

Evolution des génomes de métazoaires.

Metazoan genome evolution.

Etienne DANCHIN* Laurent ABI-RACHED+ André GILLES° Pierre PONTAROTTI*

* Laboratoire de Phylogénomique, INSERM U119, 27 bd. Leï Roure 13009 Marseille, France.

+Department of Structural Biology 299, Campus Dr. Fairchild Building Stanford University, CA 94305, USA.

° UPRES Biodiversité 2202 Université de Provence, Pl. V. HUGO 13331 Marseille cedex 3, France.

e-mail : danchin@marseille.inserm.fr, pontarotti@marseille.inserm.fr

Résumé

Notre groupe étudie l'histoire évolutive des génomes dans le but de comprendre comment les génomes actuels ont pu devenir ce qu'ils sont à partir d'un génome ancestral. Pour cela nous nous intéressons plus particulièrement à deux aspects : 1–les duplications dans les génomes et leur incidence sur l'évolution. 2–la conservation de l'organisation entre différents génomes de métazoaires. Nos travaux ont permis de démontrer pour la première fois d'une manière rigoureuse l'existence (pour une région), de duplications en blocs dans le génome des vertébrés [1]. Nous avons aussi prouvé l'existence de micro synténies entre les vertébrés et les céphalocordés pour cette même région, et ainsi pu inférer l'organisation de cette région génomique chez l'ancêtre des vertébrés. En comparant cette région avec le génome de la drosophile nous avons mis en évidence une synténie conservée vraisemblablement ancestrale entre l'homme et la drosophile. De même nous avons pu inférer l'organisation de la région ancestrale de tous les bilatériens.

Mots-clés : *évolution, synténie, phylogénie, duplications, orthologie, paralogie.*

Abstract

Our group studies the evolutionary History of genomes. Our goal is to understand how today genomes have evolved from a common ancestral genome and have become what they are. We are particularly interested in two aspects : 1–genomes duplications and their role on evolution. 2–conservation of genomes organization between metazoan. Our work allowed us to evidence for the first time “en–bloc” duplications for a genomic region in vertebrates genome[1]. We also showed the existence of conserved microsyntenies between vertebrates and cephalochordates genomes for this region. This work also allowed us to infer the genomic organization of this region in the vertebrates ancestor. Comparisons of this region with the drosophila genome led us to unveil a conserved synteny (likely ancestral) between this arthropod and the vertebrates. Similarly the inference of the genomic organization of this region has been made possible for the putative ancestor of all bilaterians : Urbilateria [2].

Keywords : *evolution, synteny, phylogeny duplications, orthology, paralogy.*

Introduction

Comprendre comment un organisme aussi simple qu'une bactérie, apparu sur terre il y a environ 3,5 milliards d'années a pu évoluer pour donner naissance à toute une variété d'espèces dont l'homme, reste encore aujourd'hui un des plus grands mystères de la biologie.

Les tous derniers travaux de notre équipe en collaboration avec des chercheurs japonais révèlent une des clés de ce phénomène : le stock de chromosomes des espèces vivantes actuelles les plus complexes, dont l'homme, résulterait du doublement de la quantité d'information génétique présente chez leurs ancêtres. Les résultats de cette étude [1] constituent pour l'heure un premier élément confirmant que certaines parties du génome humain sont issues d'un processus de duplication en bloc du matériel génétique de nos ancêtres.

L'analyse des régions dupliquées mises en évidence dans ces travaux a permis de déduire l'organisation putative de la région génomique correspondante chez l'ancêtre commun de tous les euchordés. Nous avons pu mettre en évidence lors de la comparaison de ces mêmes régions avec le génome de la drosophile, l'existence d'une synténie conservée entre l'homme et la drosophile. Il est vraisemblable que cette synténie conservée ait un caractère ancestral. Grâce aux résultats de cette étude comparative nous pouvons remonter plus loin vers la reconstruction du génome ancestral, et proposer pour cette région une organisation génomique chez l'ancêtre commun de tous les bilatériens: “Urbilateria”[2].

