

## LE GÉNOME DES VERTÉBRÉS: ORGANISATION, FONCTION ET ÉVOLUTION

Giorgio BERNARDI\*

Le génome des vertébrés (et des eucaryotes en général) n'est pas simplement formé de gènes situés au hasard au sein d'ADN « déchet », mais il est organisé en un système qui obéit à des règles qui constituent un véritable code génomique. De plus, les gènes ne sont pas distribués d'une manière uniforme. Dans le génome humain, les gènes sont concentrés dans des régions de composition particulière (les plus riches en guanine et cytosine), qui semblent correspondre aux régions télomériques de certains chromosomes (bandes T). L'étude de l'organisation du génome des différentes classes de vertébrés a permis d'aborder de façon nouvelle certains problèmes fondamentaux de l'évolution du génome.

**C**réé en 1920 pour désigner l'ensemble des gènes d'un organisme (1), le mot génome n'a été populaire, pendant longtemps, ni auprès des généticiens, ni auprès des cytogénéticiens. Ces chercheurs n'avaient pas besoin de ce terme, les généticiens s'occu-

pant de gènes individuels et les cytogénéticiens de chromosomes.

En 1948, on découvrit que la quantité d'ADN par cellule était constante pour une espèce donnée. Cette découverte mena à la définition de la quantité d'ADN par cellule

haploïde comme taille du génome. Le mot génome s'est ensuite répandu grâce à des travaux sur l'organisation des génomes eucaryotes, travaux fondés sur les cinétiques de réassociation de fragments d'ADN dénaturé, et sur le fractionnement de l'ADN par sédimentation à l'équilibre en gradient de densité.

Plus récemment, le développement spectaculaire de la génétique moléculaire a conduit à une augmentation exponentielle de nos connaissances des gènes. Mais le dernier développement en date, qui a réellement rendu populaire le mot génome, est le projet Génome humain.

### Le sens du mot génome

La compréhension de ce que le génome représente réellement ne s'est pas pour autant beaucoup améliorée. C'est encore pour beaucoup la somme des gènes, qui seraient, dans le cas des génomes eucaryotes, éparpillés au hasard dans une immensité d'ADN « déchet » formée pour l'essentiel par les séquences intergéniques, qui chez les mammifères représentent au moins 90% du génome.

Des données récentes montrent, toutefois, que le génome est beaucoup plus que cela. C'est un système intégré du point de vue de sa structure, de ses fonctions et de son évolution, un système qui obéit à des règles précises que l'on pourrait baptiser code génomique. Ce concept s'est développé à partir de recherches fondées sur les propriétés compositionnelles des génomes (2).

### Les caractéristiques compositionnelles des génomes

Les préparations d'ADN de vertébrés à sang chaud (mammifères, oiseaux) sont constituées par des fragments qui peuvent être séparés (par centrifugation à l'équilibre en gradient de densité, en présence de ligands séquence spécifiques) en familles caractérisées par leurs niveaux de GC, c'est-à-dire par leur compo-

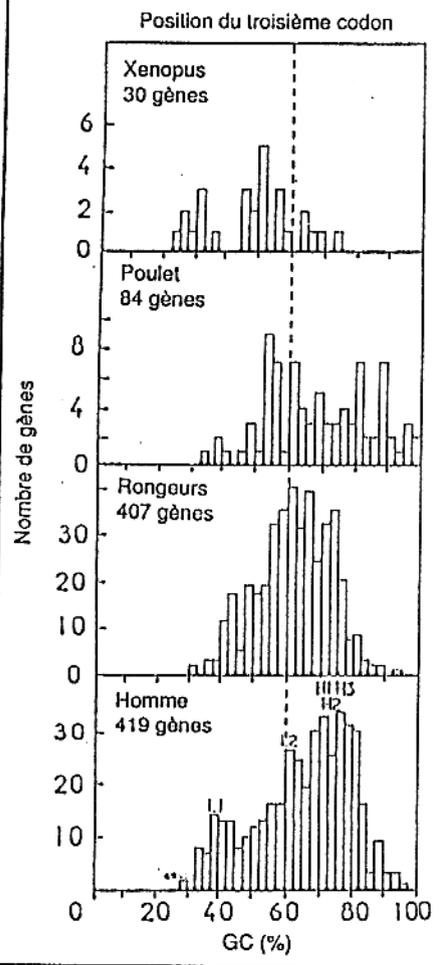
\* Laboratoire de génétique moléculaire, institut Jacques-Monod, Paris.

