

## L'impossible « fossile génétique »

### L'ADN ancien...

La publication en 1997, d'une séquence de l'ADN mitochondrial de *Homo neanderthaliensis* est à la fois une extraordinaire prouesse technique et une avancée scientifique fascinante [1-3]. Elle souligne la puissance de l'approche moléculaire dans le domaine de la paléanthropologie. Cette approche a malheureusement ses limites techniques, qui sont fixées par la disponibilité du matériel ancien, rare et précieux, la faible quantité et la médiocre stabilité de l'ADN extrait, et par la performance des techniques actuelles. L'évolution technologique pourra certainement un jour relever le défi posé par ces limites. Elle est aussi limitée par une indispensable prudence dans les interprétations [4].

### ... et l'ADN très ancien

Avec les méthodes disponibles, l'analyse moléculaire n'est possible que sur des échantillons qui remontent à moins de 50-100 000 ans [2, 3]. Il n'est pas envisageable d'étudier des fossiles d'espèces disparues au-delà de cette limite. Cela, combiné à l'absence de fossiles pour de nombreuses espèces, nous prive d'une certaine richesse de données utilisables au niveau moléculaire pour une analyse phylogénétique exhaustive. Ainsi, à défaut de pouvoir directement analyser l'ADN de fossiles très anciens ou inexistant, il faut continuer à se fonder sur des études moléculaires d'espèces éteintes récemment, comme le mammoth, ou d'espèces représentatives de phylums-clés ou très étudiées en biologie, comme le céphalocordé amphioxus, la mouche drosophile ou le nématode *C. elegans*.

Par ces études, peut-on pousser l'analyse jusqu'à déduire l'organisation d'un gène, voire d'un génome ances-

tral ou, tout du moins, d'une partie d'un génome ancestral? Des études pilotes ont été réalisées. Ainsi, des études phylogénétiques ont permis de déduire la séquence du promoteur ancestral des séquences de type Line-1. Ce promoteur s'est révélé actif après reconstitution synthétique [5]. Un autre exemple de reconstitution d'une séquence ancestrale a été réalisé à partir de l'étude phylogénétique des isocitrate déshydrogénases [6]. Mais ces études sont limitées à des gènes. Est-il maintenant envisageable de diriger les recherches vers la reconstitution de régions entières, c'est-à-dire de segments chromosomiques de plusieurs mégabases?

En analysant de façon poussée l'organisation du génome d'espèces actuelles, il est possible de reconnaître une similitude entre plusieurs espèces au niveau de certaines régions. Cela peut être considéré comme un reste de l'organisation primordiale, un stigmate du génome ancestral, un « fossile génétique ». Prenons quelques exemples. Il existe chez l'homme des régions dites de paralogie [7-9]. Ces régions seraient issues de la duplication en série d'une même région. Le cas des gènes de la famille *Hox*, groupés au niveau de quatre régions chromosomiques chez les vertébrés est sans doute le meilleur exemple de duplication en masse [10-12]. Il en existe d'autres. Nous avons ainsi observé que, chez l'homme, les régions chromosomiques suivantes: 4p16, 5q33-35, 8p11-21, et 10q24-26 comportaient une série de gènes qui appartiennent à des familles multigéniques (*figure 1*). On retrouve, par exemple, un gène codant pour un récepteur de facteurs de croissance des fibroblastes dans chacune de ces régions, ainsi que des gènes codant pour des récepteurs adrénérgiques et des transporteurs. Chez les espèces d'invertébrés qui ont été analysées

(*D. melanogaster*, *C. elegans*), on ne trouve qu'une seule région – que l'on peut dire orthologue – mais les gènes sont les mêmes (*figure 1*). Il semble donc qu'il ait existé une région originelle portant les gènes ancêtres de ces séries de gènes [13]. Cette région est restée unique chez certaines espèces, a dupliqué chez d'autres. Pour ces dernières, l'augmentation du nombre de régions pourrait être le résultat d'une tétraploïdisation du génome [14].

L'étude de ces familles de gènes chez plusieurs espèces permet, d'une part, d'émettre des hypothèses sur la date à laquelle ont eu lieu les événements de duplication. D'autre part, il est possible de déduire de façon assez grossière une ébauche de région ancestrale de ces études [9, 13]. On dispose désormais de beaucoup d'informations sur les séquences de plusieurs génomes, ainsi que de méthodes et d'ordinateurs puissants pour les analyser. Si l'on analysait systématiquement le génome de plusieurs espèces à la recherche de régions de paralogie (issues d'événements de duplication) et d'orthologie (issues d'événements de spéciation), peut-être pourrions nous déduire un assez grand nombre de régions originelles, et pourquoi pas, reconstituer une partie d'un génome ancestral virtuel, en particulier celui du premier animal triploblastique, à l'origine des protostomes et deutérostomes.

Ce type d'étude nécessite une combinaison d'analyses phylogénétiques, de comparaisons de séquences et de localisations chromosomiques chez plusieurs espèces qui, seule, permet d'établir à la fois un schéma séquentiel et une datation plausibles des duplications [15, 16]. Nous avons ainsi tenté de déduire la liaison d'une série de gènes au niveau d'une région ancêtre hypothétique [13]. Il serait intéressant, dès maintenant, de construire une carte du génome

