

Rapport semestriel d'activité n°2 datant du 15 janvier 2007

A. Identification

Programme – année	Programme non thématique 2005
Projet (acronyme)	REGLIS
Coordonnateur du projet (société/organisme - laboratoire ou entité de rattachement)	Équipe BAOBAB – Laboratoire de Biométrie et Biologie Évolutive
Référence convention/décision	NT05-3_45205
Période du projet (date début – date fin)	15 décembre 2005 – 15 décembre 2008
Période faisant l'objet du rapport d'activité (date début – date fin)	15 juin 2006 – 15 janvier 2007
Rédacteur de ce rapport : nom	SAGOT Marie-France
téléphone	04-72-44-82-38
adresse électronique	marie-france.sagot@inria.fr
Date du rapport	15 janvier 2007

B. Pour les projets multi-partenaires, rappel des tâches allouées par partenaire pour l'ensemble du projet

Le projet REGLIS n'a qu'un seul partenaire, le partenaire principal (équipe BAOBAB-HELIX).

Eléments qualitatifs

C. Description des travaux effectués pour la période concernée et conformité de l'avancement aux prévisions

Nous n'indiquons dans le présent rapport que les travaux pertinents aux six derniers mois du projet. Certains sont la continuation de travaux initiés et déjà en partie introduits lors du rapport précédent. Nous avons essayé dans ce cas de ne présenter en détail que ce qui se réfère à la période couverte par ce rapport (15 Juin 2006 à 15 Janvier 2007). D'autres travaux décrits ci-dessous ont été initiés depuis le dernier rapport d'activités du projet (datant de Juin 2006). Cela concerne principalement les sections a1 (premier point), a4 (tous les points sauf le premier qui avait été initié dès le premier semestre du projet), a5, b2, et e.

C1. Inférence et mise en évidence de régularités :

a. Au niveau génomique

1. Amélioration des algorithmes de recherche de régularités dans les séquences nucléotidiques

- **Filtrage de séquences en vue d'un alignement multiple** : Dans le but de proposer des méthodes permettant la détection rapide et exacte de régularités (approximatives, c'est-à-dire, pouvant contenir des substitutions, insertions ou suppressions de bases) dans les génomes, nous avons travaillé sur la création de filtres. Un filtre permet de supprimer d'une séquence génomique, ou d'un ensemble de séquences génomiques, de larges portions ne pouvant pas appartenir à une répétition. Ainsi, lorsqu'un utilisateur souhaite détecter de manière exacte et rapide les éléments d'une répétition, l'utilisation de ce type de filtre permet de réduire considérablement la masse de données à traiter sans altérer le résultat attendu. Nous nous sommes focalisés sur des filtres pour des alignements multiples, c'est-à-dire, pour des répétitions possédant possiblement plus de deux occurrences. Ces travaux sont réalisés en collaboration avec Alair Pereira do Lago, professeur à l'Université de São Paulo (USP).
- **Segmentation de séquences** : La recherche d'un même type de régularités dans une séquence pose l'hypothèse que cette séquence est homogène vis-à-vis du type d'information que l'on veut extraire par cette recherche. Or, il est clair qu'une telle hypothèse est extrêmement forte, surtout quand on considère de grandes séquences nucléotidiques (telles les chromosomes). Ainsi, la recherche de régularités dans de grandes séquences ne peut en général se faire en dehors de la recherche de zones homogènes (selon des critères ad hoc) dans ces séquences, soit en amont, soit en aval de l'étude des régularités. C'est la raison principale pour laquelle des méthodologies telles que les chaînes de Markov à états cachés ont été appliquées à l'étude de séquences nucléotidiques. En parallèle à cette approche, nous avons mis au point un algorithme pour calculer, étant donnée un modèle, la probabilité pour une séquence d'être structurée en un certain nombre de segments. Cet algorithme est un complément indispensable de la méthode de partitionnement maximale prédictif développée il y a quelques années (L. Guéguen, LNCS 2066, 32-45, 2001).

2. Recherche de régularités dans les régions de cassure suite à un réarrangement

La comparaison de l'ordre des gènes entre plusieurs génomes permet l'identification des ruptures de synténie. Ces ruptures constituent des régions du génome qui ont potentiellement subi un ou plusieurs réarrangements génomiques. L'analyse fine de ces régions, au niveau génomique, pourrait permettre de mieux comprendre les causes et les mécanismes de ces événements évolutifs au niveau de la cellule. L'identification de régularités dans ces régions permettrait de localiser de manière systématique les réarrangements ou les régions susceptibles d'être réarrangées.

Ce travail a nécessité le développement d'une méthode pour identifier très précisément les points de cassure. Celle-ci est basée sur la comparaison de l'ordre des gènes entre plusieurs espèces et sur l'alignement des séquences intergéniques impliquées dans les régions de cassure. Cette méthode a ensuite été appliquée à plusieurs jeux de données, pour comparer son efficacité en fonction des caractéristiques des données.

3. Étude des liens entre éléments répétés de différents types entre eux ainsi qu'avec d'autres facteurs, dont le niveau d'expression des gènes

Nous avons commencé à mener une large étude des liens entre éléments répétés de différents types, ainsi qu'entre ces éléments et d'autres facteurs, dont le niveau d'expression des gènes. Dans une première étude, nous nous sommes intéressés à l'échelle génomique au niveau d'activation de certains types d'éléments répétés, les éléments transposables. De façon un peu surprenante, des études de ce type sont rares. Nous avons choisi de mener une telle analyse sur le génome de *Drosophila melanogaster*. Ce travail a été effectué avec Cristina Vieira de l'équipe « Génomes et Populations » de notre laboratoire. Des données d'EST disponibles publiquement ont été utilisées. La difficulté du problème réside dans l'assignation d'une EST au « bon » élément transposable. En effet, le séquençage d'ESTs est susceptible d'erreurs alors même que les transposons d'une même famille sont souvent des copies quasi-exactes les uns des autres, rendant ainsi une assignation correcte particulièrement difficile.

4. Régularités et évolution : Reconstitution de scénarios évolutifs par remaniements chromosomiques

Les régularités dans l'organisation génomique observées dans différentes espèces nous permet d'essayer d'inférer des scénarios évolutifs de cette organisation. Étant donné leur interconnexion, nous avons choisi d'aborder en parallèle plusieurs questions relatives à de tels scénarios. Ces questions sont les suivantes.

