

## Détection *in silico* des origines de réplication chez les eucaryotes supérieurs : analyse et évolution

### CONTEXTE

La réplication de l'ADN chez les eucaryotes, utilise de nombreuses origines réparties le long du génome. L'« allumage » des différentes origines est coordonné au déroulement du cycle cellulaire, et fonctionnellement relié à l'état de la chromatine ainsi qu'à la chronologie de la transcription. La connaissance des origines est donc essentielle pour mieux comprendre les liens reliant ces mécanismes. Toutefois, chez les eucaryotes supérieurs, seul un petit nombre d'origines ont été identifiées expérimentalement et aucun consensus de séquence n'a pu être établi. Plusieurs points de vue coexistent aujourd'hui pour décrire l'initiation de la réplication dans ces organismes : un démarrage pouvant utiliser des séquences spécifiques, comme chez *S. cerevisiae*, ou un démarrage s'appuyant plutôt sur des éléments de nature épigénétique.

### OBJECTIFS

Dans cette problématique, l'objectif de ce projet est la détection des origines de réplication actives dans la lignée germinale, et la caractérisation statistique des réplicons à l'aide de méthodologies originales d'analyse des séquences génomiques.

Les variations de biais de composition (biais T/A et G/C) ont été utilisées avec succès pour trouver les origines de réplication chez de nombreuses bactéries, organelles et virus, mais à ce jour, cette méthode n'a pas donné de résultat chez les eucaryotes supérieurs. Pour surmonter certaines difficultés rencontrées chez ces organismes (hétérogénéité de la taille des origines et des réplicons, biais de composition lié à la transcription), nous avons entrepris l'analyse des biais à l'aide d'une méthodologie multi-échelle, la transformée en ondelettes (TO). L'examen à différentes échelles des régions transcrites et non-transcrites, ainsi que des origines de réplication connues, indique que la transcription et la réplication participent toutes deux aux profils de biais observés. La possibilité de séparer ces deux contributions sera au cœur de notre projet et exploitera les formes distinctes des profils de biais de transcription (crêteaux) et de réplication (dents de scie dues à la nature diffuse de la terminaison) via un choix de formes d'ondelettes optimales. Sur la base de nos récentes observations sur les profils de biais autour des origines de réplication chez l'homme et chez plusieurs mammifères (souris, rat), nous développerons des approches comparatives dans la représentation espace-échelle fournie par la TO, pour étudier la conservation des origines chez les mammifères.

### PROJET

Objectifs :

- (1) Montrer l'existence d'un biais réplicatif chez les mammifères.
- (2) Détection *in silico* de nouvelles origines de réplication dans le génome humain.
- (3) Automatiser la détection *in silico* des origines de réplication dans des génomes entiers par l'utilisation de formes d'ondelettes analysatrices adaptées.
- (4) Examiner les liens entre réplicons et domaines d'expression génique.
- (5) Etudier les origines par approche phylogénétique (séquences, distribution spatiale)
- (6) Etudier les profils de biais et détermination *in silico* des origines de réplication chez divers mammifères (souris, rat, chien).

### RESULTATS

- (1) Mise en évidence d'un biais réplicatif dans le génome humain et d'un profil caractéristique des biais de composition autour des origines de réplication. Conservation des profils de biais autour des origines des mammifères (Figure 1).
- (2) Détection de plus 1000 origines de réplication dans le génome humain par une méthodologie basée sur l'analyse en ondelettes (Figure 2).

