IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Lyon

Université Claude Bernard (Lyon 1)

Activity Report 2012

Project-Team BAMBOO

An algorithmic view on genomes, cells, and environments

IN COLLABORATION WITH: Laboratoire de Biométrie et Bologie Evolutive (LBBE)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Computational Biology and Bioinformatics

Table of contents

Project-Team BAMBOO

Keywords: Computational Biology, Genomics, Systems Biology, Analysis Of Algorithms, Evolution, Network Modeling

Creation of the Project-Team: January 01, 2012 .

1. Members

Research Scientists

Yvan Rahbé [Senior Researcher Inra (DR), HdR]
Marie-France Sagot [Team leader, Senior Researcher Inria (DR), HdR]
Fabrice Vavre [Senior Researcher (DR) Cnrs, HdR]
Alain Viari [Senior Researcher Inria (DR)]

Faculty Members

Etienne Birmelé [Associate Professor, Univ. Évry, délégation Inria for one year from Sept. 1st, 2011 to August 31st, 2012]
Hubert Charles [Full Professor (from Sept. 2012), Insa-Lyon, HdR]
Christian Gautier [Professor, University Claude Bernard, HdR]
Vincent Lacroix [Associate Professor, University Claude Bernard]
Cristina Vieira [Professor, IUF, University Claude Bernard, HdR]

Engineer

Alice Julien-Lafferière [ADT Inria, started December 1st, 2012]

PhD Students

Beatrice Donati [scholarship ERC, supervisors: Marie-France Sagot and Pierluigi Crescenzi]
Pierre-Antoine Farnier [scholarship ERC, supervisors: Marie-France Sagot, Laurence Mouton and Fabrice Vavre]
Mariana Ferrarini [scholarship ERC, supervisors: Marie-France Sagot and Arnaldo Zaha]
Susan Higashi [scholarship ERC, supervisors: Christian Gautier, Stefano Colella and Marie-France Sagot]
Cecília Coimbra Klein [scholarship ERC, supervisors: Marie-France Sagot and Ana Tereza Vasconcelos]
Paulo Vieira Milreu [scholarship ANR, supervisors: Vincent Lacroix, Christian Gautier and Marie-France Sagot, PhD defended December 19, 2012, will stay on until April 30, 2013]
Gustavo Sacomoto [scholarship ERC, supervisors: Marie-France Sagot, Vincent Lacroix and Pierluigi Crescenzi]
Patrícia Simões [scholarship FCT, Ministry of Science and Technology, Portugal, supervisors: Marie-France Sagot and Sylvain Charlat, PhD defended in March 14, 2012]
Martin Wannagat [scholarship ERC, supervisors: Marie-France Sagot, Leen Stougie and Alberto Marchetti-Spaccamela, PhD will start Jan. 2, 2013]

Post-Doctoral Fellows

Lilia Brinza [CDD ERC, started October 1st, 2012]
Christian Baudet [CDD ANR]
Matteo Brilli [CDD ERC, until September 30, 2012]
Blerina Sinaimeri [CDD ERC, started February 1st, 2012]

Visiting Scientists

Paulo Alvarez [Master student, Univ. of Brasilia, Brasilia, Brazil, visit 3 months, supervisors: Maria Emilia Walter Telles, Cecília C. Klein and Marie-France Sagot]
Rui Ferreira [PhD student (Supervisor: Roberto Grossi), University of Pisa, Italy, various visit of 1 week]
Andrea Marino [PhD student (Supervisor: Pierluigi Crescenzi), University of Florence, Italy, visit of 3 months plus various visits of 1-2 weeks]

Maria Cristina Motta [Univ. Federal Rio de Janeiro, Rio de Janeiro, Brazil, visit 10 days]
Susana Vinga [Professor, INESC-ID, IST Lisbon, Portugal, visit of 1 week]
Arnaldo Zaha [Univ. Federal Rio Grande do Sul, Porto Alegre, Brazil, visit 10 days]

Administrative Assistant

Florence Bouheddi [Secretary (SAR) Inria]

Others

Pierluigi Crescenzi [Professor, University of Florence, Italy, external collaborator]
Manolo Gouy [Senior Researcher (DR) Cnrs, external collaborator, HdR]
Roberto Grossi [Professor, University of Pisa, Italy, external collaborator]
Alberto Marchetti-Spaccamela [Professor, University La Sapienza, Rome, Italy, external collaborator]
Anne Morgat [Scientist, Swiss Institute of Bioinformatics, external collaborator]
Nadia Pisanti [Associate Professor, University of Pisa, Italy, external collaborator]
Leen Stougie [Free University Amsterdam and CWI, Amsterdam, the Netherlands, external collaborator]
Ana Tereza Vasconcelos [Lab Nacional de Computação Científica, Petrópolis, Brazil, external collaborator]

2. Overall Objectives

2.1. Highlights of the Year

One highlight, both scientific and organisational, for 2012 concerns the setting up of a CNRS-UCBL-Inria Laboratoire International Associé (LIA) with the Laboratório Nacional de Computação Científica (LNCC), Petrópolis, Brazil. The LIA has for acronym LIRIO ("Laboratoire International de Recherche en BInformatique") and is coordinated by Ana Tereza Vasconcelos from the LNCC and Marie-France Sagot from BAMBOO. The LIA is created for 4 years, renewable once. A preliminary web page for the LIA LIRIO is available at this address: https://team.inria.fr/bamboo/en/cnrs-lia-laboratoire-international-associe-lirio/.

3. Scientific Foundations

3.1. Formal methods

The study of symbiosis and of biological interactions more in general is the motivation for the work conducted within BAMBOO, but runs in parallel with another important objective. This concerns to (re)visit classical combinatorial (mainly counting / enumerating) and algorithmic problems on strings and (hyper)graphs, and to explore the new variants / original combinatorial and algorithmic problems that are raised by the main areas of application of this project. As the objectives of these formal methods are motivated by biological questions, they are briefly described together with those questions in the next section.

3.2. Symbiosis

The study we propose to do on symbiosis decomposes into four main parts - (1) genetic dialog, (2) metabolic dialog, (3) symbiotic dialog and genome evolution, and (4) symbiotic dynamics - that are however strongly interrelated, and the study of such interrelations will represent an important part of our work. Another biological objective, larger and which we hope within the ERC project SISYPHE just to sketch for a longer term investigation, will aim at getting at a better grasp of species identity and of a number of identity-related concepts. We now briefly indicate the main points that have started been investigated or should be investigated in the next five years.

Genetic dialog

We plan to study the genetic dialog at the regulation level between symbiont and host by addressing the following mathematical and algorithmic issues:

1. model and identify all small RNAs from the bacterium and the host which may be involved in the genetic dialog between the two, and model/identify the targets of such small RNAs;
2. infer selected parts of the regulatory network of both symbiont and host (this will enable to treat the next point) using all available information;
3. explore at both the computational and experimental levels the complementarity of the two networks, and revisit at a network level the question of a regulatory response of the symbiont to its host's demand;
4. compare the complementarities observed between pairs of networks (the host's and the symbiont's); such complementarities will presumably vary with the different types of host-symbiont relationships considered, and of course with the information the networks model (structural or dynamic); Along the way, it may become important at some point to address also the issue of transposable elements (abbreviated into TEs, that are genes which can jump spontaneously from one site to another in a genome following or not a duplication event). It is increasingly believed that TEs play a role in the regulation of the expression of the genes in eukaryotic genomes. The same role in symbionts, and in the host-symbiont dialog has been less or not explored. This requires to address the following additional task:
5. accurately and systematically detect all transposable elements (*i.e.* genes which can jump spontaneously from one site to another in a genome following or not a duplication event) and assess their implication in their own regulation and that of their host genome (the new sequencing technologies should facilitate this task as well as other data expression analyses, if we are able to master the computational problem of analysing the flow of data they generate: fragment indexing, mapping and assembly);
6. where possible, obtain data enabling to infer the PPI (Protein-Protein Interaction) for hosts and symbionts, and at the host-symbiont interface and analyse the PPI networks obtained and how they interact.

Initial algorithmic and statistical approaches for the first two items above are under way and are sustained by a well-established expertise of the team on sequence and microarray bioinformatic analysis. Both problems are however notoriously hard because of the high level of missing data and noise, and of our relative lack of knowledge of what could be the key elements of genetic regulation, such as small and micro RNAs.

We also plan to establish the complete repertoire of transcription factors of the interacting partners (with possible exchanges between them) at both the computational and experimental levels. Comparative biology (search by sequence homology of known regulators), 3D-structural modelling of putative domains interacting with the DNA molecule, regulatory domains conserved in the upstream region of coding DNA are among classical and routinely used methods to search for putative regulatory proteins and elements in the genomes. Experimentaly, the BiaCore (using the surface plasmon resonance principle) and ChIP-Seq (using chromatin precipitation coupled with high-throughput sequencing from Solexa) techniques offer powerful tools to capture all the protein-DNA interactions corresponding to a specific putative regulator. However, these techniques have not been evaluated in the context of interacting partners making this task an interesting challenge.