- **Exploration des solutions du tri par inversions.** Le plus gros problème des méthodes combinatoires actuelles pour les reconstitutions de scénarios d'inversions est le très grand nombre de solutions qu'elles fournissent, sans critère pour les départager. Nous avons développé un algorithme qui permet d'explorer l'espace des solutions sans le parcourir entièrement, ce qui prendrait trop de temps et trop de mémoire. Cet algorithme représente une des étapes importantes pour aborder la question suivante.
- **Reconstitution de l'histoire moléculaire de la différenciation sexuelle.** En examinant les chromosomes actuels humains X et Y, nous tentons de découvrir les traces de leur histoire, fortement marquée par les remaniements, et en particulier par les inversions. Nous approchons ainsi par des méthodes combinatoires la question de savoir si ces inversions ont joué le premier rôle dans la différenciation, ou si elles l'ont simplement accompagnée.
- **Nouvelles approches du tri par inversion en présence d'éléments dupliqués.** Les chromosomes X et Y comportent des duplications de gènes. Or la majorité des méthodes de calcul de distance de réarrangements entre génomes (notamment, par inversions) ignorent ces duplications dans le seul but de simplifier le problème. Nous avons donc commencé à considérer différentes façons algorithmiques d'aborder ce problème de manière biologiquement plus réaliste. Bien que motivée au départ par les chromosomes X et Y, la méthode est considérée dans le cas général et générique du calcul d'une distance de réarrangements entre génomes comportant des familles de gènes. Ce travail est mené en collaboration avec des chercheurs mathématiciens de l'Université de São Paulo, Brésil et de l'Université Fédérale de Mato Grosso do Sul à Campo Grande, São Paulo.
- **Phylogénie des primates.** Les données moléculaires ne permettent pas d'obtenir une phylogénie définitive de certains groupes de primates, qui semblent avoir évolué de façon complexe, par spéciations rapides et successives, ou par spéciations progressives. Les réarrangements chromosomiques, en raison de leur relative rareté dans l'évolution, sont des marqueurs précieux, mais ne permettent pas une reconstitution arborée dans tous les cas. Il s'agit de trouver une bonne définition formelle de l'objet phylogénique qui permettra de représenter les cas d'évolution complexe, et de développer des algorithmes qui construisent ces objets à partir des données de remaniements. Ces travaux sont menés avec la collaboration d'une équipe de biologistes, dirigée par Bernard Dutrillaux et Florence Richard, du Muséum National d'Histoire Naturelle de Paris, et bénéficient donc d'un grand ensemble de données cytogénétiques originales.
- **Reconstitution de l'organisation des génomes de mammifères ancestraux.** L'utilité de cette reconstruction est d'étudier qualitativement et quantitativement les événements évolutifs importants (du point de vue de l'organisation des molécules d'ADN) qui ont marqué l'histoire évolutive des mammifères. Cela nécessite de construire des algorithmes qui minimisent les hypothèses *a priori* sur ces événements. Nous menons une large étude méthodologique des publications récentes sur le sujet, afin de proposer une méthode fondée sur des propriétés qui élimineront plusieurs biais ayant une influence significative sur les résultats.
- **Construction d'un modèle d'évolution par remaniements.** Nous rassemblons les données et méthodes qui pourront permettre d'affiner le seul modèle actuellement disponible, où les remaniements arrivent au hasard sur le génome, ce qui implique qu'ils ont peu de contraintes

structurelles et peu d'impacts fonctionnels. Ce modèle n'a jamais été réfuté de façon indiscutable chez les mammifères. Nous entreprenons une étude des contraintes et des conséquences fonctionnelles qui pourront - ou non - faire dévier les observations de ce qu'on attend sous un modèle aléatoire.

5. Modélisation de dépendances entre sites voisins dans l'évolution des séquences d'ADN

L'évolution de séquences nucléiques est généralement modélisée par des substitutions ponctuelles, sous l'hypothèse que les sites évoluent indépendamment les uns des autres. Cette hypothèse est maintenue pour des raisons mathématiques et n'a pas de fondement biologique : il est clair que de nombreux mécanismes évolutifs mettent en jeu des bases adjacentes. L'exemple le plus typique au niveau des mécanismes de substitution est celui du processus de méthylation-désamination des dinucléotides de type CpG, observé sur les séquences méthylées. Bérard, Gouéré et Piau ont montré que, dans une certaine classe de modèles incorporant des dépendances entre sites voisins (dont le mécanisme des CpG), les distributions à l'équilibre du modèle peuvent être écrites analytiquement. Nous prenons la suite de ce travail, et l'avons appliqué à certains modèles intéressants d'un point de vue biologique. Ce travail est principalement le fruit d'une collaboration avec Didier Piau (Université Joseph Fourier – Grenoble).

En particulier, nous avons commencé à analyser le taux de substitution de type CpG de deux organismes : l'homme et l'abeille. Le patron de méthylation du génome humain est largement étudié et est relié à de nombreuses fonctions, telles que le « genome imprinting », la mise sous silence de certains gènes, et la régulation de manière générale. Nous analysons la variation du taux de substitution de type CpG le long du génome humain, et étudions sa relation à la structure en mosaïque du génome humain. Une autre application est l'étude du génome très récemment séquencé de l'abeille. Ce génome révèle un patron de méthylation très intéressant, et différent de ceux déterminés sur *Drosophila melanogaster* et *Anopheles gambiae*. Nous étudions ce patron de méthylation à travers l'analyse de son contenu en dinucléotides et l'estimation des substitutions de type CpG.

b. Au niveau du réseau métabolique

1. Recherche de motifs dans les réseaux métaboliques

Afin de mettre en évidence des propriétés structurelles des réseaux métaboliques, plusieurs modèles peuvent être considérés. Dans un premier temps, nous avons choisi de modéliser un réseau métabolique par un graphe et donc de travailler à la détection de motif coloré dans un graphe non orienté dans lequel les nœuds correspondent aux réactions, les arêtes aux métabolites qui lient ces réactions et les couleurs aux classes de mécanismes réactionnels auxquelles les réactions appartiennent (EC number). Nous nous sommes appuyés sur ce modèle simple pour définir la notion de motif qui correspond à la notion de brique structurelle élémentaire d'un réseau métabolique.

Ce travail initial a soulevé de nouvelles questions ainsi que des idées d'amélioration. Ainsi, il est apparu que pour donner une signification statistique aux propriétés structurelles dégagées, un modèle de graphe aléatoire réaliste était nécessaire. Le développement d'un tel modèle est donc un des axes que nous avons poursuivi.