1–La duplication, élément moteur de l'évolution ?

Il y a trente ans, l'hypothèse fut formulée que les grands sauts de l'évolution tels que le passage des invertébrés aux vertébrés étaient la conséquence de duplications de segments entiers de génomes, ces événements aboutissant à l'apparition de milliers de nouvelles copies de gènes capables d'engendrer de nouvelles fonctions spécifiques au monde des vertébrés. Mais, malgré les avancées récentes en génétique qui ont permis le décryptage de génomes de nombreuses espèces, aucune démonstration rigoureuse n'était venue étayer cette hypothèse.

Au départ, la supposition suivante a été avancée : si des multiplications (duplications) de séquences du génome ou polyploïdisations ont eu lieu chez nos ancêtres pré-vertébrés, on doit retrouver la trace de ces régions chromosomiques homologues dans le génome de vertébrés et plus particulièrement chez l'être humain (vertébré à mâchoires).

Notre équipe s'est intéressée à une catégorie de gènes humains particulier : celle des gènes du complexe majeur d'histocompatibilité (CMH), impliqués dans les réponses immunitaires. Par ailleurs, avec le soutien de nos collaborateurs nous avons caractérisé la région équivalente chez l'*Amphioxus*, un invertébré de forme élançée modèle pour la compréhension de l'histoire évolutive des vertébrés : c'est en effet le plus proche cousin de tous les vertébrés.

Or, des études antérieures avaient mises en évidence l'existence, à proximité des gènes du CMH humain, de gènes dits "gènes-ancres". Ces gènes ont connu une évolution relativement plus lente et sont pour cela repérés au sein de différentes espèces. Quatre exemplaires de régions homologues à la région comportant ces gènes-ancres ont été identifiés chez l'homme. Chez l'*Amphioxus*, une seule région de ce type est présente.

Ce sont des analyses comparées de segments homologues du CMH et de leur région équivalente chez l'*Amphioxus*, qui nous ont amenés à montrer que seules des duplications en bloc de ces régions pouvaient expliquer la conservation observée au sein des deux génomes. Fruit de la collaboration avec l'équipe japonaise, le séquençage des portions d'intérêt du génome d'*Amphioxus* (représentant un segment de 400 kilobases d'ADN) a contribué à l'avancement de notre démarche scientifique. Puis la confrontation de bases de données génomiques des régions du CMH, gène par gène, avec les régions homologues d'*Amphioxus*, nous a permis de démontrer que les quatre régions homologues retrouvées chez l'homme sont les stigmates de deux duplications successives du génome survenues chez les ancêtres des vertébrés à mâchoires.

Par ces travaux, il nous a été possible d'établir que les quatre régions homologues du CMH identifiées, résultent bien de la duplication d'une région ancestrale. Les datations effectuées nous permettent en outre d'affirmer que cette duplication du matériel génétique est survenue en une seule fois, au cours de la même période, c'est-à-dire après la séparation entre le groupe des céphalocordés, auquel appartient l'*Amphioxus*, et celui des vertébrés, mais avant l'apparition de l'embranchement des vertébrés à mâchoires, soit entre 766 et 528 millions d'années.

L'analyse d'autres segments homologues est en cours. Elle pourrait venir conforter ces premiers résultats en apportant la preuve que d'autres parties de notre génome actuel résultent d'épisodes de polyploïdisation.

2–L'organisation des génomes est-elle conservée au cours de l'évolution ?