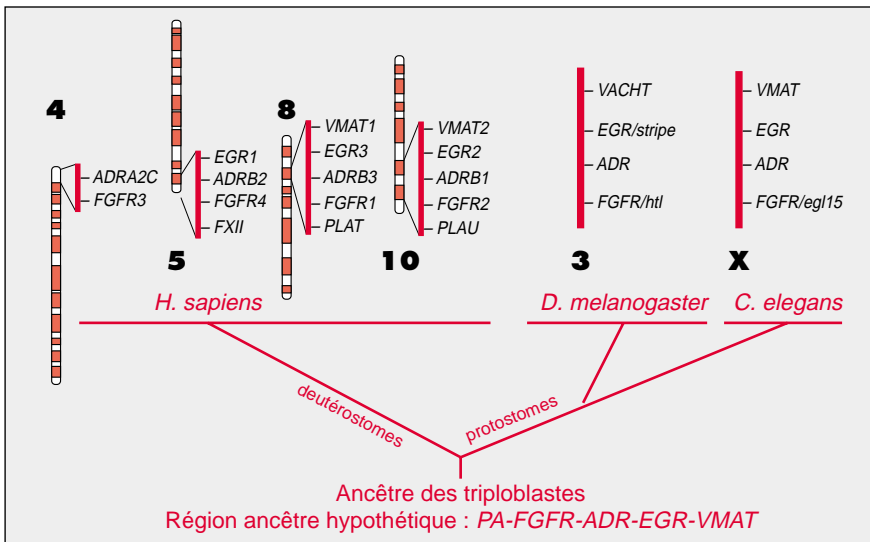


Figure 1. **Reconstitution schématique et hypothétique de l'organisation et de l'évolution d'un segment du génome de l'ancêtre du triploblaste.** Un arbre phylogénétique très simplifié montre trois espèces. Le branchement commun entre *D. melanogaster* et *C. elegans* est indiqué d'après [17]. Pour chaque espèce est représentée une région chromosomique comportant une série de gènes. Chez l'homme on trouve des paralogues de certains de ces gènes (codant pour des récepteurs des FGF, FGFR, des récepteurs adrénergiques, ADR, des activateurs du plasminogène, PLAT et PLAU et un facteur de coagulation apparenté, FXII, des protéines early growth response, EGR, ou des transporteurs, VMAT) au niveau de plusieurs régions. Chez la mouche *D. melanogaster* et le nématode *C. elegans*, on trouve un gène de chaque famille (VACHT est apparenté à VMAT; chez la drosophile, non loin de FGFR/heartless, se trouve également un deuxième gène FGFR, *breathless*). La région ancestrale hypothétique présente chez l'ancêtre des triploblastes, contenant les ancêtres des gènes évoqués plus haut chez les espèces actuelles, est indiquée à la base de l'arbre (PA : activateur du plasminogène). L'ordre et la distance entre les gènes sont arbitraires. Des informations supplémentaires peuvent être obtenues en consultant [13].

humain en indiquant, d'une part, les régions de paralogie et, d'autre part, les régions de synténie conservées dans le génome d'autres espèces, non seulement la souris comme cela est désormais bien établi [15], mais aussi des non-vertébrés, comme la mouche drosophile et le nématode *C. elegans*.

Une fois établie une hypothétique carte génétique ancestrale, il serait alors intéressant de se demander si l'organisation des gènes est aléatoire ou fondée sur des impératifs fonctionnels. En d'autres termes, si le groupement de certains gènes est le fruit du hasard ou d'une évolution commune associée à une fonction coordonnée dans un processus particulier ■

#### François Coulier

Chargé de recherches à l'Inserm.

#### Marie-Josèphe Pébusque

Chargée de recherches à l'Inserm.

#### Pierre Pontarotti

Chargé de recherches au Cnrs.

#### Daniel Birnbaum

Directeur de recherches à l'Inserm.

Inserm U. 119, Institut Paoli-Calmettes, 27, boulevard Leï-Roure, 13009 Marseille, France.

#### TIRÉS À PART

D. Birnbaum.

#### RÉFÉRENCES

- Krings M, Stone A, Schmitz R, Krainitz H, Stoneking M, Pääbo S. Neandertal DNA sequences and the origin of modern humans. *Cell* 1997; 90: 19-30.
- Lindahl T. Facts and artifacts of ancient DNA. *Cell* 1997; 90: 1-3.
- Audic S, Béraud-Colomb E. Ancient DNA is thirteen years old. *Nat Biotechnol* 1997; 15: 855-8.
- Denamur E, Lecomte G. L'ADN de l'homme de Neanderthal. *Med Sci* 1997; 12: 1488-90.
- Adey N, Tollefsbol T, Sparks A, Edgell MH, Hutchinson III C. Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc Natl Acad Sci USA* 1994; 91: 1569-73.
- Dean A, Golding GB. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc Natl Acad Sci USA* 1997; 94: 3104-9.
- Lundin L. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 1993; 16: 1-19.
- Spring J. Vertebrate evolution by interspecific hybridisation. Are we polyploid? *FEBS Lett* 1997; 400: 2-8.
- Katsanis N, Fitzgibbon J, Fisher MC. Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel *PBX* and *NOTCH* loci. *Genomics* 1996; 35: 101-8.
- Kappen C, Ruddle FH. Evolution of a regulatory gene family: HOM/HOX genes. *Curr Opin Genet Dev* 1993; 3: 931-8.
- Sidow A. Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev* 1996; 6: 715-22.
- Holland PW, Garcia-Fernandez J. Hox genes and chordate evolution. *Dev Biol* 1996; 173: 382-95.
- Pébusque MJ, Coulier F, Birnbaum D, Pontarotti P. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. *Mol Biol Evol* 1998; 15: 1145-59.
- Ohno S. *Evolution by gene duplication*. Berlin/Heidelberg/New York: Springer Verlag, 1970.
- Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. Gene duplications and the origin of vertebrate development. In: Akam M, Holland P, Ingham P, Wray G, eds. *The evolution of developmental mechanisms*. The Company of Biologists, Cambridge, Development; 1994: 125-33.
- DeBry R, Seldin M. Human/mouse homology relationships. *Genomics* 1996; 33: 337-51.
- Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 1997; 387: 489-93.