(1) H. Winkler (1920) Vererbung und Ursache der Parthenogenese in Pflanzen- und Tierreich, Iena, Fischer.

(2) G. Bernardi *et al.* (1985) The mosaic genome of warm-blooded vertebrates *Science* 228, 953-958.

**FIGURE 2 – DISTRIBUTION COMPOSITIONNELLE DES TROISIÈMES POSITIONS DES CODONS DE GÈNES DES VERTÉBRÉS**  
(d'après Bernardi [1989] [12])

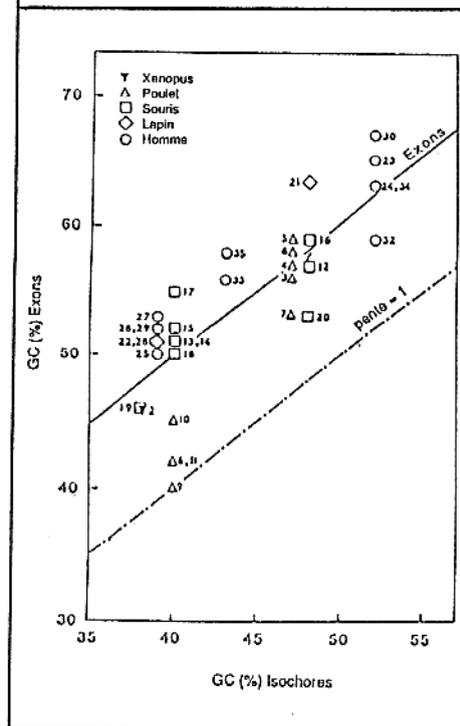
Chez les vertébrés à sang chaud, les séquences codantes riches en GC sont les plus abondantes.



concentrées dans les isochores pauvres en GC, tandis que le million de séquences répétées courtes (Sines), riches en GC, sont localisées principalement dans les isochores riches en GC. La réinsertion est donc ciblée vers des régions génomiques de composition similaire, et/ou l'intégration est plus stable dans ces régions. Un comportement analogue a été observé pour les séquences intégrées des virus de la leucémie bovine (BLV), de l'hépatite B humaine (HBV), de la tumeur mammaire de la souris (MMTV) et du sarcome de Rous (RSV). Les isochores qui abritent ces séquences sont riches en GC pour le BLV, le HBV et le RSV, dont les génomes sont riches en GC, et pauvres en GC pour le MMTV, dont le génome est pauvre en GC.

2 – Les bandes R et les frontières entre bandes R et G sont les endroits préférentiels des translocations spontanées (voir figure 4), des échanges induits de chromatides-sœurs, des anomalies chromosomiques détectées après action des rayons X et de produits chimiques, des points chauds pour la formation des chiasmats mitotiques et des sites fragiles. Ces observations soulignent le rôle joué dans ces

**FIGURE 3 – CORRÉLATION ENTRE LES NIVEAUX DE GC DES EXONS ET LES NIVEAUX DE GC DES ISOCHORES DANS LESQUELLES ILS ONT ÉTÉ LOCALISÉS**  
(d'après Bernardi [1989] [12])



phénomènes par les discontinuités de composition aux frontières G/R et dans les bandes R, ainsi que le rôle de la distribution génomique des séquences recombinogènes, telles que les Sines, par exemple.

3 – L'utilisation des codons et des acides aminés est très différente selon que les gènes correspondants sont pauvres ou riches en GC. La composition en acides aminés des protéines est ainsi corrélée avec la localisation des gènes correspondants dans les isochores. On peut alors conclure que la distribution des gènes dans les isochores a une contrepartie fonctionnelle.

