

Evolution des structures répétées dans les génomes

<http://www.lirm.fr/~rivals/RESEARCH/REPEVOL>

CONTEXTE

Une des énigmes majeures de l'évolution des génomes reste le rôle des phénomènes de duplication de l'ADN. Les événements de duplication sont à l'origine de l'accroissement de la taille des génomes, des familles de gènes et des locus de répétitions en tandem de tout le génome humain.

Malgré cela, le rôle et les raisons de leur existence et les mécanismes qui contrôlent leur développement sont encore peu connus. En outre, on sait que des répétitions en tandem influent sur le contrôle de l'expression de gènes ou le développement de maladies telles que la chorée de Huntington ou le diabète de type I.

OBJECTIFS

Ce projet vise à avancer dans la compréhension de l'évolution de diverses structures répétées. Il exige d'abord de se doter d'outils informatiques et de modèles performants pour :

- localiser ces structures répétées tant en tandem que distantes,
- comparer différentes répétitions ou allèles issus d'un site de répétition,
- reconstruire leur histoire évolutive en tenant compte des événements de duplication.

L'autre objectif est d'appliquer ces outils aux données biologiques pour étudier diverses structures répétées et plusieurs aspects évolutifs. Tout d'abord, recenser les répétitions distantes à l'intérieur des génomes procaryotes, puis eucaryotes, pour découvrir des régions conservées ayant un rôle régulateur ou des familles de gènes. Ensuite, caractériser les microsatellites de génomes eucaryotes par une approche de génomique comparative pour élucider les forces qui contrôlent leur existence. Retracer l'histoire évolutive de familles de gènes importantes du point de vue fonctionnel, comme les quelques 200 protéines de la plus grande famille de l'arabette. Enfin, comparer des allèles de sites de répétition variables pour deux types d'applications, soit pour examiner un rôle fonctionnel, soit pour inférer des relations évolutives entre individus, populations ou souches bactériennes. Ces applications serviront aussi à valider les algorithmes conçus préalablement.

PROJET

Améliorer les algorithmes existants pour prendre en compte toute la complexité évolutive de ces structures.

Appliquer ces outils aux données biologiques pour étudier diverses structures répétées et plusieurs aspects évolutifs.

RESULTATS

- STAR: Logiciel de détection de répétitions en tandem approximatives [1],
- Etude des problèmes algorithmiques liés à l'évolution des répétitions en tandem, [2] et [5],
- Recherche des répétitions et des similarités par les méthodes de filtration [3],
- Nouvelle méthode de recherche locale pour les arbres de duplication en tandem, optimisant les critères de parcimonie et d'évolution [4],
- Modélisation par des modèles markoviens de la population de microsatellites et comparaisons dans plusieurs génomes [6],

CALENDRIER

- 1 Octobre / Novembre (Paris) Démarrage
- 2 Janvier (Montpellier) Réunion générale
- 2006 Visite de Gary Benson en France (Montpellier)

LABORATOIRES IMPLIQUES

CEFE / UMR 5175 / Montpellier
IGH/ UPR 1142 / Montpellier
LIRMM / UMR5506 / Montpellier
LORIA / UMR7503 / Nancy
Laboratory for Biocomputing and Informatics, Boston, USA

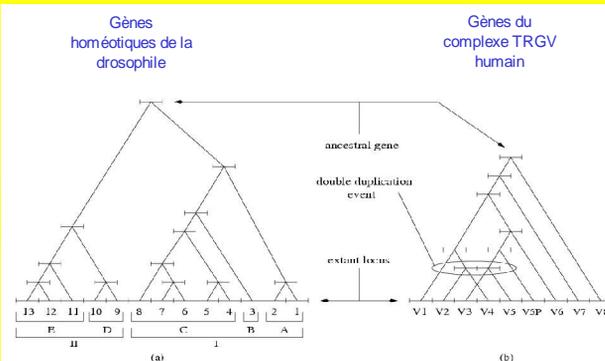
MOTS CLES

Répétitions distantes et en tandem, micro-, mini-satellites, familles de gènes, histoire de duplications, algorithmes de recherche, d'alignement de répétitions.

PUBLICATIONS PRINCIPALES

- [1] O. Delgrange, E.Rivals STAR : An algorithm to Search for Tandem Approximate Repeats, *Bioinformatics*, 20 (16):2812-2820, 2004
- [2] E.Rivals.A.Survey on Algorithmic Aspects of Tandem Repeats Evolution. *International Journal of Foundations of Computer Science*, 15 (2):225-257,2004.
- [3] L. Noe, G. Kucherov, YASS: enhancing the sensitivity of DNA similarity search, *Nucl. Acids Res.* 2005 33: W540-543
- [4] Bertrand and O. Gascuel. Topological rearrangements and local search method for tandem duplication trees. *Transactions on Computational Biology and Bioinformatics*, 2005. In press.
- [5] O.Gascuel, D. Bertrand, O. Elemento: Reconstructing the duplication history of tandemly repeated sequences. In O. Gascuel, editor, *Mathematics of Evolution and Phylogeny*, 205–235. Oxford University Press, 2005.
- [6] F.Munoz, E.Le Roy, E. Rivals, P. Jarne : Models of dinucleotide microsatellite evolution in the yeast, worm and human genome, In preparation.
- [7] C. Alkan, E. Tuzun, J. Buard, F. Lethiec, E. Eichler, J., Bailey, C., Sahinalp, Manipulating multiple sequence alignments via MaM and WebMaM. *Nucl. Acids Res.* 2005 33: W295-298

ARBRES DE DUPLICATION



Arbres de duplication retraçant l'évolution de clusters de gènes adjacents sur un chromosome. Ces arbres ont été inférés à partir des séquences contemporaines. Le gènes actuels sont représentés par les feuilles de l'arbre et sont placés dans l'ordre chromosomique. Un nœud interne de l'arbre est un événement de duplication simple qui à partir d'une copie d'un gène ancestral ajoute une nouvelle copie supplémentaire de ce gène. Un ensemble de nœuds encerclés représente une duplication multiple portant sur plusieurs (ici 2) gènes ancestraux adjacents. La plupart des duplications sont simples, comme en atteste l'arbre des gènes homéotiques ; dans ce cas l'arbre de duplication est formellement équivalent à un arbre de recherche de l'informatique. Lorsque des duplications multiples se produisent, les arbres correspondant constituent une classe plus large d'objets aux propriétés combinatoires surprenantes et encore mal comprises.