

## Impact et évolution des répétitions dans les génomes : interactions entre innovation génétique et fonction biochimique

<http://www.wabi.snv.jussieu.fr/evolrep/index.html>

### 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 de la création de la nouveauté fonctionnelle.

Malgré cela, le rôle et les raisons de leur existence et les mécanismes qui contrôlent leur développement sont encore peu connus. L'étude de l'évolution des duplications nécessite leur caractérisation aux 3 niveaux pertinents de description - séquences nucléiques, séquences protéiques et structures 3D des protéines - afin d'appréhender l'importance relative de ces 3 niveaux.

### OBJECTIFS

Du point de vue algorithmique, on considère, pour chaque type de méthode explorée (par exemple algorithmes de recherche de motifs flexibles répétés), différentes représentations correspondant aux trois niveaux de description, de manière à garder autant que possible les méthodes de recherche de duplications en invariant et à ne faire varier que la représentation.

Nous utilisons ces méthodes dans le but d'analyser les interactions que ces duplications intra-génomiques peuvent avoir aux trois niveaux de description, et d'explorer l'impact de ces interactions sur leur évolution. Deux équipes de biologistes expérimentalistes testeront les hypothèses qui concernent la recombinaison et l'évolution des duplications.

Ce projet recouvre deux des axes définis comme prioritaires :

(i) Structure des génomes / génomique comparative (et évolutive) : les études sur les duplications nucléiques ont des implications directes dans la compréhension de la structure des génomes, et l'étude comparative de ces duplications sur des génomes proches peut fournir une explication historique des mécanismes impliqués dans la dynamique du génome.

(ii) Etude structurale, dynamique et fonctionnelle des molécules du vivant : la compréhension de la relation entre les duplications de l'ADN et la structure des protéines permettra d'avancer dans le domaine de l'étude des relations entre séquences nucléiques et structure/fonction des protéines.

Nous mettons à la disposition de la communauté un ensemble intégré des programmes et méthodes développées, incluant les résultats obtenus, dans le cadre de ce projet.

### PROJET

Étudier les duplications directes proches en prenant en compte à la fois la séquence du gène, celle du polypeptide et quand cela est possible sa structure tridimensionnelle. Étudier l'évolution de ces duplications par une analyse comparative d'espèces proches. Développer et implémenter les algorithmes et méthodes statistiques appropriés.

### RESULTATS

- Identification et caractérisation des répétitions directes proches (DPPI) dans les génomes bactériens [1]
- Programme pour cribler efficacement une banque de structures [2]
- Identification et caractérisation des systèmes de recombinaison chez les génomes bactériens [3]
- Programme pour chercher des répétitions longues non-strictes dans des longues séquences d'ADN [4]
- Définition algorithmique et implémentation d'un modèle de recherche de motifs relationnels [5, 6].
- Mise en œuvre d'un programme de recherche intégrée des DPPI aux 3 niveaux: structure protéique, séquence protéique et séquence nucléique [7]

### CALENDRIER

2003	Paris, démarrage
2004	Paris, réunion générale
2005	Paris, réunion générale (à faire)

### LABORATOIRES IMPLIQUES

Atelier de Bioinformatique, Université Pierre et Marie Curie  
Unité Génétique des Génomes Bactériens, URA2171, Institut Pasteur  
Equipe Structure et Dynamique du Génome, UMR7592, Institut Jacques Monod  
Equipe Helix, INRIA Rhône-Alpes  
Laboratoire d'Informatique de Paris Nord, UMR7030, Université Paris XIII  
Laboratoire Adaptation et Pathogénie des Microorganismes

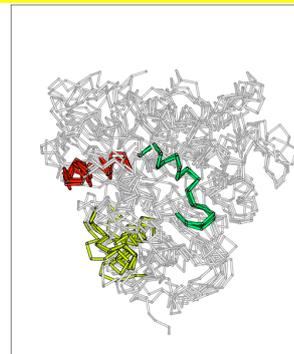
### MOTS CLES

structure du génome / dynamique du génome / génomique comparative / évolution / duplications / recombinaison / conversion / bactérie / levure / structure 3D des protéines / algorithmique de la recherche de répétitions / motifs répétés flexibles / changements de représentations

### PUBLICATIONS PRINCIPALES

- [1] Rocha, E.P.C. (2003) An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: from duplications to genome reduction. *Genome Res*, 13, 1123-1132.
- [2] M. Carpentier, S. Brouillet et J. Pothier (2005), YAKUSA : a fast structural databases scanning method. *Proteins : Structures, Fonctions and Bioinformatics*, in press
- [3] Rocha EPC, Cornet E, Michel B (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems *Plos Genetics*, in press
- [4] Achaz G, Boyer F, Rocha EPC, Viari A, Coissac E, Extracting approximate repeats from large DNA sequences. soumis
- [5] Pisanti, N., Soldano, H., and Carpentier, M. 2005. Incremental Inference of Relational Motifs with a Degenerate Alphabet. In 16th Annual Symposium on Combinatorial Pattern Matching (CPM 2005), Jeju Island, Korea, LNCS 3537. 229-240 Springer-Verlag
- [6] Pisanti N, Soldano H, Carpentier M, Pothier J Implicit and Explicit Representation of Approximated Motifs, soumis.
- [6] Abraham, AL, Carpentier M, Pothier J, Rocha EPC N, «Détection de duplications internes dans les structures des protéines », JOBIM 2005, 6-8 juillet 2005, Lyon.

### MOTIFS RELATIONNELS ET STRUCTURES 3D



Motifs relationnels "structuraux" de longueur 18 pour les 5 cytochromes P450 4CP4, 1ROM, 1CPT, 2HPD chain B et 3CPP. Les relations entre éléments ( $C\alpha$  des acides aminés) sont ici leurs distances internes.

Analyse intégrée d'une duplication directe proche interne (DDPI) dans la protéine 1a9x (carbonyl phosphate synthetase) et son gène. La duplication apparaît en vert et en orange dans la figure. La duplication se traduit par deux sous domaines structuraux similaires et symétriques. La duplication 3D est plus étendue que la duplication trouvée en acides aminés, cette dernière étant environ deux fois plus étendue que la duplication nucléique détectée.