Metabolic dialog

Our main plan for this part, where we have already many results, some obtained this last year, is to:

1. continue with and improve our work on reconstructing the metabolic networks of organisms with sequenced genomes, taking in particular care to cover as much as possible the different types of hosts and symbionts in interaction;
2. refine the network reconstructions by using flux balance analysis which will in turn require addressing the next item;
3. improve our capacity to efficiently compute fluxes and do flux balance analysis; current algorithms can handle only relatively small networks;

4. analyse and compare the networks in terms of their general structural, quantitative and dynamic characteristics;
5. develop models and algorithms to compare different types of metabolic interfaces which will imply being able, by a joint computational and experimental approach, to determine what is transported across interacting metabolisms;
6. define what would be a good null hypothesis to test the statistical significance, and therefore possible biological relevance of the characteristics observed when analysing or comparing (random network problem, a mostly open issue despite the various models available);
7. use the results from item 5, that is indications on the precursors of a bacterial metabolism that are key players in the dialog with the metabolism of the host, to revisit the genetic regulation dialog between symbiont and host.

Computational results from the last item will be complemented with experiments to help understand what is transported from the host to the symbiont and how what is transported may be related with the genetic dialog between the two organisms (items 5 and 6).

Great care will also be taken in all cases (metabolism- or regulation-only, or both together) to consider the situations, rather common, where more than two partners are involved in a symbiosis, that is when there are secondary symbionts of a same host.

The first five items above have started being computationally explored by our team, as has the last item including experimentally. Some algorithmic proofs-of-concept, notably as concerns structural, flux, precursor and chemical organisation studies (see some of the publications of the last year and this one), have been established but much more work is necessary. The main difficulties with items 3 and 4 are of two sorts. The first one is a modelling issue: what are the best models for analysing and comparing two or more networks? This will greatly depend on the biological question put, whether evolutionary or functional, structural or physiologic, besides being a choice that should be motivated by the extent and quality of the data available. The second sort of difficulty ,which also applies to other items notably (item 2), is computational. Most of the problems related with analysing and specially comparing are known to be hard but many issues remain open. The question of a good random model (item 6) is also largely open.

Symbiotic dialog and genome evolution
Genomes are not static. Genes may get duplicated, sometimes the duplication affects the whole genome, or genes can transpose, while whole genomic segments can be reversed or deleted. Deletions are indeed one of the most common events observed for some symbionts. Genetic material may also be transferred across sub-species or species (lateral transfer), thus leading to the insertion of new elements in a genome. Finally, parts of a genome may be amplified through, for instance, slippage during DNA replication resulting in the multiplication of the copies of a repeat that appear tandemly arrayed along a genome. Tandem repeats, and other types of short or long repetitions are also believed to play a role in the generation of new genomic rearrangements although whether they are always the cause or consequence of the genome break and gene order change remains a disputed issue.

Work on this part will involve the following items:

1. extend the theoretical work done in the past years (rearrangement distance, rearrangement scenarios enumeration) to deal with different types of rearrangements and explore various types of biological constraints;
2. develop good random models (a largely open question despite some initial work in the area) for rearrangement distances and scenarios under a certain model, i.e. type of rearrangement operation(s) and of constraint(s), to assess whether the distances / scenarios observed have statistically notable characteristics;
3. extensively use the method(s) developed to investigate the rearrangement histories for the families of symbionts whose genomes have been sequenced and sufficiently annotated;

4. investigate the correlation of such histories with the repeats content and distribution along the genomes;
5. use the results of the above analyses together with a natural selection criterion to revisit the optimality model of rearrangement dynamics;
6. extend such model to deal with eukaryotic (multi-chromosomal) genomes;
7. at the interface host-symbiont, investigate the relation between the rearrangement histories in hosts and symbionts and the various types of symbiotic relationships observed in nature;
8. map such histories and their relation with the genetic and metabolic networks of hosts and symbionts, separately and at the interface;
9. develop methods to identify and quantify rearrangement events from NGS data.

Symbiotic dynamics
In order to understand the evolutionary consequences of symbiotic relations and their long term trajectories, one should be able to assess how tight is the association between symbionts and their hosts.

The main questions we would like to address are:

1. how often are symbionts horizontally transferred among branches of the host phylogenetic tree?
2. how long do parasites persist inside their host following the invasion of a new lineage?
3. what processes underlie this dynamic gain/loss equilibrium?

Mathematically, these questions have been traditionally addressed by co-phylogenetic methods, that is by comparing the evolutionary histories of hosts and parasites as represented in phylogenetic trees.

Currently available co-phylogenetic algorithms present various types of limitations as suggested in recent surveys. This may seriously compromise their interpretation with a view to understanding the evolutionary dynamics of parasites in communities. A few examples of limitations are the (often wrong) assumption made that the same rates of loss and gain of parasite infection apply for every host taxonomic group, and the fact that the possibility of multi-infections is not considered. In the latter case, exchange of genetic material between different parasites of a same host could further scramble the co-evolutionary signal. We therefore plan to:

1. better formalise the problem and the different simplifications that could be made, or inversely, should be avoided in the co-phylogeny studies; examples of the latter are the possibility of multi-infections, differential rate of loss and gain of infection depending on the host taxonomic group and geographic distance between hosts, etc., and propose better co-phylogenetic algorithms;
2. elaborate series of simulated data that will enable to (i) get a better grasp of the effect of the different parameters of the problem and, more pratically, (ii) evaluate the performance of the method(s) that exist or are proposed (see next item);
3. apply the new methods to address the three questions above.

3.3. Intracellular interactions

The interactions of a symbiont with others sharing a same host, or with a symbiont and the cell of its host in the case of endosymbionts (organism that lives within the body or cells of another) are special, perhaps more complex cases of intracellular interactions that may concern different types of genetic elements, from organelles to whole chromosomes. The spatial arrangement of those genetic elements inside the nucleus of a cell is believed to be important both for gene expression and exchanges of genetic material between chromosomes. This question goes beyond the symbiosis one and has been investigated in the team in the last few years. Work on this will continue in future and concern developing algorithmic and statistical methods to analyse the interaction data that is starting to become available, in particular using NGS methods, in order to arrive at a better understanding of transcription, regulation both classical and epigenetic (inherited changes in phenotype or gene expression caused by mechanisms other than changes in the underlying DNA sequence), alternative splicing and trans-splicing phenomena, as well as study the possible interactions between an eukaryotic cell and its organelles or other cytoplasmic structures.

4. Application Domains

4.1. Biology with a focus on symbiosis

The main area of application of BAMBOO is biology, with a special focus on symbiosis (ERC project) and on intracellular interactions.

5. Software

5.1. AcypiCyc

Participants: Hubert Charles [EPI], Patrice Baa Puyoule [Contact, Patrice.Baa-Puyoulet@lyon.inra.fr], Stefano Colella [Contact, stefano.colella@lyon.inra.fr], Ludovic Cottret, Marie-France Sagot [EPI], Augusto Vellozo [Contact, augusto@cycadsys.org], Amélie Véron.

Database of the metabolic network of *Acyrthosiphon pisum*.
http://acypicyc.cycadsys.org/

5.2. AlViE

Participants: Pierluigi Crescenzi [Contact, pierluigi.crescenzi@unifi.it, ext. member EPI], Giorgio Gambosi, Roberto Grossi, Carlo Nocentini, Tommaso Papini, Walter Verdese.

ALVIE is a post-mortem algorithm visualization Java environment, which is based on the interesting event paradigm. The current distribution of ALVIE includes more than forty visualizations. Almost all visualizations include the representation of the corresponding algorithm C-like pseudo-code. The ALVIE distribution allows a programmer to develop new algorithms with their corresponding visualization: the included Java class library, indeed, makes the creation of a visualization quite an easy task (once the interesting events have been identified).
http://piluc.dsi.unifi.it/alvie/

5.3. Cassis

Participants: Christian Baudet [EPI, Contact, christian.baudet@univ-lyon1.fr], Christian Gautier [EPI], Claire Lemaitre [Contact, claire.lemaitre@inria.fr], Marie-France Sagot [EPI], Eric Tannier.

Algorithm for precisely detecting genomic rearrangement breakpoints.
http://pbil.univ-lyon1.fr/software/Cassis/

5.4. Cravela

Participants: Ana Teresa Freitas, Nuno Mendes [Contact, ndm@kdbio.inesc-id.pt], Marie-France Sagot [EPI, Contact, marie-france.sagot@inria.fr].

Framework for the identification and evaluation of miRNA precursors (finished), targets (in development) and regulatory modules(in development).
http://www.cravela.org/

5.5. C3P

Participants: Frédéric Boyer, Anne Morgat [EPI, ext. member], Alain Viari [EPI, Contact, alain.viari@inria.fr].