4 – Les doublets CpG (9) sont les seuls sites où la méthylation est possible dans les génomes des vertébrés. Les « îlots de CpG » (10) sont des séquences habituellement localisées près des extrémités 5' des gènes ; ils sont caractérisés par des niveaux élevés de GC, par des regroupements de doublets CpG non méthylés et par des « boîtes G/C » qui correspondent à des promoteurs riches en GC. La fréquence des doublets CpG et des « îlots de CpG » augmente jusqu'à atteindre de hauts niveaux en allant des isochores pauvres en GC aux isochores riches en GC. Ainsi la méthylation pourrait être plus élevée dans les isochores riches en GC, et les promoteurs riches en GC seraient plus fréquemment associés aux gènes riches en GC.

5 – Les bandes G et les isochores pauvres en GC se répliquent tard et se condensent tôt dans le cycle cellulaire, à l'inverse des bandes R et des isochores riches en GC. Ces phénomènes semblent toutefois plutôt liés à l'organisation de base en chromomères-interchromomères (qui correspondent aux bandes G et R respectivement) qu'aux différences de composition.

(9) Doublets cytosine-phosphate guanine, dans lesquels la cytosine se trouve en 5' par rapport à la guanine.

(10) A. Bird (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321, 209-213.

**TABEAU – LE GÉNOME HUMAIN (\*)**

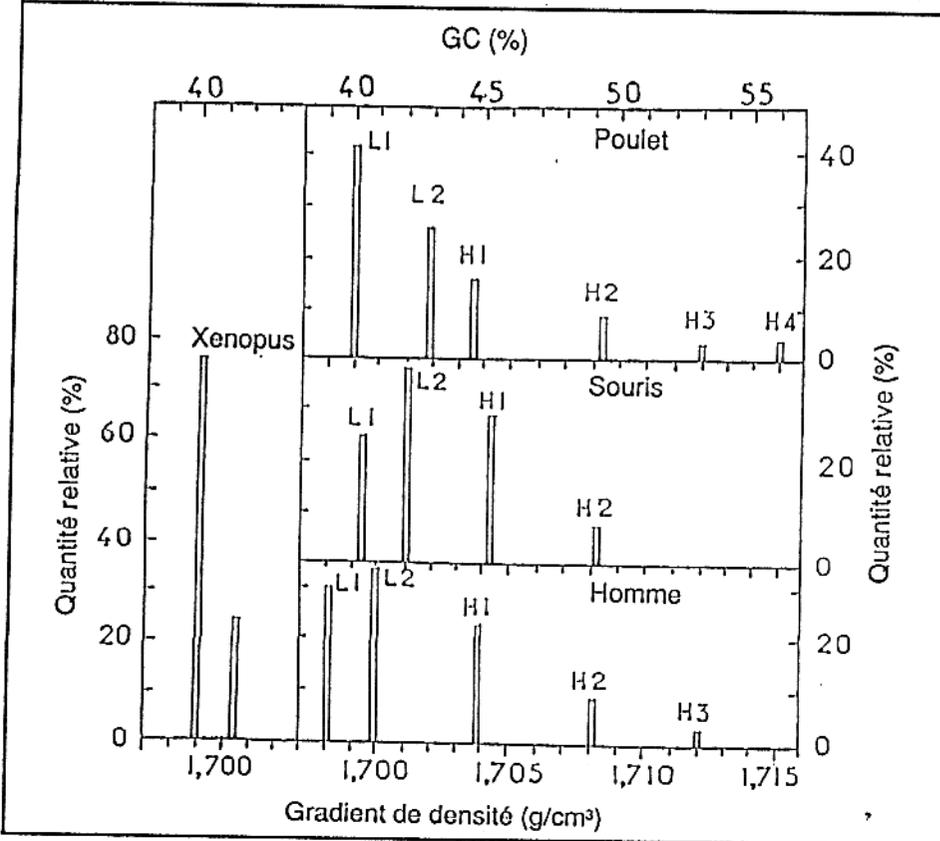
Le paléogénome : isochores pauvres en GC, bandes G	Le néogénome : isochores riches en GC, bandes R
Pénurie des gènes, gènes pauvres en GC, pénurie d'« îlots de CpG », promoteurs à TATA box, résistance à la DNase I, recombinaison peu fréquente	Abondance de gènes (H3), gènes riches en GC, abondance d'« îlots » de CpG, promoteurs à G/C box, sensibilité à la DNase I, recombinaison fréquente
Triples hélices riches en AT, digestion par HaellI	DNA Z, digestion par DNase après chromomycine A3, abondance de Sines, hétérogénéité de composition
Abondance de Llines, homogénéité de composition	Interchromomères, réplication précoce, condensation tardive
Chromomères, réplication tardive, condensation précoce	