D'autre part, le modèle simple de graphe peut être suffisant pour concevoir et appliquer des méthodes telles que la recherche de motifs mais trouve ses limites lorsqu'on veut pousser plus loin l'analyse des résultats obtenus. Une extension naturelle consiste à représenter un réseau par un hypergraphe qui permet alors de capturer de manière plus réaliste les liens qui associent les métabolites entre eux, et dès lors de dégager des propriétés structurelles plus fines. De plus, la formulation de problèmes d'analyse structurelle des réseaux métaboliques sous forme d'hypergraphes semble pouvoir permettre un rapprochement intéressant avec d'autres méthodes

d'analyses des réseaux métaboliques basées sur une décomposition de la matrice stœchiométrique (modèles basés sur les contraintes). Nous avons ainsi commencé à nous intéresser au problème de la recherche de modes élémentaires dans un réseau métabolique et à des questions connexes. Ce travail est réalisé en collaboration avec le groupe d'Alberto Marchetti-Spaccamela de l'Université de Rome et celui de Leen Stougie de l'Institut Technologique de Eindhoven et du CWI (« Centrum voor Wiskunde en Informatica ») à Amsterdam.

Enfin, nous nous sommes intéressés au développement d'une nouvelle méthode de visualisation de graphes permettant de dessiner un réseau métabolique en prenant en compte sa structure en voies. Nous collaborons pour cela avec des spécialistes en visualisation de structures complexes : Fabien Jourdan de l'INRA de Toulouse, David Auber et Romain Bourqui du LABRI à Bordeaux.

2. Recherche de régularités au sein des réseaux métaboliques de bactéries en fonction de leur style de vie

Le fonctionnement métabolique d'un organisme est intimement lié à son environnement, c'est-à-dire aux composés disponibles autour de lui. Ces composés peuvent avoir une source minérale (cas de la plupart des plantes par exemple) mais également biologique. Dans ce cas, plusieurs phénomènes peuvent être à l'origine du flux de matière d'un organisme vers l'autre : certains impliquent la mort de l'organisme source (prédation ou décomposition), d'autres au contraire impliquent la survie des deux organismes : c'est le cas de la symbiose, définie ici comme une interaction durable entre deux organismes. C'est à ce dernier phénomène que nous nous sommes intéressés plus particulièrement. Ce travail se fait en collaboration avec Hubert Charles et le laboratoire « Biologie Fonctionnelle, Insectes et Interactions (BF2I) » de l'unité mixte INSA-INRA de Lyon.

Dans un premier temps, nous nous concentrons sur les bactéries endosymbiotes mutualistes et sur les bactéries endosymbiotes parasites vivant à l'intérieur de cellules spécialisées de certains insectes. Les mutualistes ont un effet positif sur la valeur sélective de leur hôte tandis que les parasites ont un effet négatif. L'objectif est de modéliser le réseau métabolique de plusieurs bactéries libres, parasites ou mutualistes, endosymbiotes ou non, afin de distinguer des régularités propres aux différents styles de vie.

La première étape a été de récupérer et de formater les données métaboliques de plusieurs organismes.

La seconde étape, en cours, est de comparer ces différents réseaux à plusieurs niveaux :

- Au niveau structurel, en comparant certaines mesures classiquement appliquées sur les réseaux métaboliques : connectivité, modularité, diamètre...
- Au niveau fonctionnel :
 - Comparaison des motifs de réactions grâce au programme Motus (voir la section « Livrables »)
 - Comparaison des capacités métaboliques propres à chaque organisme.
 - Comparaison de la robustesse des réseaux métaboliques.
- Au niveau fonctionnel et structurel : en développant une méthode d'alignement de réseaux métaboliques qui prenne en compte les similarités locales à la fois structurelles et fonctionnelles. Le développement de cette méthode se fait en parallèle avec le développement d'une méthode efficace de comparaison de réactions biochimiques (dans le cadre d'un stage de Master M2 Pro en bioinformatique de l'Université de Rouen).

La troisième étape est la création d'un site internet permettant l'accès aux bases de données créées et aux résultats des différentes mesures effectuées.

c. Au niveau du réseau génique

1. Modélisation par réseaux bayésiens de données de puce de différents types, différentes maladies et différents patients, et recherche de régularités dans de tels réseaux.

Nous nous intéressons au développement d'une modélisation probabiliste intégrant des jeux de données de différents types pour un même système biologique. Nous voulons pour cela utiliser et adapter la méthode des réseaux bayésiens. Afin de tester les modèles, nous collaborons avec la plateforme de pharmacogénomique Lyon-Est qui disposera pour plusieurs centaines de patients (notre système), des données cliniques, de traitement, et de biopuces (génomiques et transcriptomiques), le tout pour plusieurs maladies cancéreuses (leucémie, cancer du sein, myélome multiple et cancer du poumon). Les échantillons issus de biopsies sont en cours de réception par la plateforme de pharmacogénomique. Nous contribuons actuellement à ce projet en analysant des expériences de puces dans ces domaines (fournies par nos partenaires de la plateforme), ce qui permet aussi de nous familiariser avec certaines molécules et certains processus métaboliques qui intéressent ce champ de la pharmacologie.

Les données pour chacune des maladies seront très similaires : les études par les biopuces seront quasiment identiques dans leur forme et très comparables car faites au même endroit, par les mêmes personnes et en utilisant le même matériel. Elles seront aussi parfois semblables du point de vue des données cliniques (sérologies ou autres analyses biologiques par exemple). De plus, toutes ces maladies sont des cancers, ce qui sous-entend qu'elles peuvent avoir des caractéristiques communes (prolifération, croissance cellulaire, dérèglement de l'apoptose, etc.). L'idée est de comparer des réseaux bayésiens de plusieurs maladies, pour en extraire des motifs de relations causales communs, tenant compte des propriétés bayésiennes des réseaux dont ils sont issus. La comparaison de réseaux bayésiens n'existant pas dans la littérature, ceci est une approche originale, à la fois par la méthode employée et par son application.