Merging two or more graphs representing biological data (e.g. pathways, ...).
http://www.inrialpes.fr/helix/people/viari/cccpart

5.6. CycADS

Participants: Hubert Charles [EPI], Patrice Baa Puyoule [Contact, Patrice.Baa-Puyoulet@lyon.inra.fr], Stefano Colella [Contact, stefano.colella@lyon.inra.fr], Ludovic Cottret, Marie-France Sagot [EPI], Augusto Vellozo [Contact, augusto@cycadsys.org].

Cyc annotation database system.
http://www.cycadsys.org/

5.7. Gobbolino

Participants: Vicente Acuña [EPI], Etienne Birmelé [EPI, délégation], Ludovic Cottret, Pierluigi Crescenzi, Fabien Jourdan, Vincent Lacroix, Alberto Marchetti-Spaccamela [EPI, ext. member], Andrea Marino, Paulo Vieira Milreu [EPI, Contact, pvmilreu@gmail.com], Marie-France Sagot [EPI], Leen Stougie [EPI, ext. member].

Algorithm to enumerate all metabolic stories in a metabolic network given a set of metabolites of interest.
Code available on request.

5.8. kisSNP

Participants: Vincent Lacroix [EPI], Pierre Peterlongo [Contact, pierre.peterlongo@inria.fr], Nadia Pisanti, Marie-France Sagot [EPI], Nicolas Schnel.

Algorithm for identifying SNPs without a reference genome by comparing raw reads.
http://alcovna.genouest.org/kissnp/

5.9. kisSplice

Participants: Rayan Chikhi, Janice Kielbassa [EPI], Vincent Lacroix [Contact, EPI], Pierre Peterlongo [Contact, pierre.peterlongo@inria.fr], Gustavo Sacomoto [EPI], Marie-France Sagot [EPI], Raluca Uricaru.

Algorithm for de-novo calling alternative splicing events from RNA-seq data.
http://alcovna.genouest.org/kissplice/

5.10. LASAGNE

Participants: Pierluigi Crescenzi [Contact, pierluigi.crescenzi@unifi.it, ext. member EPI], Roberto Grossi, Michel Habib, Claudio Imbrenda, Leonardo Lanzi, Andrea Marino.

LASAGNE is a Java application which allows the user to compute distance measures on graphs by making a clever use either of the breadth-first search or of the Dijkstra algorithm. In particular, the current version of LASAGNE can compute the exact value of the diameter of a graph: the graph can be directed or undirected and it can be weighted or unweighted. Moreover, LASAGNE can compute an approximation of the distance distribution of an undirected unweighted graph. These two features are integrated within a graphical user interface along with other features, such as computing the maximum (strongly) connected component of a graph.
http://amici.dsi.unifi.it/lasagne/

5.11. MetExplore

Participants: Michael Barrett, Hubert Charles [EPI], Ludovic Cottret [Contact, Ludovic.Cottret@toulouse.inra.fr], Fabien Jourdan, Marie-France Sagot [EPI], Florence Vinson, David Wildridge.

Web server to link metabolomic experiments and genome-scale metabolic networks.
http://metexplore.toulouse.inra.fr/metexplore/

5.12. Migal

Participants: Julien Allali [Contact, julien.allali@labri.fr], Marie-France Sagot [EPI, Contact, marie-france.sagot@inria.fr].

RNA, tree comparison
Algorithm for comparing RNA structures.
http://www-igm.univ-mlv.fr/~allali/logiciels/index.en.php

5.13. MotusWEB

Participants: Ludovic Cottret, Fabien Jourdan, Vincent Lacroix [EPI, Contact, vincent.lacroix@univ-lyon1.fr], Odile Rogier, Marie-France Sagot [EPI].

Algorithm for searching and inferring coloured motifs in metabolic networks (web-based version - offers different functionalities from the downloadable version).
http://pbil.univ-lyon1.fr/software/motus_web/

5.14. Motus

Participants: Ludovic Cottret, Fabien Jourdan, Vincent Lacroix [EPI, Contact, vincent.lacroix@univ-lyon1.fr], Odile Rogier, Marie-France Sagot [EPI].

Algorithm for searching and inferring coloured motifs in undirected graphs (downloable version - offers different functionalities from the web-based version).
http://pbil.univ-lyon1.fr/software/motus/

5.15. PhEVER

Participants: Christian Gautier [EPI], Vincent Lotteau, Leonor Palmeira [Contact, mlpalmeira@ulg.ac.be], Chantal Rabourdin-Combe, Simon Penel.

Database of homologous gene families built from the complete genomes of all available viruses, prokaryotes and eukaryotes and aimed at the detection of virus/virus and virus/host lateral gene transfers.
http://pbil.univ-lyon1.fr/databases/phever/

5.16. PepLine

Participants: Jérôme Garin, Alain Viari [EPI, Contact, alain.viari@inria.fr].

Pipeline for the high-throughput analysis of proteomic data.
http://www.grenoble.prabi.fr/protehome/software/pepline

5.17. Pitufo and family

Participants: Vicente Acuña [EPI], Ludovic Cottret [Contact, Ludovic.Cottret@toulouse.inra.fr], Alberto Marchetti-Spaccamela [EPI, ext. member], Paulo Vieira Milreu [EPI, Contact, pvmilreu@gmail.com], Marie-France Sagot [EPI], Leen Stougie [EPI, ext. member], Fabio Viduani-Martinez.

Algorithms to enumerate all minimal sets of precursors of target compounds in a metabolic network.
http://sites.google.com/site/pitufosoftware/

5.18. PSbR

Participants: Yoan Diekmann, Marie-France Sagot [EPI, Contact, marie-france.sagot@inria.fr], Eric Tannier.

Algorithm for testing the evolution and conservation of common clusters of genes.
http://pbil.univ-lyon1.fr/members/sagot/htdocs/team/software/PSbR/

5.19. Repseek

Participants: Guillaume Achaz [Contact, achaz@abi.snv.jussieu.fr], Eric Coissac, Alain Viari [EPI].

Finding approximate repeats in large DNA sequences.
http://wwwabi.snv.jussieu.fr/public/RepSeek/

5.20. Smile

Participants: Laurent Marsan, Marie-France Sagot [EPI, Contact, marie-france.sagot@inria.fr].

Motif inference algorithm taking as input a set of biological sequences.

5.21. Tuiuiu

Participants: Alair Pereira do Lago, Pierre Peterlongo [Contact, pierre.peterlongo@inria.fr], Nadia Pisanti, Gustavo Sacomoto [EPI], Marie-France Sagot [EPI].

Multiple repeat search filter with edit distance.
http://mobyle.genouest.org/cgi-bin/Mobyle/portal.py?form=tuiuiu

5.22. UniPathway

Participants: Eric Coissac, Anne Morgat [EPI, Contact, anne.morgat@inria.fr], Alain Viari [EPI].

Database of manually curated pathways developed with the Swiss-Prot group.
http://www.unipathway.org

6. New Results

6.1. Partial Enumeration of Traces

Traditional algorithms to solve the problem of sorting by signed reversals output just one optimal solution while the space of all optimal solutions can be huge. A so-called trace represents a group of solutions which share the same set of reversals that must be applied to sort the original permutation following a partial ordering. By using traces, we therefore can represent the set of optimal solutions in a more compact way. Algorithms for enumerating the complete set of traces of solutions were developed. However, due to their exponential complexity, their practical use is limited to small permutations. A partial enumeration of traces is a sampling of the complete set of traces and can be an alternative for the study of distinct evolutionary scenarios of big permutations. Ideally, the sampling should be done uniformly from the space of all optimal solutions. This is however conjectured to be #P-complete.

We proposed and evaluated three algorithms for producing a sampling of the complete set of traces that instead can be shown in practice to preserve some of the characteristics of the space of all solutions [7]. We analysed the distribution of the enumerated traces with respect to their height and average reversal length.

6.2. De-novo calling alternative splicing events from RNA-seq data

We addressed the problem of identifying and quantifying polymorphisms in RNA-seq data when no reference genome is available, without assembling the full transcripts. Based on the fundamental idea that each polymorphism corresponds to a recognisable pattern in a De Bruijn graph constructed from the RNA-seq reads, we proposed a general model for all polymorphisms in such graphs. We then introduced an exact algorithm, called KISSPLICE, to extract alternative splicing events. The first version of KISSPLICE appeared in 2011, but several important improvements were implemented in 2012 [24]. The first improvement was the memory consumption, the new version is much more memory efficient and can handle datasets of approximately 10^8 reads. The second was in the running time, the enumeration step can now be done in parallel, which results in a significant speedup in the overall running time. Finally, an improved event quantification step was added to the method.

Application-wise, we showed that KISSPLICE enables to identify more correct events than general purpose transcriptome assemblers. Additionally, on a 71 M reads dataset from human brain and liver tissues, KISSPLICE identified 3497 alternative splicing events, out of which 56% are not present in the annotations, which confirms recent estimates showing that the complexity of alternative splicing has been largely underestimated so far.