(\*) D'après la référence 12 : voir cette référence pour des renseignements complémentaires.

# GÉNÉTIQUE FONDAMENTALE

FIGURE 1 – DISTRIBUTION COMPOSITIONNELLE DES ISOCHORES DES GÉNOMES DES VERTÉBRÉS (d'après Bernardi [1989] [12])

L'ADN des différentes espèces étudiées a été fractionné par centrifugation à l'équilibre en gradient de densité en présence de ligands séquence-spécifiques. Dans le génome humain, et dans celui des vertébrés à sang chaud en général, les isochores pauvres en GC sont les plus abondantes.



sition en bases guanine et cytosine. Ces fragments dérivent de segments beaucoup plus grands (> 300 kilobases), qui sont caractérisés par une remarquable homogénéité de composition (3) et qui, pour cette raison, furent appelés isochores (ce qui signifie régions identiques). Chez l'homme, les familles d'isochores pauvres en GC représentent environ deux tiers du génome et possèdent des niveaux de GC très voisins. Les familles d'isochores riches en GC correspondent au tiers restant, avec des niveaux de GC très divers.

Les distributions compositionnelles des isochores et des séquences codantes (qui représentent les deux caractéristiques compositionnelles les plus importantes du génome) sont montrées dans les figures 1 et 2. Ces distributions sont très différentes, la première étant caractérisée par une abondance de segments pauvres en GC, la seconde par une abondance de séquences riches en GC. On peut considérer que les caractéristiques compositionnelles définissent un phénotype du génome.

Le fractionnement de l'ADN associé à l'hybridation de sondes spécifiques de gènes donnés a révélé que les gènes ne sont pas distribués d'une manière uniforme dans le génome des vertébrés à sang chaud, mais qu'ils sont concentrés dans les isochores les plus riches en GC. De plus, cette méthodologie a permis d'établir que des corrélations compo-

sitionnelles existent entre les gènes et les isochores qui les contiennent (2). Ces corrélations (figure 3) ainsi que les corrélations compositionnelles entre les trois positions des codons (4) constituent le code génomique qui résulte de contraintes compositionnelles dues à la sélection ou à la pression de mutation.

## Isochores, bandes chromosomiques et cartographie compositionnelle

Les isochores sont reliées aux bandes chromosomiques G (Giemsa-positives) et R (bandes réverses, équivalentes aux bandes Giemsa négatives) [5] : les isochores pauvres en GC correspondent à l'ADN des bandes G et les isochores riches en GC (riches en gènes) aux bandes R (voir tableau).

Les relations entre isochores et bandes chromosomiques peuvent être étudiées plus en détail par une nouvelle méthode, la cartographie compositionnelle. Chaque fois que des cartes physiques sont disponibles, on peut déterminer les niveaux de GC autour de séquences de référence (localisées sur la carte physique) qui peuvent être repérées par hybridation avec des sondes appropriées. La cartographie compositionnelle peut définir la composition en bases de segments d'ADN environ deux fois plus grands que les fragments

utilisés dans le fractionnement.