2. Inférence de modules d'expression à partir de données hétérogènes

Nous avons continué à travailler sur le problème de l'inférence de modules d'expression à partir de données hétérogènes. Ces données sont : des données de puces, des informations de séquences (motifs potentiellement liés à des séquences régulatrices), et des empreintes laissées par l'évolution dans les génomes. L'objectif final est d'être capable d'inférer de façon automatique des modules cellulaires complets, c'est-à-dire, en fait, des réseaux géniques de petite taille ou, à plus long terme, des réseaux de grande taille. La première difficulté à laquelle nous avons été confrontés a été de trouver un moyen d'obtenir des mesures normalisées, et ainsi comparables de ces différentes informations.

d. Au niveau du réseau intégré

L'objectif est de proposer un cadre pour l'analyse (à court terme) et la modélisation (à moyen terme) des relations entre le génotype et le phénotype. Un premier axe des travaux en cours, menés sur la levure, porte sur l'étude des liens entre les réseaux de régulation et les réseaux métaboliques. Un objectif final étant la décomposition des réseaux biologiques en modules fonctionnels, un second axe de nos analyses se concentre sur l'analyse biostatistique des données d'expression en lien avec les réseaux métaboliques et de régulation.

Dans un premier temps, nous nous sommes focalisés sur une analyse biologique de la topologie et de la répétition des motifs dans les réseaux métaboliques, dont une définition a été proposée dans le cadre de ce projet (voir partie b1 ci-dessus), en corrélation avec l'expression et le mode de régulation des enzymes impliquées dans ces motifs. Les données d'expressions sont issues de plusieurs études de micro-arrays en accès libre sur la base de données SGD (*Saccharomyces Genome Database* <http://www.yeastgenome.org>) et les données de régulation sont fournies par Ana Teresa Freitas et Arlindo Oliveira du Technical University de Lisbonne, Portugal (récupérées à travers internet à l'URL : <http://www.yeastract.com>).

e. Relation entre dynamique des génomes et réseau métabolique

Nous avons initié une étude de l'impact de certains types de réarrangements sur le métabolisme d'un organisme. Dans un premier temps, nous nous intéressons au phénomène de duplication, en particulier aux duplications de génomes complets. Après avoir initialement considéré travailler sur la levure (dont la plupart des souches ont subi une duplication complète au cours de leur évolution), nous avons décidé de déplacer notre intérêt vers le génome de la paramécie, cela pour plusieurs raisons. La première est que d'autres personnes de notre laboratoire sont impliquées dans l'annotation et l'analyse fonctionnelle du génome de cet organisme. Cela nous permet ainsi d'avoir un accès direct et expert aux données. La seconde raison est que l'information pertinente aux trois duplications que le génome de la paramécie aurait subi est beaucoup plus claire et dénuée d'ambiguïté que l'information concernant l'unique duplication de la levure. En particulier, il est apparemment possible d'avoir une information relativement fiable sur quels gènes ont gardé leur copie suite à une des trois duplications complètes. De même, il est relativement facile de distinguer les copies issues de duplications complètes des copies issues de duplications localisées survenues de façon dynamique au cours de l'évolution. Cela est important car une des hypothèses est que l'impact des dupliqués du premier type sur le métabolisme est tout à fait différent de l'impact que peuvent avoir les duplications dynamiques. Notre approche vise à la fois effectuer une étude exploratoire de ces différents types d'impacts, et essayer de tester des hypothèses biologiques très précises. Deux exemples de telles hypothèses sont : 1. les gènes dupliqués codant pour des enzymes et ayant conservé leur copie seraient proches dans le réseau, 2. on s'attend à ce que ces gènes représentent préférentiellement les étapes initiales d'une voie de synthèse. Ces travaux sont effectués en collaboration avec Laurent Duret, Daniel Kahn et Jean-François Gout, chercheurs et doctorant respectivement dans l'équipe « Bioinformatique et Biologie Évolutive » de notre laboratoire.

D. Résultats obtenus pour la période concernée, dégager notamment les faits marquants

Nous ne présentons ici que les résultats nouveaux (publications, implantations, analyses bioinformatiques) obtenus depuis le dernier rapport d'activités (datant de Juin 2006).

a. Au niveau génomique

1. Amélioration et extension des algorithmes de recherche de régularités dans les séquences nucléotidiques

- **Filtrage de séquences en vue d'un alignement multiple** : Ces travaux ont donné lieu à une publication soumise dans la revue *Journal of Discrete Algorithms* (voir section « Livrables »). De nouvelles améliorations et extensions sont en cours d'analyse et de développement, et sont suffisamment bien avancées pour que les résultats puissent être soumis à la conférence WABI à la fin Avril 2007. Une implantation préliminaire qui ne comporte pas toutes les dernières améliorations est disponible dans le logiciel Ed'Nimbus (<http://igm.univ-mlv.fr/~peterlon/officiel/ednimbus/>). Une implantation de la version finale, appelée TUIUIU, sera rendue disponible au moment de la soumission de l'article à WABI.
- **Segmentation de séquences** : Un algorithme pour calculer, étant donné un modèle, la probabilité pour une séquence d'être structurée en un certain nombre de segments, a été soumis à la revue *Computational Statistics & Data Analysis*. Cette méthode est également en cours d'application aux données de cassures génomiques.

2. Recherche de régularités dans les régions de cassure

La méthode de détection des points de cassure a été appliquée sur les régions de cassures de l'homme par comparaison avec le génome de la souris, et sur des cassures simulées dans des régions synténiques. Le résultat principal de ce travail est une très nette différence d'efficacité de la méthode entre ces deux types de données, ce qui suggère une évolution différente et complexe des séquences ayant subi un réarrangement par rapport aux autres séquences du génome. Ce résultat et la méthode ayant permis de l'obtenir devront être soumis dans le courant du premier semestre 2007. En amont de cela, un travail important d'exploration bibliographique du problème avait été effectué et a été rédigé sous forme d'un article soumis à *Theoretical Computer Science* qui a, depuis le précédent rapport, été accepté.