6.3. Efficient bubble and/or cycle enumeration in directed/undirected graphs

Polymorphisms in DNA- or RNA-seq data lead to recognisable patterns in a de Bruijn graph representation of the reads obtained by sequencing. Such patterns have been called mouths, or bubbles in the literature. They correspond to two vertex-disjoint directed paths between a source s and a target t. Due to the high number of such bubbles that may be present in real data, their enumeration is a major issue concerning the efficiency of dedicated algorithms. We developed the first linear delay algorithm to enumerate all bubbles with a given source [31].

By combining the insights from the most efficient but not optimal solution presented by Johnson [SIAM J. Computing, 1975] for simple cycle enumeration in undirected graphs and an amortisation technique previously established by our collaborators Roberto Grossi and Rui Ferreira [ESA, 2011] from the University of Pisa, Italy, we obtained the first optimal solution to list all the simple cycles in an undirected graph G (paper accepted at SODA 2013, to appear). Moreover, we also obtained the first optimal solution to list all the simple paths from s to t in an undirected graph G. This work benefited also from discussions and work from Pierluigi Crescenzi and Marie-France Sagot. The method is not naturally extendable to directed graphs, and the challenge is now to obtain optimal solutions in this case also.

6.4. Simulating RNA-seq experiments

RNAseq experiments now enable to characterise the RNA complement of a cell. However, the series of steps (fragmentation, reverse transcription, sequencing) that separate the initial RNA molecules from the short DNA reads obtained in fine are not well understood although it is widely accepted that they contribute to generating noise in the signal. We introduced the FLUXSIMULATOR [14], a computer program able to reproduce the biases seen in RNAseq data. This pipeline should prove useful both to produce realistic data on which to test programs which aim at reconstructing RNA from short reads, and suggest ways of improving the experimental steps so that they produce less noise.

6.5. Chimeric Transcripts may be Translated

There is now increasing evidence for the existence of so-called Chimeric Transcripts. In contrast to regular transcripts, which are composed of exons located close to each other on the genome, these chimeric transcripts can be composed of exons which are located megabases away, or even on different chromosomes. We showed that these chimeras are lowly expressed, are tissue specific, and that some of them may be translated, yielding proteins with altered function or localisation [13].

6.6. Transcriptomics of symbiosis in the Asobara tabida-Wolbachia association

Wolbachia has evolved a very peculiar phenotype in the host *Asobara tabida* where it is obligatory for oogenesis. Transcriptomics approaches were developed first using Sanger sequencing of mRNA [19]. It has now been complemented by RNAseq analyses on two lines, which exhibit different ovarian phenotypes in absence of *Wolbachia*. We have currently analysed these data both to isolate genes that are differentially expressed, but also that exhibit polymorphism between the lines. Interesting candidates were detected that are under further investigation and that are involved in the regulation of early oogenesis, apoptosis, autophagy and oxidative stress. This part is in direct connection with the algorithms developed by BAMBOO for the analysis of NGS data without a reference genome (KISSPLICE).

6.7. MicroRNA predictor

We developed a microRNA predictor using structural and target information. The method shows 97% sensitivity and 90% specificity for the *Acyrthosiphon pisum* genome. Comparing to the results of the previous method we developed in 2010 (available in the software CRAVELA) we obtained a better performance (sensitivity 90% and specificity 88%). However, as we are working on a genome wide scale, it is important to obtain even better specificity (obviously, maintaining a reasonable sensitivity). This work is currently in development.

On the other hand, the computational search for novel miRNA precursors often involves also some sort of structural analysis with the aim of identifying which type of structures are recognised and processed by the cellular miRNA-maturation machinery. A natural way to tackle this problem is to perform clustering over the candidate structures along with known miRNA precursor structures. Mixed clusters allows then the identification of candidates that are similar to known precursors. Given the large number of pre-miRNA candidates that can be identified in single-genome approaches, even after applying several filters for precursor robustness and stability, a conventional structural clustering approach is unfeasible. We proposed a method to represent candidate structures in a feature space which summarises key sequence/structure characteristics of each candidate [21]. We showed that proximity in this feature space is related to sequence/structure similarity, and we selected candidates which have a high similarity to known precursors. Additional filtering steps were then applied to further reduce the number of candidates to those with greater transcriptional potential.

6.8. Genomics of symbiosis

Insect symbioses are model systems for studying the effect of symbionts and the evolution of bacterial genomes. Members from the LBBE described the symbiotic complement of different biotypes of the insect *Bemisia tabaci* in Western Africa. We further obtained the complete genome of different symbionts that co-exist in *Bemisia tabaci*, among which the the primary symbiont *Portiera* [25], *Hamiltonella*, *Rickettsia* and *Wolbachia*. Analyses are underway, that concern the possible complementation between *Hamiltonella* and *Portiera* and the comparative analyses of different *Hamiltonella* genomes.

6.9. Representation and curation of metabolic data: UniPathway, Rhea and MNX

These activities are carried out in collaboration with the SwissProt group at the Swiss Institute for BioInformatics (SIB). UniPathway (http://www.unipathway.org) is a manually curated database of metabolic pathways. It provides the official controlled vocabulary for pathway annotation within UniProtKB records since 2009. A complete description of the UniPathway database and of its relationship with UniProtKB has been published in *Nucleic Acids Research* (Jan. 2012 Database Issue) [22]. Rhea (http://www.ebi.ac.uk/rhea) is developed jointly with the European Institute for Bioinformatics (EBI) and the SIB. It provides a comprehensive resource of expert-curated biochemical reactions, for use in a large spectrum of applications, including metabolic network reconstruction and pathway inference. The complete description of the Rhea database appeared in the Jan. 2012 *NAR* Database issue [5]. The MNX project is developed in the context of the MetaNetX project (http://www.metanetx.org). It attempts to automate the reconciliation of discrepancies between metabolite or reaction information from distinct resources (BiGG, BRENDA, ChEBI/Rhea, KEGG, MetaCyc, UniPathway, The SEED, Reactome), thereby alleviating a major bottleneck in the construction of genome-scale metabolic network models. The MNXref namespace is available at http://www.metanetx.org/mnxdoc/mnxref.html and the method to compute the MNXref namespace is described in [8].

6.10. Annotation of the proteins of Angomonas deanei and Strigomonas culicis

Angomonas deanei and *Strigomonas culicis* are trypanosomatids that harbour only one beta-proteobacterial endosymbiont and this mutualistic association is an interesting model to study eukaryotic cell evolution. The genomes of these organisms were sequenced by our collaborators at LNCC / MCT (Brazil) and we participated in the functional annotation of these genomes as concerns their metabolism which enabled to reveal new aspects of the *Trypanosomatidae* family. This work has been submitted for publication. It was done with Ana Tereza Vasconcelos in a collaboration with Maria Cristina Machado Motta (UFRJ - Brazil).

6.11. Finding candidate genes for orphan enzymes

Of all biochemically characterized metabolic reactions formalized by the IUBMB, over one out of four have yet to be associated with a nucleic or protein sequence, *i.e.* are sequence-orphan enzymatic activities. Few bioinformatics annotation tools are able to propose candidate genes for such activities by exploiting

context-dependent rather than sequence-dependent data, and none are readily accessible and propose result integration across multiple genomes. We introduced CANOE (Candidate genes for Orphan Enzymes), a four-step bioinformatics strategy that proposes ranked candidate genes for sequence-orphan enzymatic activities (or orphan enzymes for short) [26]. Our strategy found over 60,000 genomic metabolons in more than 1,000 prokaryote organisms from the MICROSCOPE platform developed by the group of Claudine Médigue from the Génoscope with whom this work was done, generating candidate genes for many metabolic reactions, of which more than 70 distinct orphan reactions. A computational validation of the approach was discussed and we presented a case study on the anaerobic allantoin degradation pathway in *Escherichia coli* K-12.

6.12. Metabolic cooperation of symbionts and their host trypanosomatids

Trypanosomatids that harbour a symbiotic bacterium (SHTs) are known to have less nutritional requirements when compared to their counterparts without symbionts (RTs). Nutritional and biochemical data indicated that the symbionts largely contributed to the routes for amino acid and vitamin biosynthesis. We analysed the genomic data of 5 SHTs and their respective symbionts and 2 RTs and we found most of the genes related to those pathways in the symbionts. This work will soon be submitted for publication. It is being done with Ana Tereza Vasconcelos in a collaboration with Erney P. Camargo, Marta M.G. Teixeira (USP - Brazil), João M.P. Alves, Gregory A. Buck (VCU - USA), and Maria Cristina Machado Motta (UFRJ - Brazil).

6.13. Structural and dynamical analysis of biological networks

We published a review on the structural and dynamical analysis of biological networks with as main focus explaining the cares that should be taken when this kind of analysis is performed [18]. Correctly distinguishing between potential metabolic networks and their realisations is necessary in choosing the right methods to be used and in the interpretation of their outcomes. In our review, we covered several different techniques, both static and dynamic, for the analysis of metabolic networks such as centrality techniques, flux-balance analysis and kinetic modelling of full-scale networks.