L'identification de synténies conservées entre les génomes de deux espèces peut révéler leur organisation ancestrale. Au plus les événements de spéciation sont anciens, au moins les informations restantes sont disponibles, et au plus la difficulté de réaliser des comparaisons est grande. L'évolution implique dans les génomes des changements dans la séquence génomique, dans le nombre de gènes, ainsi que dans l'ordre des gènes sur les chromosomes. Les gènes sont redistribués suite à des réarrangements chromosomiques tels que des mutations, des délétions, des inversions, des insertions, des translocations, et des transpositions. Jusqu'à aujourd'hui, des synténies conservées n'ont été trouvées que dans des phylum ayant divergé relativement récemment. A l'exception du cluster Hox qui est conservé chez tous les Bilatériens, la plus vieille synténie conservée décrite jusqu'alors est celle que nous avons identifiée entre les vertébrés et les céphalochordés (voir ci-dessus). L'existence de synténies conservées entre les protostomiens et les deutérostomiens a été clamée mais cela n'a jamais été soutenu par des analyses statistiques ni par des phylogénies, et donc cette hypothèse reste incertaine. Dans ce travail nous avons testé l'existence de synténies conservées entre les deux phylum les plus distants chez les métazoaires: les protostomiens (*Drosophila*) et les deutérostomiens (homme et *Amphioxus*). Cela nous a permis de décrire une synténie ancestrale qui était vraisemblablement présente chez Urbilateria, et nous fournit un « premier pas » vers la reconstruction du génome d'Urbilateria.

3–Matériel et méthodes.

3.1 Annotation des cosmides d'amphioxus:

La détection des gènes dans les cosmides d'amphioxus a été réalisée par trois logiciels de prédiction utilisant des méthodes différentes (Genscan, Fgenes, et GrailEXP) les données provenant de ces logiciels ont été couplées à une analyse directe par recherche de similarité (blastx) contre des banques de protéines connues.

3.2 Détection d'orthologues :

Dans le but d'identifier les orthologues entre les différentes espèces étudiées (Homme, Drosophile, Amphioxus), nous avons dans un premier temps recherché les séquences similaires entre ces espèces par blast. Afin d'extraire les relations d'orthologie et de paralogie de manière fiable entre les gènes nous avons réalisé une étude phylogénétique pour chaque gène dans un deuxième temps.

3.3 Étude phylogénétique :

Pour chaque gène étudié nous avons recherché le maximum de gènes similaires dans la plus grande variété d'espèces possibles par blast (tblastn) contre différentes banques génomiques. Les séquences significativement similaires ont ensuite été alignées avec le programme ClustalX (utilisant l'algorithme de ClustalW). Les alignements obtenus ont ensuite été retravaillés afin d'éliminer les biais d'alignement éventuels. Les reconstructions phylogénétiques ont été réalisées à partir de ces alignements par l'utilisation de trois méthodes distinctes (Neighbor Joining, Maximum Parsimony, et Maximum Likelihood). Les topologies données par les trois méthodes ont ensuite été comparées puis validées en cas de congruence.

3.4 Tests statistiques :

Afin d'évaluer la signification des résultats, la distribution des orthologues entre les espèces sur leurs génomes respectifs a été analysée. Chaque distribution a été comparée à une distribution aléatoire, et le degré de différence avec une distribution aléatoire nous a permis de valider ou de réfuter l'hypothèse d'une synténie conservée entre ces différentes espèces pour les régions étudiées.

Conclusion.

Les résultats de ces études révèlent l'importance de la phylogénie (étude de l'histoire évolutive des espèces, lignées et des groupes d'organismes) et de la génomique comparative, pour retracer l'histoire évolutive de notre génome. En outre ces travaux contribuent également à une meilleure compréhension de l'organisation actuelle de nos génomes. Les avancées en terme de puissance de calcul et de performances des outils bioinformatiques, ainsi que la disponibilité de nouvelles données, nous ont permis pour la première fois de tester des hypothèses qui avaient été avancées depuis des décennies de manière rigoureuse. De telles hypothèses ne pouvaient être testées que par des études à la fois très fine et à grande échelle ce qui était impossible jusqu'alors.

Références

- [1] Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. & Inoko, H. Evidence for en bloc duplication in vertebrate genomes. *Nature Genet.* In press.
- [2] DeRobertis, E.M. & Sasai, Y. A common plan for dorsoventral patterning in Bilateria. *Nature.* **380**, 37–40 (1996).