3. Reconstitution de scénarios évolutifs par remaniements chromosomiques

- **Exploration des solutions du tri par inversions.** Nous avons développé un algorithme qui permet l'exploration de l'espace des solutions du tri par inversions, avec un gain de temps très important par rapport aux méthodes existantes. L'implantation de cet algorithme est disponible à l'adresse <http://www.geocities.com/mdvbraga/baobabLuna.html> et un article est soumis.
- **Reconstitution de l'histoire moléculaire de la différenciation sexuelle.** Plusieurs traits de l'histoire évolutive des chromosomes sexuels jusqu'ici inconnus sont apparus grâce à notre approche, par exemple un probable déplacement du centromère du chromosome Y ou plusieurs mécanismes de coupure des gènes du chromosome Y en plusieurs régions qui évoluent de façon indépendante. Ces résultats sont encore en cours d'exploitation en collaboration avec Gabriel Marais, chercheur dans l'équipe « Bioinformatique et Biologie Évolutive » de notre laboratoire.
- **Nouvelles approches du tri par inversion en présence d'éléments dupliqués.** Une première approche a été proposée qui correspond à la recherche d'une plus longue sous-séquence commune libre de répétitions. La complexité du problème a été étudiée montrant que ce dernier est APX-difficile (il est difficile d'obtenir même une approximation de la solution). Pour des instances pas trop grandes (comme cela est le cas, par exemple, des chromosomes X et Y), une formulation en termes de programmation linéaire par nombres entiers n'est cependant pas irréaliste en pratique. L'algorithme est actuellement en cours d'implantation par nos collaborateurs à l'Université Fédérale de Mato Grosso do Sul, Brésil. Une version préliminaire de la partie théorique du travail a été rédigée. Un premier article sur ce sujet devrait donc être soumis avant la fin de l'été 2007.
- **Phylogénie des primates.** Nous avons commencé à travailler sur des données cytogénétiques de cercopithèques fournies par nos collaborateurs du Muséum d'Histoire Naturelle de Paris, initialement en utilisant une méthode (SplitsTree, <http://bibiserv.techfak.uni-bielefeld.de/splits/>) disponible publiquement. Le résultat de cette application conduit à une représentation de l'évolution des cercopithèques (pour lesquels nous disposons de données) sous forme d'un réseau qui est très proche de celui obtenu par une analyse manuelle effectuée par nos collaborateurs, malgré le fait que SplitsTree ne permet pas de prendre en compte toutes les informations en notre possession. Par exemple, nous n'avons pas pu pour le moment modéliser les déplacements inférés de centromères. Comment le faire dans une méthode que nous serons amenés à développer est donc le prochain point à discuter avec nos collaborateurs.
- **Reconstitution de l'organisation des génomes de mammifères ancestraux.** Une étude méthodologique des publications récentes sur le sujet de la reconstitution de l'organisation de mammifères ancestraux est déjà bien avancée. Cette étude suggère clairement que les différences observées entre ces méthodes sont dues non au type de données utilisées (cytogénétiques versus génomiques) comme cela est suggéré ou affirmé dans de nombreux papiers récemment parus, mais plutôt à l'algorithme utilisé pour un même ensemble

d'organismes (le choix de ces organismes n'est, bien sûr, lui, pas du tout innocent vis-à-vis du résultat obtenu). Ce travail est en cours de rédaction.

4. Étude des liens entre éléments répétés de différents types entre eux ainsi qu'avec d'autres facteurs, dont le niveau d'expression des gènes

Les résultats initiaux que nous avons obtenus pointent vers un profil d'expression différent pour les transposons du chromosome X par rapport aux autres chromosomes de la drosophile, ainsi que vers une corrélation entre le nombre de copies dans une famille et le nombre de transposons dans une famille auxquels un EST a pu être associé. Ce dernier résultat est en contradiction avec ceux que l'on peut trouver dans la littérature. Les deux sont en cours d'exploitation.

5. Dépendances entre sites voisins dans les séquences d'ADN

- **Étude de la sur- et sous-représentation de dinucléotides.** Des statistiques non paramétriques pour l'étude de la sur- et sous-représentation de dinucléotides ont été développées et implantées dans la bibliothèque SeqinR du logiciel R (http://pbil.univ-lyon1.fr/software/SeqinR/seqinr1_angl.php), aussi disponible sur le CRAN à <http://cran.univ-lyon1.fr/src/contrib/Descriptions/seqinr.html>. Le détail de ces statistiques est publié dans l'article qui était sous presse lors du dernier rapport et est maintenant paru dans la revue *Molecular Biology and Evolution* (voir la section « Livrables »). La plupart des fonctions développées dans le cadre de ce projet ont été ajoutées à la version 1.0-6 de SeqinR (en collaboration avec D. Charif, J. Lobry et A. Necsulea de l'équipe « Bioinformatique et Biologie Évolutive » - <http://cran.univ-lyon1.fr/src/contrib/Descriptions/seqinr.html>).
- **Modélisation de dépendances entre sites voisins dans l'évolution des séquences d'ADN.** Il a été montré qu'à l'équilibre, il est possible d'estimer les taux de substitution agissant sur une séquence donnée, ce qui va permettre de nombreuses analyses ultérieures. Un programme est en cours d'implantation pour simuler l'évolution de séquences selon cette classe de modèles. Par ailleurs, les premiers résultats de notre analyse du patron de méthylation chez l'abeille paraissent confirmer que celui-ci est plus proche du patron de l'homme que de celui de la drosophile. En outre, il semblerait que seuls certains gènes soient méthylés, et donc possèdent un faible taux en CpG, et que CpG soit plutôt surreprésenté dans les séquences intergéniques, ce que nous ne savons pas encore expliquer.

b. Au niveau du réseau métabolique

1. Recherche de motifs dans les réseaux métaboliques

- L'interface web pour l'algorithme Motus (pour l'instant en accès restreint à <http://pbil.univ-lyon1.fr/software/motus> en utilisant le login : baobab et le mot de passe : baobab, et bientôt accessible publiquement) a été enrichie depuis le dernier rapport d'activités. En particulier, une méthode de simulation de graphes aléatoires a été implantée (par randomisation du graphe réel). Cette méthode est maintenant jointe à Motus et permet désormais de savoir si un motif est sur ou sous représenté dans le réseau. Il est déjà possible également de visualiser les occurrences d'un motif grâce à une applet java développée par notre collaborateur Fabien Jourdan et jointe également à Motus. Cette interface est en phase finale de test et sera très prochainement rendue disponible avec Motus.
- L'algorithme Motus lui-même a été interfacé avec Nauty, programme disponible publiquement qui permet de déterminer si deux graphes sont isomorphes. Le but est de pouvoir travailler avec des définitions alternatives de motifs, en particulier, d'introduire des contraintes topologiques *a posteriori*.

2. Recherche de régularités au sein des réseaux métaboliques de bactéries en fonction de leur style de vie

Pour la récupération et la mise en forme des données métaboliques de plusieurs organismes, nous avons utilisé : 1. les librairies Java et Perl initiées pendant le premier semestre du projet que nous avons continué à développer lors de ce second semestre et qui sont maintenant regroupées sous le nom provisoire de parseBioNet, et 2. les pathway Tools développés au sein de l'Université de Stanford qui permettent de reconstruire ses propres bases de données métaboliques dans le même format que celles déjà présentes dans BioCyc (<http://biocyc.org>) à partir de génomes annotés. L'ensemble des outils permettant de lire, de filtrer et d'analyser ces données est disponible sur la page <http://biomserv.univ-lyon1.fr/baobab/parsebionet>.

