B. Diagramme cumulatif des quantités relatives d'ADN et de gènes en fonction du GC de l'ADN ou du GC₃ des séquences codantes.



savoir les plus résistantes à la dénaturation, et sont localisées essentiellement au niveau des télomères.

Dans le génome humain, les isochores des familles H2 et H3 sont localisées dans les bandes T (10, 11) alors que les bandes R' (c'est-à-dire les bandes R à l'exclusion des bandes T) comprennent aussi bien des isochores riches en GC (famille H1)

que pauvres en GC. Enfin, les bandes chromosomiques G(iemsa) sont formées presque exclusivement d'isochores pauvres en GC (voir figure 5). Le fait que la majorité des gènes soit localisée au niveau des bandes T présente vraisemblablement un avantage fonctionnel : il faciliterait les télomères étant associés à la matrice et à l'enveloppe

Les isochores et les bandes chromosomiques

Les chromosomes métaphasiques humains présentent divers types de bandes selon le traitement auquel ils sont soumis (colorants, température, enzymes protéolytiques ou DNases). Les bandes G(iemsa) positives représentent environ 50% des bandes chromosomiques, l'autre moitié étant constituée de bandes G(iemsa) négative. Ces dernières ne sont autres que les bandes R(verse), obtenues par dénaturation thermique. Quant aux bandes T(élomériques), identifiées par B Dutrillaux (9), elles forment un sous-ensemble des bandes R, à

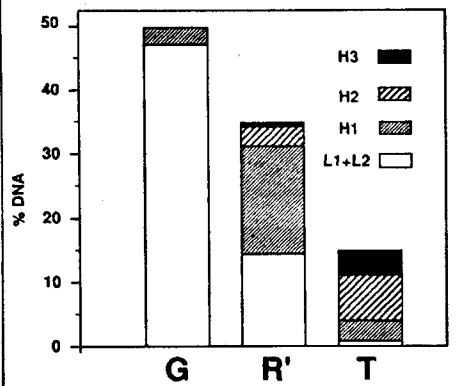
(9) B Dutrillaux (1973) *Chromosoma* 41, 395-402.

(10) S Saccone *et al* (1992) *Proc Natl Acad Sci* 89, 4913-4 17.

(11) S Saccone *et al* (1993) *Proc Natl Acad Sci* 90, 11929-11933.

FIGURE 5 - LES ISOCHORES ET LES BANDES CHROMOSOMIQUES

Schéma des quantités relatives des familles d'isochores L1+L2, H1, H2 et H3 dans les bandes G(iemsa), les bandes R' et T (élomériques) ; les bandes R' sont les bandes R(verse) à l'exclusion des bandes T qui en forment un sous-ensemble.



du noyau – l'exportation des produits de transcription dans le cytoplasme où ils seront traduits sur les ribosomes.

Les données les plus complètes sur la cartographie du génome humain recouvrent 80% de la carte physique (12), les 20% restants correspondant, pour l'essentiel, aux bandes T. On peut donc estimer, en se basant sur l'évaluation du nombre relatif de gènes par niveau de GC (données accessibles par la *figure 4B*) que la carte physique actuelle du génome humain ne recouvre, au mieux, que 40% des gènes (13). Cette situation évolue cependant rapidement grâce aux progrès de la cartographie

génétique et, en particulier, à l'utilisation de micro-satellites des isochores H2 et H3 (14).

Génome humain et évolution

La distribution des gènes trouvée dans le génome humain a été largement conservée au cours de l'évolution. Chez tous les vertébrés, en effet, on retrouve une concentration des gènes au niveau des isochores les plus riches en GC (3). Cette conservation de la distribution des gènes s'accompagne toutefois de changements compositionnels importants. Ainsi, les génomes des vertébrés à sang froid présentent une gamme très étroite de composition en bases, aussi bien de l'ADN que des gènes. Leurs niveaux de GC sont presque toujours plus faibles que ceux des génomes des vertébrés à sang chaud (*voir figure 2*). De plus, leurs chromosomes métaphasiques ne présentent pas de bandes R. Le *pattern* du génome humain, typique des génomes de la plupart des mammifères et semblable à celui des génomes des oiseaux, est caractérisé par des changements de composition (accroissement de GC) qui

sont intervenus entre les reptiles d'une part et les mammifères et les oiseaux d'autre part, et qui n'ont touché que les régions riches en gènes et de façon proportionnelle à cette richesse. Ainsi, sont nées les familles d'isochores H1, H2 et H3. Quant à la plus grande partie du génome, pauvre en gènes, elle n'a pas subi de changement de composition notable. Ce phénomène d'enrichissement en GC a mené à la situation décrite dans le cas du génome humain, où la densité des gènes est en corrélation avec les niveaux de GC₃ des gènes et de GC des isochores. Il est fort possible, et de récentes données vont dans ce sens (15, 16), que la cause des enrichissements en GC du génome des vertébrés à sang chaud soit liée à des avantages sélectifs. □

(12) D Cohen *et al* (1993) *Nature* 366, 698-701.

(13) S Zoubak & G Bernardi, résultats non publiés.

(14) G Gyapay *et al* (1994) *Nature Genetics* 7, 246-249.

(15) S Cacciò *et al* (1995) *J Mol Evol* 40, 280-292.

(16) S Zoubak *et al* (1995) *J Mol Evol* 40, 293-307.