### CALENDRIER

Résultats attendus:

- 1<sup>ère</sup> année : objectifs (1) et (2)
- 2<sup>ème</sup> année : objectifs (3) et (4)
- 3<sup>ème</sup> année : objectifs (5) et (6)

### LABORATOIRES IMPLIQUES

Centre de Génétique Moléculaire/Gif-sur-Yvette  
Laboratoire Joliot-Curie/ENS de Lyon

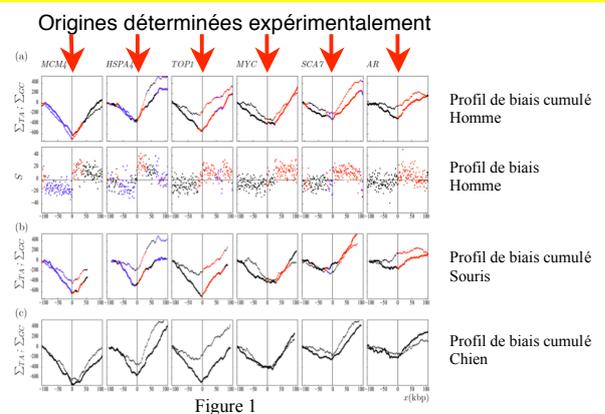
### MOTS CLES

Replication origins, mammalian genomes, strand compositional asymmetries, wavelet transform

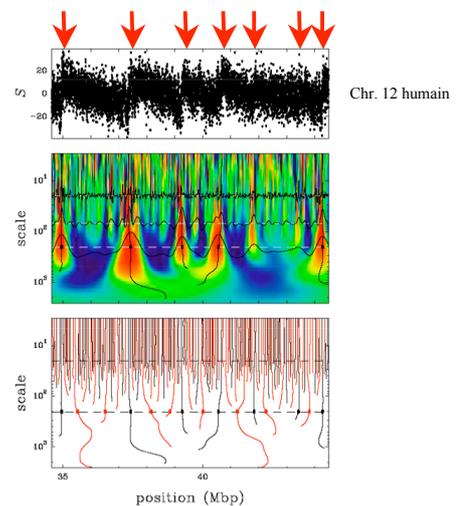
### PUBLICATIONS PRINCIPALES

- [1] M. Touchon, S. Nicolay, B. Audit, E.B. Brodie of Brodie, Y. d'Aubenton-Carafa, A. Arneodo & C. Thermes. Replication-associated strand asymmetries in mammalian genomes: towards detection of replication origins. (2005) *Proc. Natl. Acad. Sci. USA*, **102**, 9836-9841.
- [2] E.-B. Brodie of Brodie, S. Nicolay, M. Touchon, B. Audit, Y. d'Aubenton-Carafa, C. Thermes & A. Arneodo. From DNA sequence analysis to modelling replication in the human genome. (2005) *Phys. Rev. Lett.*, **94**, 248103.

### Détection des origines de réplication par l'analyse des profils de biais



### Détection de nouvelles origines



Au cours de l'évolution, les mutations n'affectent pas de façon identique les deux brins d'ADN génomique. Ceci est principalement dû à des asymétries des mécanismes de mutation et de réparation associés à la réplication et à la transcription.

Chez les procaryotes, un excès de G par rapport à C (biais GC) et de T par rapport à A (biais TA) sont fréquemment observés sur le brin avancé. Ces biais changent brutalement de signe lorsque l'on traverse l'origine de réplication et le site de terminaison, produisant ainsi un profil en forme de crêteaux, caractéristique de la réplication.

Chez les mammifères, un biais couplé à la transcription a été détecté, mais jusqu'à maintenant, aucun biais associé à la réplication n'a été mis en évidence. L'analyse des régions intergéniques et transcrites bordant les origines de réplication identifiées expérimentalement démontre l'existence d'une asymétrie de composition de brin associée à la réplication. L'analyse multi-échelle du profil de biais du génome humain nous a permis d'identifier un groupe de plus de mille régions candidates d'initiation de la réplication. Autour de ces origines, le profil de biais présente une forme caractéristique en dents de scie également observée dans les génomes de la souris et du chien.

Ces résultats nous permettent de proposer un modèle de terminaison de la réplication reposant sur l'existence de multiples sites de terminaison distribués aléatoirement entre deux origines de réplication adjacentes. L'ensemble de ces analyses constitue une avancée à l'échelle génomique, dans la compréhension des mécanismes de la réplication.