La récupération et la mise en forme des données a été réalisée pour une quinzaine d'organismes et quelques mesures sur les réseaux ont déjà été réalisées. Ces données sont d'ores et déjà disponibles, via la création d'un site internet permettant l'accès aux bases de données créées et aux résultats des différentes mesures effectuées. Cette base, appelée SymBioCyc est disponible dans sa version préliminaire à l'adresse <http://biomserv.univ-lyon1.fr/baobab/symbiocyc>, pour le moment en accès restreint (login : baobab, mot de passe : baobab).

La comparaison de motifs de réactions grâce au programme Motus (voir plus haut) est actuellement en cours. Les résultats initiaux montrent une différence au niveau des motifs inférés dans *Buchnera* par rapport à ceux identifiés dans *Escherichia coli* mais nous ne savons pas encore comment interpréter cette observation.

Afin de pouvoir effectuer la comparaison des capacités métaboliques propres à chaque organisme, nous avons implanté dans parseBioNet la méthode utilisée dans les travaux de Reinhart Heinrich et de ses collaborateurs (Handorf *et al.*, *J Mol Evol*, 61(4) :498-512, 2005), qui consiste à retourner l'ensemble des composés qu'il est possible de synthétiser dans un réseau métabolique, à partir d'un ensemble de composés d'entrée. Cette méthode nous permettra également, dans le cas des symbiotes, de mettre en évidence les fonctions métaboliques qui seraient assurées par l'hôte. Il est possible déjà grâce à elle de simuler l'élimination d'une ou de plusieurs réactions biochimiques et de tester ainsi la robustesse des réseaux.

c. Au niveau du réseau génique

Les analyses effectuées sur les puces de la plateforme de pharmacogénomique ont permis la détection de certains gènes (confirmée par PCR en temps réel, « RTQ-PCR ») qui réagissent avec un traitement antimitotique (s'opposant à l'accomplissement des mitoses, c'est-à-dire à la division et donc à la multiplication de certaines cellules) grâce à l'exploitation d'informations d'homologies homme-souris. Notre analyse de ces biopuces nous sera utile pour concevoir prochainement des réseaux bayésiens bien adaptés aux spécificités du problème et de chaque cancer traité.

E. Difficultés rencontrées et solutions de remplacement envisagées

La récolte des données est un des facteurs qui ralentissent souvent les études que nous menons. Le cas des endosymbiotes obligatoires est une illustration particulière de ce type de difficultés. Les endosymbiotes obligatoires ne sont pas cultivables à l'extérieur de leur hôte. Il existe donc très peu de sources expérimentales sur le fonctionnement métabolique de ces organismes. Nous ne disposons ici que de l'ensemble des réactions catalytiques, inféré à partir des annotations fonctionnelles des gènes de l'organisme. Cette reconstruction demande beaucoup de corrections manuelles mais implique également une confiance limitée dans les données générées. En outre, le plus souvent, nous ne disposons pas des reconstructions des réseaux métaboliques des hôtes, le génome de ceux-ci n'ayant pas été encore annoté.

Le manque total ou partiel de données n'empêche pas de développer des méthodes, par exemple de comparaison de réseaux dans le cas des endosymbiotes, mais cela peut rendre leur évaluation, et donc la révision en retour du modèle et de la méthode difficiles, voire dans certains cas impossibles pour le moment.

F. Livrables externes réalisés

a. Articles acceptés, sous presse ou soumis entre le 15 Juin 2006 et le 15 Janvier 2007

1. Articles dans des revues internationales

L. Palmeira, L. Guéguen, and J. Lobry. UV-targeted dinucleotides are not depleted in light-exposed Prokaryotic genomes. *Mol. Biol. Evol.*, 2006, 23(11):2214-2219 (**indiqué comme sous presse dans le rapport précédent**).

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 (**indiqué comme sous presse dans le rapport précédent**).

C. Lemaitre and M.-F. Sagot. A Small Trip in the Untranquil World of Genomes - A survey on the detection and analysis of genome rearrangement breakpoints. *Theoretical Computer Science*, 2007, sous presse (**indiqué comme soumis dans le rapport précédent**).

Y. Diekmann, M.-F. Sagot, and E. Tannier, Evolution under reversals: parsimony and conservation of common intervals, accepté à *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2006, sous presse.

L. Guéguen. Probability of a sequence segmentation under a Markovian model, soumis 2006.

R. Bourqui, L. Cottret, V. Lacroix, D. Auber, P. Mary, M.-F. Sagot, and F. Jourdan. Metabolic network visualization eliminating node redundancy and preserving metabolic pathways, soumis fin 2006.

C. Melodelima, L. Guéguen, C. Gautier, and D. Piau. A Markovian Approach for the Analysis of the Gene Structure, soumis fin 2006.

P. Peterlongo, J. Allali and M.-F. Sagot. The Gapped-Factor Tree, soumis fin 2006.

P. Peterlongo, N. Pisanti, F. Boyer, A. Pereira do Lago and M.-F. Sagot. Lossless filter for multiple repetitions, soumis fin 2006.

2. Chapitres de livres

M.-F. Sagot, E. Tannier, Sorting by reversals, soumis à *The Encyclopedia of Algorithms*, 2006.

3. Articles dans des conférences internationales à comité de lecture

R. Bourqui, D. Auber, V. Lacroix, and F. Jourdan. Metabolic network visualization using a constraint planar graph drawing. In *10th conf. on Information Visualization*, pages 489-496, 2006.

M. Braga, M.-F. Sagot, C. Scornavacca, E. Tannier, The solution space of sorting by reversals, soumis fin 2006.

F. Chierichetti, V. Lacroix, A. Marchetti-Spaccamela, M.-F. Sagot, L. Stougie. Modes and Cuts in Metabolic Networks: Complexity and Algorithms, soumis début 2007 (SPOR-Report 2007-01, Technische Universiteit Eindhoven, 2007, <http://www.win.tue.nl/bs/spor/2007-01.pdf>).

4. Articles dans des conférences nationales à comité de lecture

J.-J. Daudin, V. Lacroix, F. Picard, S. Robin, and M.-F. Sagot. Uncovering structure in biological networks. In *JOBIM06, Bordeaux, France*, 2006.