6.14. Network distance analysis

We addressed the diameter computation problem in the case of undirected unweighted graphs, where the diameter D is defined as the maximum distance among all the pairs of nodes and the distance $d(u, v)$ between two nodes u and v is defined as the number of edges contained in the shortest path from u to v. In the context of real-world networks, the textbook method based on performing a breadth-first search (in short, BFS) from *every* node of the graph, requires a prohibitive cost of $O(nm)$ time, where n is the number of nodes and m is the number of edges of the graph. Our main contribution consists of showing that BFS can indeed be an extremely *powerful* tool to compute the *exact* value of the diameter, whenever it is used in a more clever way. In particular, we have developed the *iterative* Fringe Upper Bound (in short, iFUB) algorithm to calculate the exact value of the diameter. This work has been accepted for publication in *Theoretical Computer Science* (to appear).

We then successively generalised the idea of the iFUB algorithm, by presenting the directed iFUB (in short, DiFUB) algorithm, in order to calculate the diameter of the strongly connected components of directed graphs [33]. As far as we know, DiFUB is the first algorithm which is able to compute exactly the diameter of the strongly connected components of huge real-world directed graphs. The DiFUB algorithm can also return a pair of nodes whose distance is exactly equal to the diameter, and a natural adaptation of it works also for weighted graphs.

6.15. Information spreading in dynamic graphs

We showed how a technique used to analyse the flooding completion time in the case of a special class of random evolving graph model, that is, the *edge-Markovian model*, can be used in order to prove that the flooding completion time of a random evolving graph $(G_t)_{t\geq 0}$ is bounded by $kD + 2C$, where intuitively (1) k is the smallest time necessary for the rising of a giant component, (2) D is the diameter of the giant

component, and (3) C is the time required for the nodes outside the giant component to eventually get an edge connecting them to the giant component [30]. Then, based on this result, we developed a general methodology for analysing flooding in sequences of random graphs and we applied this general methodology to the case of power-law evolving graphs (that is, sequences of mutually independent random graphs such that the number y of nodes of degree x distributes like $1/x^\beta$ for some $\beta > 0$), and to the case of an arbitrary given degree distribution.

6.16. Metabolic network comparison

Previous works on minimal gene sets, when analysing host-dependent bacteria, found small common sets of metabolic genes. When such analyses are restricted to bacteria with similar lifestyles, larger portions of metabolism are expected to be shared and their composition is worth investigating. Comparing the small molecule metabolism of 58 bacteria carefully selected and representing a range of lifestyles, we found not a single enzymatic reaction common to all of them. While obligate intracellular symbionts have no core of reactions within their group, extracellular and cell-associated symbionts do have a small core enriched in biosynthetic processes composed of disconnected fragments. As more genomes are added, we expect, based on our simulations, that the core of cell-associated and extracellular bacteria continues to diminish, converging to approximately 60 reactions. These results were in preparation in 2011 and are now published [17]. The work was done with Ana Tereza Vasconcelos and in a collaboration with Ludovic Cottret (INSA Toulouse).

6.17. Core and periphery of metabolic networks

The core metabolism can be defined as the reactions present in every organism, however it is not robust considering that adding or removing one organism in the study will modify the resulting set. An alternative way is to include in the core the reaction that is present in a large enough proportion of species. For that, we proposed a method where the threshold to decide what is large enough is not set by the user (thus relying on a subjective choice), but rather automatically selected by the method, relying on the information contained in the data. Two approaches are being proposed, one is EM (Expectation Maximization) which relies only on the information of presence / absence of a reaction in a species while the second (NEM - Neighboring Expectation Maximization) relies on a neighbouring relation between reactions. The latter tends to classify in a same group (core or periphery) a reaction for which a majority of neighbours belong to a same group. The work is being done with Ana Tereza Vasconcelos in a collaboration with Catherine Matias, Christophe Ambroise, Yolande Diaz (Genopole, CNRS).

6.18. Metabolic stories

Enumerating stories, *i.e.*, enumerating maximal directed acyclic graphs with sets of sources and targets contained in a given subset of the nodes, is an algorithmic approach we proposed for interpreting metabolomics experiments. The modelling, algorithms and complexity results were recently accepted for publication [2]. The complexity of the enumeration problem remains unknown. There are also further modelling issues that could be dealt with in a near future. Both considerations were also detailed in a talk given in August at St. Petersburg, in the First RECOMB Satellite Conference on Open Problems in Algorithmic Biology.

We then applied our enumerating method on real data. We analysed data on the detoxification process of yeast cells exposed to cadmium. Our method allowed to recover known pathways involved in the process but also to propose alternative scenarios. The method was also investigated in order to automatically propose metabolic pathways through an experiment in which we try to recover known metabolic pathways using only minimal information (*e.g.*, their entries and endpoints). A paper is in prepartion and should soon be submitted for publication. This work is being done in collaboration with Fabien Jourdan and Ludovic Cottret from the INRA at Toulouse, and with Christophe Junot from the CEA in Paris.

6.19. Minimal precursor sets

We proposed two new, more efficient algorithms for the enumeration of minimal precursor sets: PITUFINA and PAPA PITUFO [3]. The model of minimal precursor sets we had previously published was the first to formally take into account cycles, which are a common event in metabolic networks. The new methods avoid the memory issues of our previous approach by traversing directly the metabolic network structure instead of building a secondary tree representation. PAPA PITUFO additionally saves pre-computed solutions by a local modification of the network.

6.20. Minimum ratio cover of matrix columns by extreme rays of its induced cone

Given a matrix S and a subset of columns R, we studied the problem of finding a cover of R with extreme rays of the cone $\mathcal{F} = \{v \in \mathbb{R}^n \mid Sv = \mathbf{0}, v \geq \mathbf{0}\}$, where an extreme ray v covers a column k if $v_k > 0$ [34]. In order to measure how proportional a cover is, we introduced two different minimisation problems, namely the MINIMUM GLOBAL RATIO COVER (MGRC) and the MINIMUM LOCAL RATIO COVER (MLRC) problems. In both cases, we applied the notion of the *ratio* of a vector v, which is given by $\frac{\max_i v_i}{\min_{j|v_j>0} v_j}$. These problems were originally motivated by a biological question on metabolic networks. This notion of ratio is also of interest in the field of *exact linear programming*, where current algorithms for scaling a matrix have a complexity that depends on the ratio of its elements. We showed that these two problems are NP-hard, even in the case in which $|R| = 1$. We introduced a mixed integer programming formulation for the MGRC problem, which is solvable in polynomial time if all columns should be covered, and introduce a branch-and-cut algorithm for the MLRC problem. Finally, we presented computational experiments on data obtained from real metabolic networks.

6.21. Optimal flux spaces of genome-scale stoichiometric models

The metabolism of organisms can be studied with comprehensive stoichiometric models of their metabolic networks. Flux balance analysis (FBA) calculates optimal metabolic performance of stoichiometric models. However, detailed biological interpretation of FBA is limited because, in general, a huge number of flux patterns give rise to the same optimal performance. The complete description of the resulting optimal solution spaces was thus far a computationally intractable problem. We introduced COPE-FBA: Comprehensive Polyhedra Enumeration Flux Balance Analysis, a computational method that solves this problem [15]. COPE-FBA indicates that the thousands to millions of optimal flux patterns result from a combinatorial explosion of flux patterns in just a few metabolic sub-networks. The entire optimal solution space can now be compactly described in terms of the topology of these sub-networks. COPE-FBA simplifies the biological interpretation of stoichiometric models of metabolism, and provides a profound understanding of metabolic flexibility in optimal states.

6.22. Lateral gene transfer as a support for the tree of life

We published with Sophie Abby the last results of her PhD work that apply an explicit phylogenetic model of horizontal gene transfer to bacterial and archaeal phyla [1]. We showed that lateral gene transfer allows to discriminate between phylogenetic hypotheses, and that in a typical bacterial gene family, 96-98% of tree branches result from vertical descent and 2-4% from lateral gene transfer.

6.23. Comparative approximability of hybridization number and directed feedback vertex set

We showed that the problem of computing the hybridization number of two rooted binary phylogenetic trees on the same set of taxa X has a constant factor polynomial-time approximation if and only if the problem of computing a minimum-size feedback vertex set in a directed graph (DFVS) has a constant factor polynomial-time approximation. The latter problem, which asks for a minimum number of vertices to be removed from

a directed graph to transform it into a directed acyclic graph, is one of the problems in Karp's seminal 1972 list of 21 NP-complete problems. However, despite considerable attention from the combinatorial optimisation community, it remains to this day unknown whether a constant factor polynomial-time approximation exists for DFVS. Our result thus placed the (in)approximability of hybridization number in a much broader complexity context, and as a consequence we obtained that hybridization number inherits inapproximability results from the problem Vertex Cover [16]. On the positive side, we used results from the DFVS literature to give an $O(logrloglogr)$ approximation for the hybridization number, where r is the value of an optimal solution to the hybridization number problem. This work is submitted for publication.