La cartographie compositionnelle a déjà été appliquée (6) à l'analyse du bras long du chromosome 21 (figure 4) et au gène de la dystrophine (7). Elle a apporté la démonstration directe de l'homogénéité compositionnelle des bandes G et de l'hétérogénéité des bandes R. Ces dernières (qui correspondent aux familles d'isochores riches en GC et recouvrent une large gamme en composition) contiennent de nombreuses bandes G fines, seulement visibles à haute résolution. Les isochores les plus riches en GC sont localisées dans la région subtélomérique du bras long. Comme les télomères correspondent presque toujours à des bandes R et que les régions terminales d'une moitié environ d'entre eux sont les régions des chromosomes humains les plus résistantes à la dénaturation (8), on peut penser que les isochores les plus riches en GC correspondent à ces régions télomériques (6), comme c'est le cas pour le bras long du chromosome 21.

Il va sans dire que ces régions, qui contiennent une proportion importante de gènes, devraient constituer les premières cibles des projets de cartographie et séquençage.

## Isochores et fonctions génomiques

On connaît encore mal la signification fonctionnelle des isochores. Quelques points sont toutefois d'ores et déjà définis.

1 – L'intégration des séquences mobiles et virales a lieu surtout dans des isochores de composition voisine. Par exemple, dans le génome humain, les 100 000 séquences répétées longues (ou Lines), pauvres en GC, sont

(3) G. Macaya *et al.* (1976) An approach to the organization of eukaryotic genomes at a macromolecular level. *J. Mol. Biol.* 108, 237-254.

(4) G. Bernardi & G. Bernardi (1986) Compositional constraints and genome evolution. *J. Mol. Evol.* 24, 1-11.

(5) Ces bandes sont obtenues par le traitement des chromosomes métaphasiques avec des colorants fluorescents, des protéases et des agents dénaturants.

(6) K. Gardiner *et al.* (1990) A compositional map of human chromosome 21. *EMBO J.* 9, 1853-1858.

(7) T. Bettecken *et al.* (1990) A compositional map of the human dystrophin gene (en préparation).

(8) B. Dutrillaux (1973) Nouveau système de marquage chromosomique : les bandes T. *Chromosoma* 41, 395-402.



**Principes d'amélioration génétique des animaux domestiques**

F. Minvielle  
Coéd. INRA-Presses de l'Université Laval

Ce manuel vise tout d'abord à fournir aux étudiants des productions et sciences animales un exposé des principes de l'amélioration génétique qui leur donne les clés théoriques nécessaires à la compréhension de l'amélioration des animaux domestiques. Il offre aussi à celui qui entreprend des études avancées dans le domaine, une connaissance plus approfondie de la matière en montrant relativement en détail le cheminement qui conduit aux résultats principaux. Chacun des 10 chapitres comprend des exemples et des exercices adaptés et une bibliographie de base. 1990, 232 p., 210 F