5. Communications orales dans des conférences, colloques ou rencontres

L. Palmeira : Models of DNA evolution with neighbor-dependent substitutions, *Otto Warburg Summer School and Workshop 2006* (September 2006 - Berlin, Germany).

L. Palmeira. Theoretical approaches for the study of dinucleotide content in genomes. *Theoretical Approaches for the Genome* (November 2006 - Annecy, France).

C. Gautier and E. Prestat. Some bioinformatics tools for studying biodiversity (November 2006, Universidade Federal do Mato Grosso do Sul, Brazil).

6. Logiciels et sites internet rendus publics ou enrichis depuis Juin 2006

- Baobab Luna (Marília Braga) : implantation d'un algorithme qui permet l'exploration de l'espace des solutions du tri par inversions, avec un gain de temps important par rapport aux méthodes existantes.
<http://www.geocities.com/mdvbraga/baobabLuna.html>
- PSbR (Yoan Diekmann en collaboration avec Eric Tannier et M.-F. Sagot) : implantation d'un algorithme pour trier des permutations (de marqueurs) signée qui fournit une solution optimale qui respecte les intervalles communs si une telle solution existe.
<http://biomserv.univ-lyon1.fr/~tannier/PSbR/>
- Motus (Vincent Lacroix en collaboration avec Odile Rogier du PRABI) : recherche et inférence de motifs dans les réseaux métaboliques. Disponible à certains chercheurs externes au projet, bientôt disponible à toute la communauté.
<http://pbil.univ-lyon1.fr/software/motus>
- Fonction supplémentaire dans SeqinR version 1.0-6 (Leonor Palmeira avec Delphine Charif, Jean Lobry, Anamaria Necsulea) : Incorporation de statistiques non paramétriques pour l'étude de la sur- et sous-représentation de dinucléotides dans la bibliothèque SeqinR du logiciel R.
<http://cran.univ-lyon1.fr/src/contrib/Descriptions/seqinr.html>
- Ed'Nimbus (Pierre Peterlongo) : filtre de séquences qui peut être utilisé pour trouver de longues répétitions dans les séquences nucléotidiques
<http://igm.univ-mlv.fr/~peterlon/officiel/ednimbus/>
- parseBioNet (Ludovic Cottret) : ensemble de bibliothèques Java et de routines Perl et R permettant de charger et d'analyser des données métaboliques
<http://biomserv.univ-lyon1.fr/baobab/parsebionet/>

7. Thèses démarrées, en cours et/ou soutenues en relation directe avec le projet

Les thèses suivantes sont en relation directe avec le projet :

Nom du doctorant	Date de début de la thèse	Section(s) décrivant sujet principal
------------------	---------------------------	--------------------------------------

Vicente Acuña	Septembre 2006 (Lyon)	b1 (modes élémentaires) et e
Marília Braga	Septembre 2005 (Lyon)	a4
Yves-Pol Daniélou	Septembre 2006 (Grenoble)	réseaux et réarrangements, sujet débutant
Marc Deloger	Septembre 2006 (Lyon)	a3
Vincent Lacroix	Septembre 2004 (Lyon)	b1 et e
Claire Lemaitre	Septembre 2005 (Lyon)	a2, a4 et e
Nuno Mendes	Janvier 2007 (Lyon)	motifs ARNs, sujet débutant
Leonor Palmeira	Septembre 2004 (Lyon)	a5

Par ailleurs, la thèse de Pierre Peterlongo co-dirigée par la coordonatrice du projet et portant sur le thème décrit en a1 a été soutenue en Septembre 2006 à l'Université de Marne-la-Vallée. Pierre Peterlongo est actuellement postdoc à l'IRISA de Rennes.

Enfin, Paulo Gustavo Soares da Fonseca, doctorant en thèse à l'Université Fédérale de Pernambuco, Brésil et séjournant en France depuis Septembre 2005 sur une bourse dite « sandwich » d'un an, va prolonger son séjour pour une période supplémentaire d'un an au moins. Il travaille sur le thème c (second point).

G. Autres commentaires

Les personnes suivantes ont été recrutées sur des fonds de l'ANR :

Ludovic Cottret, recruté comme ingénieur en Juin 2006 : sujet principal décrit dans les sections b2, e et a4 (point 4).

Emmanuel Prestat, recruté comme ingénieur en Décembre 2006 : sujet principal décrit dans la section c.

Patricia Thébault, sera recrutée comme postdoc en Février 2006 : sujet principal décrit dans la section d.

Par ailleurs, cinq personnes ont intégré ou sont revenues travailler dans BAOBAB sur des sujets de l'ANR :

Juliet Ansel, Master M2 Pro, sur les thèmes b1 et b2.

Christelle Melo de Lima, ATER Lyon II, sur les thèmes a1 et a2.

Vincent Navratil, postdoc INRA, sur la comparaison de réseaux.

Cinzia Pizzi, postdoc BAOBAB sur le thème b1.

Alexandra Popa, stage M2 sur le thème a1.

Eléments quantitatifs

H. Liste des réunions/séminaires/colloques organisés durant la période et des missions à l'étranger

a. Missions et séminaires en France

- Visite de Ludovic Cottret, Claire Lemaitre, Eric Tannier et Marie-France Sagot rendue le 2 Octobre 2006 au Muséum d'Histoire Naturelle de Paris pour travailler avec Bernard Dutrillaux et Florence Richard sur la phylogénie des primates et l'évolution en réseau.
- Visite de Vincent Lacroix et Marie-France Sagot rendue le 5 Octobre 2006 à l'INA-PG (Paris) pour travailler avec Stéphane Robin et Sophie Schbath sur le réalisme des modèles de graphes aléatoires et sur le comptage des motifs dans ces graphes.
- Séminaire invité au laboratoire d'Alain Arnéodo (Novembre 2006 – Lyon, France) : *"Theoretical approaches for the study of dinucleotide content in genomes"* (L. Palmeira).
- Groupe de Travail en Génomique Comparative, réunion en Octobre 2006 à Nantes (<http://biomserv.univ-lyon1.fr/~tannier/GTGC>).