6.24. Influence of symbionts on antagonistic interactions

Symbionts are often key players in antagonistic interactions between their hosts and other organisms. In host-parasitoid interactions, both players can be infected by different symbionts. We investigated how a virus and *Wolbachia*, respectively infecting a parasitoid and a drosophila, can shape the host-parasitoid interaction. While only a limited effect *Wolbachia* has been detected, the virus protects the parasitoid from the immune response of *Drosophila* [20]. Protection conferred by symbionts to their insect hosts is a promising avenue for antivectorial programs, but requires a thorough analysis of the evolutionary consequences of protection. We reviewed the literature on this topic [28].

6.25. Mod/Resc Parsimony Inference

We addressed a computational biology problem that aims at understanding a mechanism that could potentially be used to genetically manipulate natural insect populations infected by inherited, intra-cellular parasitic bacteria. In this problem, that we denoted by Mod/Resc Parsimony Inference, we are given a boolean matrix and the goal is to find two other boolean matrices with a minimum number of columns such that an appropriately defined operation on these matrices gives back the input. We showed that this is formally equivalent to the Biclique Edge Cover for Bipartite Graphs problem and derive some complexity results for our problem using this equivalence. We provided a new, fixed parameter tractability approach for solving both problems that slightly improves upon a previously published algorithm for the Biclique Edge Cover for Bipartite Graphs. Finally, we presented experimental results applying some of our techniques to a real-life dataset. This is the augmented journal version [23] of the conference paper that appeared in 2011.

6.26. On the genetic architecture of cytoplasmic incompatibility

Numerous insects carry intracellular bacteria manipulating their reproduction and thus facilitating their own spread. Cytoplasmic incompatibility (CI) is a common form of such manipulation, where a (currently uncharacterized) bacterial modification of male sperm 35 induces the early death of embryos unless the fertilized eggs carry the same bacteria, inherited from the mother. The death of uninfected embryos provides an indirect selective advantage to infected ones, thus enabling the spread of the bacteria. We used and expanded recently developed algorithms (the first being the one described in the previous item) to infer the genetic architecture underlying the complex incompatibility data from the mosquito *Culex pipiens*. We showed that CI requires more genetic determinants 40 than previously believed, and that quantitative variation in gene products potentially contributes to the observed CI patterns. In line with population genetic theory of CI, our analysis suggests that toxin factors (those inducing embryo death) are present in fewer copies in the bacterial genomes than antitoxin factors (those ensuring that infected embryos survive). In combination with comparative genomics, our approach will provide helpful guidance to 45 identify the genetic basis of CI, and more generally of other toxin / anti-toxin systems that can be conceptualised under the same framework. This work is currently submitted for publication. It was done in collaboration with Sylvain Charlat from the LBBE.

6.27. Viral population structure and dynamics

The work which started a few years ago with the Pasteur Institute in Cambodia (Dr. P. Buchy) and the CIRAD at Montpellier (Dr. R. Frutos) on viral population structure and dynamics has been continued in 2012, focusing on the H5N1 and Dengue viruses. The exploratory statistical approach based on MCoA (see the Bamboo annual report for 2011) was used to identify a novel H5N1 endemic sub-clade specific to Cambodia [27] and the work performed last year on Dengue serotype 1 has been extended in 2012 to serotypes 2 and 3 [11] thus providing a more precise view of the virus population dynamics over the last 12 years and demonstrating "synchronized" replacements most probably linked to climatic disasters like flood or drought.

6.28. Charge group partitioning in biomolecular simulation

Molecular simulation techniques are increasingly being used to study biomolecular systems at an atomic level. Such simulations rely on empirical force fields to represent the intermolecular interactions. There are many different force fields availableleach based on a different set of assumptions and thus requiring different parametrization procedures. Recently, efforts have been made to fully automate the assignment of force-field parameters, including atomic partial charges, for novel molecules. In this work, we focused on a problem arising in the automated parameterisation of molecules for use in combination with the gromos family of force fields: namely, the assignment of atoms to charge groups such that for every charge group the sum of the partial charges is ideally equal to its formal charge. In addition, charge groups are required to have size at most k. We showed NP-hardness and gave an exact algorithm capable of solving practical problem instances to provable optimality in a fraction of a second [32].

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. InférenceGraphesRégulation

- Title: Inférence de graphes de régulations génétiques à partir de données d'expression
- Coordinator: H. Charles
- BAMBOO participant(s): H. Charles, L. Brinza, M.-F. Sagot
- Type: Pré-Projet de Recherche de l'IXXI (2012-2013)
- Web page: Not available

7.2. National Initiatives

7.2.1. ABS4NGS

- Title: Solutions Algorithmiques, Bioinformatiques et Logicielles pour le Séquençage Haut Débit
- Coordinator: E. Barillot
- BAMBOO participant(s): V. Lacroix
- Type: ANR (2012-2015)
- Web page: Not available

7.2.2. Adapthantroph

- Title: Adaptation des insectes aux anthroposystèmes
- Coordinator: M. Harry
- BAMBOO participant(s): C. Vieira
- Type: ANR Génoplante (2009-2012)

- Web page: Not available

7.2.3. Exomic

- Title: Functional annotation of the transcriptome at the exon level
- Coordinator: D. Auboeuf (Inserm, Lyon)
- BAMBOO participant(s): V. Lacroix, M.-F. Sagot
- Type: INSERM Systems Biology Call (2012-2015)
- Web page: Not available

7.2.4. ImmunSymbArt

- Title: Immunity and Symbiosis in Arthropods
- Coordinator: D. Bouchon
- BAMBOO participant(s): F. Vavre
- Type: ANR Blanc (2010-2014)
- Web page: Not available

7.2.5. Metagenomics of Bemisia tabaci

- Title: Metagenomics of *Bemisia tabaci* symbiotic communities
- Coordinator: L. Mouton (LBBE, UCBL)
- BAMBOO participant(s): F. Vavre, M.-F. Sagot
- Type: Genoscope Project
- Web page: Not available

7.2.6. MIRI

- Title: Mathematical Investigation of "Relations Intimes"
- Coordinator: M.-F. Sagot
- BAMBOO participant(s): V. Acuña, C. Baudet, C. Gautier, V. Lacroix, P. Milreu, C. Klein, I. Nor, M.-F. Sagot, P. Simões
- Type: ANR Blanc (2009-2012)
- Web page: http://pbil.univ-lyon1.fr/members/sagot/htdocs/team/projects/miri/miri.html

7.2.7. SpeciAphid

- Title: Evolutionary genetics and mechanisms of plant adaptation in aphids
- Coordinator: Jean-Christophe Simon (IGEPP, INRA, Rennes)
- BAMBOO participant(s): H. Charles, Y. Rahbé
- Type: ANR (2012-2014)
- Web page: Not available

7.3. European Initiatives

7.3.1. FP7 Projects

7.3.1.1. Microme

- Title: The Microme Project: A Knowledge-Based Bioinformatics Framework for Microbial Pathway Genomics
- Coordinator: P. Kersey (EBI)

- European partners: Amabiotics (France), CEA (France), CERTH (Greece), CSIC (Spain), CNIO (Spain), DSMZ (Germany), EBI (UK), HZI (Germany), Isthmus (France), Molecular Nertwork (Germany), SIB (Switzerland), Tel Aviv Univ. (Israel), Université Libre de Bruxelles (Belgium), WTSI (UK), Wageningen Univ. (The Netherlands)
- BAMBOO participant(s): Anne Morgat
- Type: Collaborative Project. Grant Agreement Number 222886-2
- Web page: http://www.microme.eu

7.3.1.2. SISYPHE

- Title: Species Identity and SYmbiosis Formally and Experimentally explored
- Coordinator: M.-F. Sagot
- BAMBOO participant(s): Whole BAMBOO team
- Type: ERC Advanced Grant (2010-2015)
- Web page: http://pbil.univ-lyon1.fr/members/sagot/htdocs/team/projects/sisyphe/sisyphe.html

7.3.1.3. Symbiox

- Title: Role of the oxidative environment in the stability of symbiotic associations
- Coordinator: F. Vavre
- BAMBOO participant(s): F. Vavre
- Type: Marie Curie IOF for Natacha Kremer (2011-2014)
- Web page: Not available

7.3.1.4. SWIPE

- Title: Predicting whitefly population outbreaks in changing environments
- Coordinator: E. Zchori-Fein
- BAMBOO participant(s): F. Vavre
- Type: European ERA-NET program ARIMNET (2012-2015)
- Web page: Not available

7.3.2. Collaborations with Major European Organizations

Partner 1: Pierluigi Crescenzi, Univ. Florence, Italy

Algorithmic (graphs, trees, sequences), complexity

Partner 2: Ana Teresa Freitas and Susana Vinga, INESC-ID, IST Lisbon, Portugal

NGS, metabolism, small RNAs, motifs

Partner 3: Alberto Marchetti-Spaccamela, Univ. Rome La Sapienza, Italy

Algorithmic (graphs, trees), complexity

Partner 4: Nadia Pisanti and Roberto Grossi, Univ. Pisa, Italy

Algorithmic (graphs, trees, sequences)