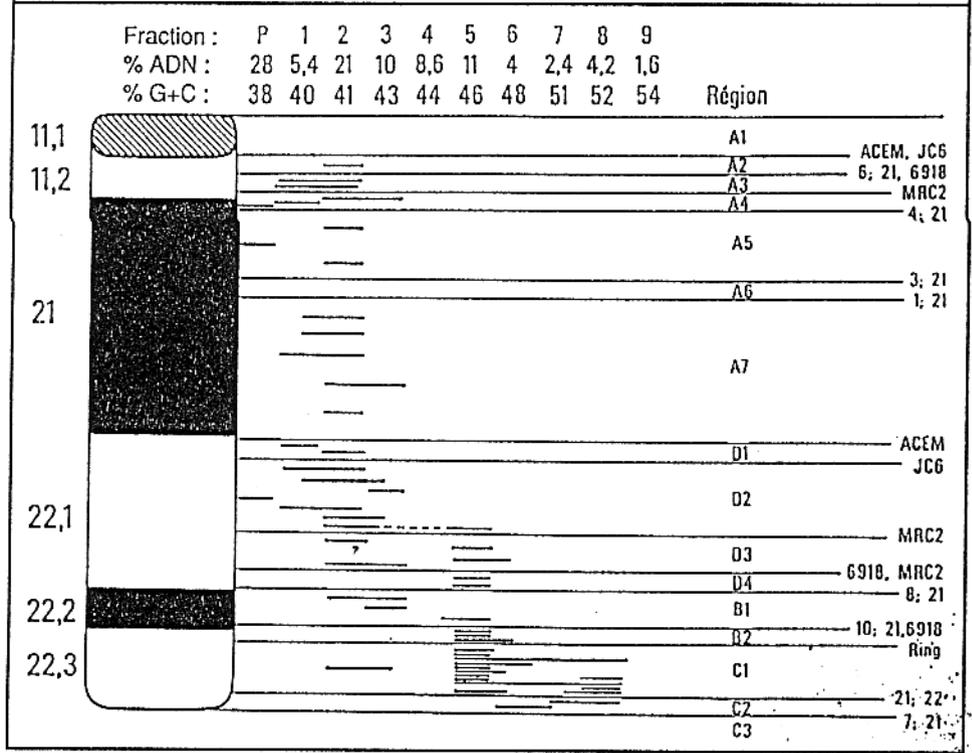


**Cytogénétique des mammifères d'élevage**

P. Popescu  
L'auteur présente, dans cet ouvrage, le caryotype normal et les anomalies chromosomiques connues à ce jour pour chacune des principales espèces domestiques : bœuf, mouton, chèvre, cheval, porc et lapin. Une part importante revient à la pathologie chromosomique et à ses conséquences sur l'élevage. 1989, 114 p., 110 F

**FIGURE 4 – CARTE COMPOSITIONNELLE DU BRAS LONG DU CHROMOSOME 21 (d'après Gardiner *et al.* [1990] [6])**

Les longues lignes horizontales indiquent les positions des points de cassure qui définissent des segments chromosomiques. Les barres horizontales indiquent les fractions d'ADN dans lesquelles s'hybrident des segments chromosomiques donnés. On obtient ainsi des renseignements sur la composition de segments d'ADN de 0,2 à 0,3 Mb (mégabases).



**Isochores et évolution du génome**

Les caractéristiques compositionnelles (ou phénotypes) des génomes des vertébrés permettent de définir deux modes d'évolution (11). Dans le mode conservatif, répandu chez les vertébrés à sang chaud, les mutations s'accumulent en induisant peu de changements de composition du génome. Dans le mode transitionnel, de grands changements de composition ont lieu, et parfois dans des intervalles de temps très courts.

Deux transitions compositionnelles majeures ont eu lieu au cours de l'évolution du génome des vertébrés : celle entre les reptiles et les mammifères, et celle entre les reptiles et les oiseaux. Ces transitions ont conduit à la formation du néogénome (12) [voir tableau]. Dans ce compartiment du génome des vertébrés à sang chaud, des isochores ancestrales pauvres en GC qui correspondent aux interchromomères ont été changées en isochores riches en GC, qui contiennent de nombreux gènes riches en GC, très fréquemment associés avec des « îlots de CpG ». L'autre com-

partiment du génome des vertébrés à sang chaud, le paléogénome (12) est caractérisé par sa similitude avec ce qu'il était, et est toujours, chez les vertébrés à sang froid. Les différences de composition liées aux transitions ont pris naissance grâce à une fixation directionnelle de mutations ponctuelles, ainsi que le démontre la comparaison des séquences codantes homologues de vertébrés à sang froid et à sang chaud (11).

L'étude approfondie de la composition des génomes, et en particulier celle des isochores des régions génomiques les plus riches en GC, devrait largement contribuer à la compréhension de l'organisation du génome humain. Les connaissances acquises ainsi pourraient s'appliquer non seulement au génome humain ou à celui des vertébrés à sang chaud, mais encore à celui d'autres génomes eucaryotes.

(11) G. Bernardi *et al.* (1988) Compositional patterns in vertebrate genomes : conservation and change in evolution. *J. Mol. Evol.* 28, 7-18.

(12) G. Bernardi (1989) The isochore organization of the human genome. *Ann. Rev. Genet.* 23, 637-661.