

**Acronyme du projet : REGLIS**

**Titre du projet : De la molécule à la cellule : développement, confrontation et intégration de modèles formels et de méthodes d'analyse**

**Nom du porteur de projet : Sagot**

**Prénom du porteur de projet : Marie-France**

**Discipline : Sciences et Technologies de l'Information (1<sup>er</sup>) et Biologie et Santé (2<sup>ème</sup>)**

**Montant du financement ANR : 310 000 euros**

**Noms, prénoms et laboratoires d'appartenance et organismes des responsables des partenaires :**

Aucun partenaire

**Mots Clés** (5 principaux):

- Bioinformatique
- Algorithmique
- Combinatoire
- Biologie moléculaire
- Biologie cellulaire

**Résumé du projet** (15 lignes maximum) :

Ce projet, largement exploratoire, avait pour but, à travers une approche comparative à partir de données disponibles publiquement, d'arriver à une meilleure compréhension du réseau complexe d'interactions spatiales et temporelles entre les divers éléments composant le vivant (gènes, métabolites, etc.) et de ces derniers, individuellement ou collectivement, avec l'environnement. Une telle compréhension obligeait à adopter un point de vue à la fois local et global, statique et dynamique d'un organisme, allant du niveau moléculaire (génome) jusqu'à une vision générale du fonctionnement d'une cellule (réseaux biologiques). Plus spécifiquement, nous souhaitions nous intéresser à trois questions biologiques très générales : 1. existe-t-il des régularités, structurales ou fonctionnelles, dans la diversité qui est observée, qui pourraient représenter des indices d'une organisation profonde du vivant; 2. pouvons-nous identifier ces régularités de façon systématique et arriver ainsi à dégager un ordre dans le réseau complexe des interactions détectées; enfin, 3. comment ce réseau s'est-il mis en place au cours de l'évolution, pour accomplir quelles fonctions ? Les résultats attendus concernaient : 1. l'obtention de meilleurs formalismes de modélisation mathématique et algorithmes d'analyse; et 2. des éléments de réponse à la fois à des questions biologiques spécifiques (tests d'hypothèses) et plus générales (exploration systématique des données disponibles).

**Verrous ou points durs** (scientifiques, technologiques, organisationnels...) à dépasser (8 lignes maximum) :

Les verrous principaux du projet étaient liés à sa nature très exploratoire, initialement sur des questions biologiques extrêmement vastes et relativement vagues - existe-t-il des régularités, structurales ou fonctionnelles, dans la diversité qui est observée, qui pourraient représenter des indices d'une organisation profonde du vivant ? – dans un domaine qui en outre est fortement compétitif depuis quelques années. Les défis étaient donc de réussir à apporter des éléments de réponse à une question aussi large, en fait de dégager des voies de réponse ou d'exploration plus précises, tout en contribuant au principal objectif méthodologique qui consistait à obtenir de meilleurs formalismes de modélisation mathématique et algorithmes d'analyse.

**Résultats majeurs** (8 lignes maximum) :

Il est clair que nous ne pouvions arriver aussi rapidement à une réponse nette et claire aux trois questions biologiques posées. Par contre, nous pensons avoir clairement rempli l'objectif de dégager de meilleurs formalismes de modélisation mathématique et algorithmes d'analyse, que ce soit au niveau génomique (lien entre recombinaison et organisation chromosomique, blocs de synténie et régions de cassure, génomes ancestraux, scénarios d'inversions) qu'à celui des réseaux biologiques (motifs, flux et ensembles minimaux de précurseurs, modules de régulation). À travers des tests d'hypothèse et une exploration systématique des données, nous avons aussi dégagé des régularités que nous sommes maintenant en mesure d'analyser selon différentes voies plus spécifiques.

Production Scientifique depuis le début du projet (6 principales publications à comité de lecture, brevets, conférences invitées, nb de colloques hors conférences invitées...)

### **Publications à comité de lecture** (6 principales) :

- M. D. V. Braga, M.-F. Sagot, C. Scornavacca and E. Tannier. Exploring The Solution Space of Sorting by Reversals With Experiments and an Application to Evolution, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(3):348-356, 2008.
- L. Cottret, P. V. Milreu, V. Acuña, F. V. Martinez, A. Marchetti Spaccamela, M.-F. Sagot, L. Stougie. Enumerating Precursor Sets of Target Metabolites in a Metabolic Network. *Proceedings of WABI'08, Lecture Notes in BioInformatics*, vol. 5251, pages 233-244, 2008.
- P. G. S. da Fonseca, C. Gautier, K. S. Guimaraes and M.-F. Sagot. Efficient representation and P-value computation for high order Markov motifs. *Proceedings of the European Conference on Computation Biology (ECCB'08). Bioinformatics*, 24:160-166, 2008.
- V. Lacroix, C. G. Fernandes, and M.-F. Sagot. The Motif Search Problem in Graphs: Application to Metabolic Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4):360-368, 2006
- C. Lemaître, E. Tannier, C. Gautier, M.-F. Sagot. Precise detection of rearrangement breakpoints in mammalian chromosomes. *BMC Bioinformatics*, 9:286-322, 2008.
- E. Tannier, C. Zheng and D. Sankoff. Multichromosomal genome median and halving problems. *Proceedings of WABI'08, Lecture Notes in BioInformatics*, vol. 5251, pages 1-13, 2008.

### **Conférences invitées** (5 principales) :

33 conférences invitées sur la période dont :

- L. Cottret, « Genome-based metabolic reconstructions and functional analysis ». Systry Meeting, Glasgow University, Scotland, 5-6 Juin 2008.
- C. Gautier, Some bioinformatic tools for studying microbial biodiversity, Meeting COST 853 « Agricultural Biomarkers for Array-Technology », Zurich, Suisse, 18-19 Septembre.
- C. Lemaître, « A method to detect precisely rearrangement breakpoints in mammalian genomes ». Dynamics of Genomes Workshop, Valparaiso, Chili, 27-28 Mars 2008.
- E. Tannier, « Genome rearrangements: the two parcimonies ». Minisymposium on computational biology at the CANADAM 2007 Conference, Alberta, Canada, 27-31 Mai, 2007.
- M.-F. Sagot, « Open combinatorial / graph problems in computational biology », Séminaire à l'Université « La Sapienza », Rome, Italie, 4 Février 2008.

### **Nombre de Colloques** (hors conférences invitées) :

14 conférences non invitées sur la période

### **Personnels (en hommes x mois)**

Exemple : un enseignant chercheur impliqué à 30% sur 36 mois aura une implication de 10.8 hommes/mois ( $36 \times 30\% = 10,8$ )

Chercheurs + enseignants chercheurs + ingénieurs de recherche : 108.6  
Personnel recruté dans le cadre du projet (CDD sur financement ANR) : 88

**Images**

Merci de nous renvoyer par mail de 1 à 4 images/photos, illustrant vos travaux de recherche en 300 dpi, de préférence en format paysage, fichiers jpeg ou gif.

Titre Image 1: Raffinement des points de cassure survenus suite à des réarrangements chromosomiques dans les génomes de mammifères

Titre Image 2: La recombinaison comme un exemple d'interaction forte entre processus moléculaires, fonctionnement cellulaire, information génomique (dont épigénétique), et génétique populationnelle

Titre Image 3: Algorithme de recherche et d'inférence de motifs colorés dans les réseaux métaboliques et d'interactions

Titre Image 4: Comparaison de réseaux métaboliques et recherche d'ensembles minimaux de précurseurs

**Nous vous remercions pour ces informations**