Partner 5: Leen Stougie, Free Univ. Amsterdam and CWI, the Netherlands

Algorithmic (graphs, trees), complexity

7.4. International Initiatives

7.4.1. DISCO

- Title: Laboratoire International de Recherche en BIoinformatique
- Coordinators: E. Zucca (Italy)
- BAMBOO participant(s): Pierluigi Crescenzi (external member BAMBOO)

- Type: Ministero dell'Istruzione, dell'Università e della Ricerca
- Web page: http://bart.disi.unige.it/DISCO/

7.4.2. LIA project with Brazil: LIRIO

- Title: Laboratoire International de Recherche en BIoinformatique
- Coordinators: M.-F. Sagot (France), A. T. Vasconcelos (LNCC, Brazil)
- BAMBOO participant(s): BAMBOO Team
- Type: LIA CNRS
- Web page: https://team.inria.fr/bamboo/en/cnrs-lia-laboratoire-international-associe-lirio/

7.4.3. Inria-Faperj (Brazil) project: RAMPA

- Title: Bioinformatics for the Reconstruction and Analysis of the Metabolism of PArasites
- Coordinators: M.-F. Sagot (France), A. T. Vasconcelos (LNCC, Brazil)
- BAMBOO participant(s): Whole BAMBOO Team
- Type: Faperj-Inria
- Web page: Not available

7.4.4. Project within CIRIC

- Title: Omics Integrative Sciences
- Coordinators: Alejandro Maass (Chile), Anne Siegel and M.-F. Sagot (France)
- BAMBOO participant(s): BAMBOO Team
- Type: Communication and Information Research and Innovation Center (CIRIC)
- Web page: Not available

7.4.5. Inria International Partners

- Acronym: AMICI
- Title: Algorithms and Mathematics for Investigating Communication and Interactions intra- and inter-organisms
- Coordinators: M.-F. Sagot (France), A. Marchetti-Spaccamela (Univ. Rome, Italy), L. Stougie (Free Univ. Amsterdam and CWI, the Netherlands), P. Crescenzi, Univ. Florence, Italy), N. Pisanti (Univ. Pise, Italy)
- BAMBOO participant(s): Whole BAMBOO Team
- Type: Inria International Partner
- Web page: http://amici.dsi.unifi.it/amici/

7.5. International Research Visitors

7.5.1. Visits of International Scientists

Andrea Marino, PhD student (Supervisor: Pierluigi Crescenzi), University of Florence, Italy, visit of 3 months and various visits of 1-2 weeks
Maria Cristina Motta, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, visit of 10 days
Susana Vinga, Professor, INESC-ID, IST Lisbon, Portugal, visit of 1 week
Arnaldo Zaha, Universidade Federal de Rio Grande do Sul, Porto Alegre, Brazil, visit of 10 days

8. Dissemination

8.1. Scientific Animation

Hubert Charles is director of studies of the "Bioinformatique et Modélisation (BIM)" track at the INSA-Lyon. He is co-director of the Biosciences Department of the INSA-Lyon, and co-director of the Doctoral School E2M2.

Marie-France Sagot is a member of the Scientific Advisory Board ("Conseil Scientifique (COS)" for the Inria Grenoble Rhône-Alpes Research Center. She is since 2012 member of the Scientific Board of the French Society of Computer Science (SFI). She was co-chair for one area track of the 20th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) 2012, and co-chair for the RECOMB Satellite Workshop Comparative Genomics (RECOMB-CG) 2012. She is Editor-in-Chief of *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, and Associate Editor of *BMC Bioinformatics*, *Algorithms for Molecular Biology*, *Journal of Discrete Algorithms*, and *Lecture Notes in BioInformatics*. She is member of the Steering Committee for the European Conference on Computational Biology (ECCB), for the International Symposium on Bioinformatics Research and Applications (ISBRA), and for the Latin American Theoretical Informatics Symposium (LATIN). She was member of the Program Committee for BSB, ISMB, PACBB, PSC, WABI. She is a member of the scientific council of the Labex EcoFect.

Fabrice Vavre is director of the GDR 2153(CNRS) "Interactions multipartenaires dans les populations et les communautés d'insectes". He is also member of the management committee and responsible of a working group in the COST Action FA0701 "Arthropod Symbiosis: from fundamental studies to pest and disease management" which ended in June 2012. He was president of a selection committee for a teaching position at the University Lyon 1 in Microbial Genetics. He was elected member of the Section 29 of the Comité National de la Recherche Scientifique. He is a member of the scientific council of the Labex EcoFect.

Alain Viari is since 2012 Deputy Scientific Director at Inria in charge of ICST for Life and Environmental Sciences. He represents Inria in several national instances related to Life Sciences and Health and is member of several scientific advisory boards (IMMI (Institut de Microbiologie et Maladies Infectieuses / Aviesan); IRT (Institut de Recherche Technologique) BioAster). He is the French coordinator of the Bioinformatics working group of the France-US joint committee on Science and Technology. He is a member of the Scientific Advisory Board ("Conseil Scientifique (COS)" for the Inria Grenoble Rhône-Alpes Research Center.

Cristina Vieira is director of the GDRE "Comparative genomics" since the GDRE was renewed in 2010.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Three members of the BAMBOO project are professors or associate professors at the University Claude Bernard in Lyon and at the INSA Lyon: Hubert Charles, Vincent Lacroix, and Cristina Vieira. They therefore have a full teaching service (at least 192 hours) except for Cristina Vieira who became since 2010 a Junior Member of the Institut Universitaire de France.

Various members of the EPI have developed over the years courses in biometry, bioinformatics and evolutionary biology at all levels of the University as well as at the "École Normale Supérieure" (ENS) of Lyon and the INSA ("Institut National de Sciences Appliquées"). Two members of the EPI have also in the past participated in, or sometimes organised courses or teaching modules at the international level: creation and support of a Master's course in Ho-Chi-Minh, Vietnam, and creation and direction of a PhD Program in Computational Biology in Lisbon, Portugal (http://bc.igc.gulbenkian.pt/pdbc/).

8.2.2. Supervision

The following are the PhDs defended in BAMBOO in 2012.

PhD: Patricia Simões, University of Lyon 1, March 14, supervisors S. Charlat and M.-F. Sagot

PhD: Paulo Vieira Milreu, University of Lyon 1, December 19, supervisors C. Gautier, V. Lacroix and M.-F. Sagot

8.2.3. Juries

M.-F. Sagot: Reviewer of the HDR of Guillaume Blin (University of Paris-Est) and of the PhDs of Stéphane Prin (Muséum National d'Histoire Naturelle) and Adam Smith (University of Évry).

F. Vavre: Member of the committee for the HDR of Franck Prugnolle (University of Montpellier), and for the PhDs of Barbara Reumer (University of Leiden), Winka Le Clec'h (University of Poitiers), Johan Decelle (University of Paris 6) and Flore Zele (University of Montpellier).

A. Viari: Member of the committee for the PhDs of Adam Smith (University of Évry) Philippe Bordron (University of Nantes).

8.3. Popularization

Fabrice Vavre gave two talks, respectively entitled "Quand les parasites utilisent nos gènes, et réciproquement" and "Sommes-nous manipulés par nos gènes", at the Université Ouverte. He also gave a talk entitled "Diversité des interactions hôtes-microorganismes et nouvelles méthodes de lute contre les maladies infectieuses" at the General Meeting of the Biotechnologies Professors.

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

[1] S. S. ABBY, E. TANNIER, M. GOUY, V. DAUBIN. *Lateral gene transfer as a support for the tree of life.*, in "Proceedings of the National Academy of Sciences", March 2012, vol. 109, n^{o} 13, p. 4962-7 [*DOI :* 10.1073/PNAS.1116871109], http://hal.inria.fr/hal-00752055.

[2] V. ACUÑA, E. BIRMELÉ, L. COTTRET, P. CRESCENZI, F. JOURDAN, V. LACROIX, A. MARCHETTI-SPACCAMELA, A. MARINO, P. V. MILREU, M.-F. SAGOT, L. STOUGIE. *Telling stories: Enumerating maximal directed acyclic graphs with a constrained set of sources and targets*, in "Theoretical Computer Science (TCS)", 2012, vol. 457, p. 1–9 [*DOI :* 10.1016/J.TCS.2012.07.023], http://hal.inria.fr/hal-00764025.

[3] V. ACUÑA, P. V. MILREU, L. COTTRET, A. MARCHETTI-SPACCAMELA, L. STOUGIE, M.-F. SAGOT. *Algorithms and complexity of enumerating minimal precursor sets in genome-wide metabolic networks.*, in "Bioinformatics", October 2012, vol. 28, n^{o} 19, p. 2474-2483 [*DOI :* 10.1093/BIOINFORMATICS/BTS423], http://hal.inria.fr/hal-00751715.

