

## Extension du modèle GATE et implémentation de la base de connaissance FlyGATE dédiés aux éléments transposables

### CONTEXTE

Les éléments transposables (ET), séquences d'ADN qui ont la capacité de changer de localisation sur les chromosomes, ont été trouvés dans des organismes aussi différents que les bactéries, les nématodes, la levure, les plantes, les poissons, les mammifères y compris l'homme. Ils représentent une part variable mais importante des génomes et sont responsables d'une forte proportion des mutations.

Les interactions des ET avec leur hôte et les relations entre ET sont complexes. Plusieurs familles d'ET (ensemble des copies plus ou moins divergées et dégradées du génome issues les unes des autres par transposition) existent à l'intérieur d'un même génome chacune reflétant des mécanismes de transposition utilisés. C'est avec les données récentes issues des séquençages à grande échelle des génomes comme ceux du nématode, de la drosophile que l'on a accès à la notion de familles d'ET en termes à la fois de diversité de séquences et de distribution le long des chromosomes. Ceci représente une grande quantité d'information qu'il convient d'organiser et de rendre facilement accessible notamment en l'intégrant dans des

### PROJET

Il est difficile d'extraire des bases de données existantes l'information pertinente pour nombre de questions biologiques. C'est pourquoi nous avons représenté les concepts nécessaires à l'analyse des ET au sein des génomes eucaryotes, tant d'un point de vue dynamique au sein d'un génome que d'un point de vue évolution de structure, par un modèle de type objet-association appelé GATE (*Genome Analysis for Transposable Elements*) en utilisant le langage AROM (Allier Relation et Objets pour Modéliser). Ce modèle est utilisé dans le cadre du développement d'une base de connaissances appelée FlyGATE dédiée à l'analyse des ET du génome séquencé de la

### OBJECTIFS

- Implémenter la base FlyGATE dédiée aux éléments transposables (ET) du génome de *Drosophila melanogaster*, et mettre cette base à la disposition de la communauté internationale grâce à une interface simple, accessible sur internet et permettant d'effectuer des requêtes sur la base.
- FlyGATE permettra une recherche efficace des relations entre les ET et les gènes, les régions régulatrices, les séquences répétées, les taux de recombinaison... des génomes, ainsi qu'une analyse comparative des différentes familles d'ET chez la drosophile.
- appliquer GATE à d'autres génomes que la drosophile et créer ainsi "MosquitoGATE", "NématodeGATE"...

### RESULTATS

- Extension et implémentation du modèle de représentation des connaissances GATE
- A ce jour, la base contient les données concernant la distribution des 1571 ET et des 14 015 gènes du génome. Une première interface web en cours de développement est disponible.
- Les formulaires implémentés concernent essentiellement la distribution des copies d'ET le long des chromosomes, en fonction de leur type, familles, voisinage et orientation. Cette interface permet de répondre à des questions biologiques, telles que :
  - Quelles sont les copies d'ET insérées dans une autre copie d'ET ?
  - Quelles sont les copies de rétrotransposons à LTR situé à moins de 1000 pb en amont d'un gène de l'hôte, sur le bras chromosomique 2L ?

### LABORATOIRES IMPLIQUES

- UMR 5558: "Biométrie et Biologie Evolutive" / Lyon
- UR INRIA Rhône-Alpes - Projet HELIX / Grenoble

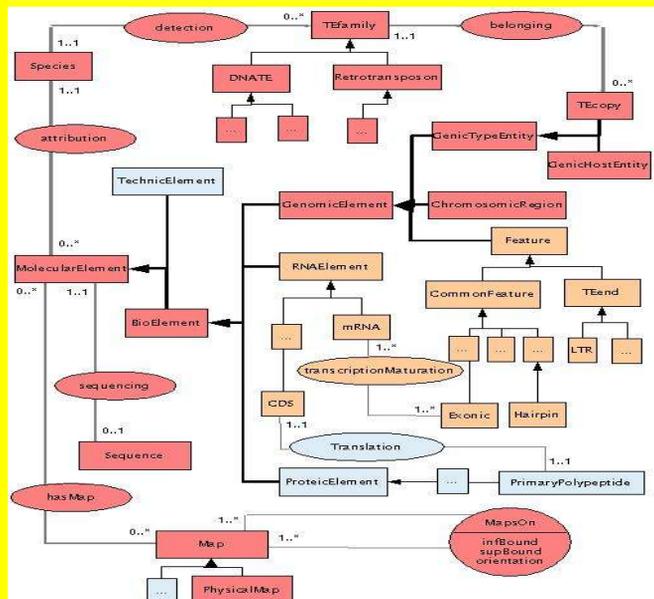
### MOTS CLES

Éléments transposables, génome, représentation des connaissances, ontologie, UML, AROM.

### PUBLICATIONS PRINCIPALES

- Rizzon, C., Marais, G., Gouy, M., Biémont, C. 2002 Transposable elements distribution in relation to recombination rate in the *Drosophila melanogaster* genome. *Genome Res.* 12: 400-407.
- Lerat, E., Rizzon, C., Biémont, C. 2003 Sequence divergence of transposable elements in the *D. melanogaster* genome. *Genome Res.* 13: 1889-1896.
- Rizzon, C., Marais, G., Gouy, M., Biémont, C. 2002 Transposable elements distribution in relation to recombination rate in the *Drosophila melanogaster* genome. *Genome Res.* 12: 400-407.
- Rizzon, C., Martin, E., Marais, G., Duret, L., Ségalat, L., Biémont, C. 2003 Patterns of selection against transposons inferred from the distribution of Tc1, Tc3, and Tc5 insertions in

### GATE : *Genome Analysis for Transposables Elements*



**Organisation générale du modèle GATE.** Seules les classes principales ainsi qu'une partie des hiérarchies de classes qui en découlent sont représentées. Aucun attribut n'est représenté. En rouge les classes principales, en jaune les arborescences d'héritage incomplètes, en orange les associations principales du modèle.

- TEfamily représente la notion de famille d'ET
- TECopy correspond aux copies d'ET.
- La hiérarchie MolecularElement représente tous les éléments moléculaires. La classe BioElement est spécialisée selon les trois natures chimiques : DNA, RNA, protéique. Au sein de la classe GenomicElement, les entités géniques (GenicTypeEntity) telles que les gènes et les copies d'ET sont distinguées des éléments structuraux (Feature).
- La classe Map permet de représenter tout élément biologique en terme de carte, et ainsi de pouvoir positionner chaque carte par rapport à une autre.