b. Missions et séminaires à l'étranger

- Visite de Vicente Acuña, Paulo Fonseca, Claire Lemaitre, et Marie-France Sagot à Bertinoro, Italie, en Juin 2006 pour un colloque, « Biological Networks III: Modularity and Genome Evolution » (<http://www.inrialpes.fr/helix/people/sagot/bertinoro2006/bertinoro.html>) co-organisé par Marie-France Sagot.
- Visite de Vincent Lacroix et Marie-France Sagot rendue fin Juin 2006 à l'Université de Rome, Italie pour travailler avec Alberto Marchetti-Spaccamela et Leen Stougie sur les modes élémentaires dans les réseaux métaboliques.
- Visite de Christian Gautier, Vicente Acuña, Marília Braga, Ludovic Cottret, Paulo Fonseca, Claire Lemaitre, Pierre Peterlongo, Emmanuel Prestat et Marie-France Sagot rendue en Novembre et Décembre 2006 (deux semaines) à l'Université de São Paulo, Brésil et l'Université Fédérale de Mato Grosso do Sul, Campo Grande, Brésil pour travailler avec Alair Pereira do Lago, Yoshiko Wakabayashi, Cristina Gomes Fernandes, Fabio Martinez et Said Sadique Adi sur la recherche de régularités dans les séquences génomiques, les tris par inversions prenant en compte les duplications, les modes élémentaires dans les réseaux métaboliques et la comparaison de réseaux.
- Séminaire invité au laboratoire d'Axel Meyer (Octobre 2006 – Constance, Allemagne) : « *Dinucleotides in evolution: analysis and modelling* » (L. Palmeira).
- Séminaire invité au laboratoire d'Arndt von Haeseler (Novembre 2006 – Vienne, Autriche) : « *Neighboring-site dependencies in genome evolution: analysis and modelling* » (L. Palmeira).
- Séminaire invité au laboratoire de Michael Laessig (Décembre 2006 – Cologne, Allemagne) : « *Theoretical approaches for the study of dinucleotide content in genomes* » (L. Palmeira).
- Séminaire invité en vue d'une collaboration avec Gianpaolo Oriolo, de l'Université de Rome Tor-Vergata (Septembre 2006) : « *Conservation and Rearrangements in Genomes* ».
- Séminaire invité à l'Université Fédérale de Mato Grosso do Sul (Novembre 2006, Mato Grosso do Sul, Brésil) : « *Some bioinformatics tools for studying biodiversity* » (C. Gautier et E. Prestat).
- Séminaire invité à l'Université de São Paulo (Novembre 2006, Brésil) : « *Analysing gene expression* » (C. Gautier et P. Gustavo).

J. *Le cas échéant et pour les programmes thématiques*, préciser les travaux réalisés par les partenaires étrangers associés au projet sans aide de l'ANR

K. Liste des personnels recrutés en CDD par des établissements publics dans le cadre du projet sur l'aide allouée par l'ANR

Nom	Prénom	Qualifications	Date de recrutement	Durée du contrat (en mois)
Cottret	Ludovic	Ingénieur d'étude	Juin 2006	2 ans
Prestat	Emmanuel	Ingénieur d'étude	Décembre 2006	2 ans
Thébault	Patricia	Postdoc	Février 2006	9 mois

L. *Le cas échéant*, indiquer les différents types d'aides complémentaires obtenues grâce à ce projet.

M. *Le cas échéant*, modalités d'utilisation du complément de financement « pôles de compétitivité »

N. CADRE RESERVE AU COORDONNATEUR DU PROJET

Dans le dernier rapport d'activités concernant le projet, nous avons noté comme une de ses difficultés principales, le fait qu'il exige d'avancer sur plusieurs fronts à la fois dans la recherche de régularités de natures diverses, au niveau à la fois du génome et du réseau d'interactions, afin d'essayer de répondre à la question très générale « est-ce que la structure des systèmes vivants est simple, ou simplifiable en des principes généraux, ou bien est-ce que la vie est faite essentiellement d'exceptions? ». Après avoir porté notre attention au départ sur des sujets relativement proches de résultats que nous avons déjà et que nous essayions d'étendre, nous avons dans les six derniers mois considérablement élargi notre champ d'action, en partie aussi grâce à l'arrivée de nouveaux membres de BAOBAB-HELIX, et donc participants du projet. Les liens entre les différents thèmes concernés par ce projet nous paraissent de plus en plus clairement importants à explorer, surtout ceux entre dynamique génomique et réseau métabolique. Nous avons également fortement renforcé les collaborations, nationales et internationales, que nous avions. En particulier, outre les collaborations indiquées dans le rapport précédent qui ont été rendues plus solides au cours de ces derniers mois, nous avons pu établir des collaborations nouvelles pour le volet « dynamique des génomes » du projet, ce qui nous faisait encore un peu défaut. Ces collaborations sont dans trois cas internes au laboratoire, et concernent des biologistes dont une expérimentaliste (Cristina Vieira). Nous espérons ainsi pouvoir augmenter également l'impact multidisciplinaire de nos travaux. La quatrième nouvelle collaboration dans le cadre de la dynamique des génomes nous donne accès à une collection de données (cytogénétiques) et à une expertise extrêmement précieuses. À l'heure actuelle, nous disposons donc déjà, à la fin de cette première année, d'un ensemble de relations de recherche avec des informaticiens, des mathématiciens et des biologistes qui est bien équilibré, et qui enrichit considérablement notre démarche.

Concernant les jeux de données, sur lesquels nous avons concentré notre attention lors des six premiers mois, nous rencontrons encore beaucoup de difficultés dans certains cas pour en obtenir qui soient fiables, ou suffisamment complètes. Nous avons pu néanmoins dans plusieurs cas (données évolutives, réseaux métaboliques d'organismes divers) établir des jeux assez importants que nous avons rendus disponibles publiquement lorsque ces données ont été rassemblées par nos propres moyens (SymBioCyc).

Nous avons commencé également à rendre les méthodes que nous développons disponibles, à travers des interfaces, soit à toute la communauté (Baobab Luna, Ed'Nimbus, parseBioNet, PSbR, fonction supplémentaire dans SeqInR), soit initialement à certains chercheurs uniquement avant une diffusion plus large (Motus). Une difficulté que nous rencontrons cependant encore, malgré les moyens que nous y avons investi, est à implanter suffisamment rapidement les algorithmes que nous développons, pour le moment, à une vitesse plus rapide que nous ne réussissons à les transcrire sous forme de code pour pouvoir les tester et, très possiblement, revoir, affiner, adapter les modèles et approches initiales. Certainement la récolte des données et le développement logiciel exigent considérablement de temps. Du côté positif, cela signifie aussi que, pour le moment, les idées algorithmiques nouvelles ne nous font pas défaut, que ce soit par rapport au volet « dynamique des génomes » ou à celui « réseaux d'interactions ».

CADRE RESERVE A l'USAR

Nom du coordinateur scientifique de l'USAR :

Date :