[4] A. AKKOUCHE, R. REBOLLO, N. BURLET, C. ESNAULT, S. MARTINEZ, B. VIGINIER, C. TERZIAN, C. VIEIRA, M. FABLET. *tirant, a newly discovered active endogenous retrovirus in Drosophila simulans.*, in "J Virol", April 2012, vol. 86, n^{o} 7, p. 3675-81 [*DOI :* 10.1128/JVI.07146-11], http://hal.inria.fr/hal-00763499.

[5] R. ALCÂNTARA, K. B. AXELSEN, A. MORGAT, E. BELDA, E. COUDERT, A. BRIDGE, H. CAO, P. DE MATOS, M. ENNIS, S. TURNER, G. OWEN, L. BOUGUELERET, I. XENARIOS, C. STEINBECK. *Rhea–a manually curated resource of biochemical reactions.*, in "Nucleic Acids Research", January 2012, vol. 40, n^{o} Database issue, p. D754-60 [*DOI :* 10.1093/NAR/GKR1126], http://hal.inria.fr/hal-00746861.

[6] J. ALLALI, C. SAULE, C. CHAUVE, Y. D'AUBENTON-CARAFA, A. DENISE, C. DREVET, P. FERRARO, D. GAUTHERET, C. HERRBACH, F. LECLERC, A. DE MONTE, A. OUANGRAOUA, M.-F. SAGOT, M. TERMIER, C. THERMES, H. TOUZET. *BRASERO: A resource for benchmarking RNA secondary structure comparison algorithms*, in "Advances in Bioinformatics", 2012, http://hal.inria.fr/hal-00647725.

[7] C. BAUDET, Z. DIAS, M.-F. SAGOT. *Sampling Solution Traces for the Problem of Sorting Permutations by Signed Reversals.*, in "Algorithms for Molecular Biology", June 2012, vol. 7, n° 1, 18 [*DOI :* 10.1186/1748-7188-7-18], http://hal.inria.fr/hal-00739414.

[8] T. BERNARD, A. BRIDGE, A. MORGAT, S. MORETTI, I. XENARIOS, M. PAGNI. *Reconciliation of metabolites and biochemical reactions for metabolic networks.*, in "Briefings in Bioinformatics", December 2012 [*DOI :* 10.1093/BIB/BBS058], http://hal.inria.fr/hal-00764412.

[9] S. BERTHOUMIEUX, M. BRILLI, D. KAHN, H. DE JONG, E. CINQUEMANI. *On the identifiability of metabolic network models*, in "Journal of Mathematical Biology", November 2012 [*DOI :* 10.1007/S00285-012-0614-X], http://hal.inria.fr/hal-00762620.

[10] S. DJEBALI, J. LAGARDE, P. KAPRANOV, V. LACROIX, C. BOREL, J. M. MUDGE, C. HOWALD, S. FOISSAC, C. UCLA, J. CHRAST, P. RIBECA, D. MARTIN, R. R. MURRAY, X. YANG, L. GHAMSARI, C. LIN, I. BELL, E. DUMAIS, J. DRENKOW, M. L. TRESS, J. L. GELPI, M. OROZCO, A. VALENCIA, N. L. VAN BERKUM, B. R. LAJOIE, M. VIDAL, J. STAMATOYANNOPOULOS, P. BATUT, A. DOBIN, J. HARROW, T. HUBBARD, J. DEKKER, A. FRANKISH, K. SALEHI-ASHTIANI, A. REYMOND, S. E. ANTONARAKIS, R. GUIGÓ, T. R. GINGERAS. *Evidence for Transcript Networks Composed of Chimeric RNAs in Human Cells*, in "PLoS ONE", 2012, vol. 7, p. 1-22, http://hal.inria.fr/hal-00698428.

[11] V. DUONG, M. R. HENN, C. SIMMONS, C. NGAN, Y. BUNTHIN, L. GAVOTTE, A. VIARI, S. ONG, R. HUY, N. J. LENNON, S. LY, S. VONG, B. W. BIRREN, J. J. FARRAR, V. DEUBEL, R. FRUTOS, P. BUCHY. *Complex dynamic of dengue virus serotypes 2 and 3 in Cambodia following series of climate disasters.*, in "Infection, Genetics and Evolution", June 2012 [*DOI :* 10.1016/J.MEEGID.2012.05.012], http://hal.inria.fr/hal-00763498.

[12] M. FRENKEL-MORGENSTERN, A. GOROHOVSKI, V. LACROIX, M. ROGERS, K. IBANEZ, C. BOULLOSA, E. ANDRES LEON, A. BEN-HUR, A. VALENCIA. *ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data.*, in "Nucleic Acids Research", November 2012 [*DOI :* 10.1093/NAR/GKS1041], http://hal.inria.fr/hal-00756097.

[13] M. FRENKEL-MORGENSTERN, V. LACROIX, I. EZKURDIA, Y. LEVIN, A. GABASHVILI, J. PRILUSKY, A. DEL POZO, M. L. TRESS, R. JOHNSON, R. GUIGÓ, A. VALENCIA. *Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts.*, in "Genome Research", July 2012, vol. 22, n° 7, p. 1231-42 [*DOI :* 10.1101/GR.130062.111], http://hal.inria.fr/hal-00744761.

[14] T. GRIEBEL, B. ZACHER, P. RIBECA, E. RAINERI, V. LACROIX, R. GUIGÓ, M. SAMMETH. *Modelling and simulating generic RNA-Seq experiments with the flux simulator.*, in "Nucleic Acids Research", September 2012 [*DOI :* 10.1093/NAR/GKS666], http://hal.inria.fr/hal-00744755.

[15] S. KELK, B. G. OLIVIER, L. STOUGIE, F. J. BRUGGEMAN. *Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks.*, in "Nature Scientific Reports", 2012, vol. 2, 580 [*DOI :* 10.1038/SREP00580], http://hal.inria.fr/hal-00763439.

[16] S. KELK, L. VAN IERSEL, N. LEKIC, S. LINZ, C. SCORNAVACCA, L. STOUGIE. *Cycle killer... qu'est-ce que c'est? On the comparative approximability of hybridization number and directed feedback vertex set*, in "Siam Journal on Discrete Mathematics", 2012, vol. 26, n° 4, p. 1635-1656, http://hal.inria.fr/hal-00765019.

[17] C. KLEIN, L. COTTRET, J. KIELBASSA, H. CHARLES, C. GAUTIER, A. T. VASCONCELOS, V. LACROIX, M.-F. SAGOT. *Exploration of the core metabolism of symbiotic bacteria.*, in "BMC Genomics", 2012, vol. 13, n^{o} 1 [*DOI :* 10.1186/1471-2164-13-438], http://hal.inria.fr/hal-00737452.

[18] C. KLEIN, A. MARINO, M.-F. SAGOT, P. VIEIRA MILREU, M. BRILLI. *Structural and dynamical analysis of biological networks.*, in "Briefings in Functional Genomics", 2012 [*DOI :* 10.1093/BFGP/ELS030], http://hal.inria.fr/hal-00737455.

[19] N. KREMER, D. CHARIF, H. HENRI, F. GAVORY, P. WINCKER, P. MAVINGUI, F. VAVRE. *Influence of Wolbachia on host gene expression in an obligatory symbiosis.*, in "BMC Microbiology", January 2012, vol. 12 Suppl 1, S7 [*DOI :* 10.1186/1471-2180-12-S1-S7], http://hal.inria.fr/hal-00763438.

[20] J. MARTINEZ, A. DUPLOUY, M. WOOLFIT, F. VAVRE, S. L. O'NEILL, J. VARALDI. *Influence of the virus LbFV and of Wolbachia in a host-parasitoid interaction.*, in "PLoS ONE", 2012, vol. 7, n^{o} 4, e35081 [*DOI :* 10.1371/JOURNAL.PONE.0035081], http://hal.inria.fr/hal-00763437.

[21] N. D. MENDES, S. HEYNE, A. T. FREITAS, M.-F. SAGOT, R. BACKOFEN. *Navigating the unexplored seascape of pre-miRNA candidates in single-genome approaches.*, in "Bioinformatics", October 2012 [*DOI :* 10.1093/BIOINFORMATICS/BTS574], http://hal.inria.fr/hal-00752408.

[22] A. MORGAT, E. COISSAC, E. COUDERT, K. B. AXELSEN, G. KELLER, A. BAIROCH, A. BRIDGE, L. BOUGUELERET, I. XENARIOS, A. VIARI. *UniPathway: a resource for the exploration and annotation of metabolic pathways.*, in "Nucleic Acids Research", January 2012, vol. 40, n^{o} Database issue, D761-9 [*DOI :* 10.1093/NAR/GKR1023], http://hal.inria.fr/hal-00746856.

[23] I. NOR, D. HERMELIN, S. CHARLAT, J. ENGELSTADTER, M. REUTER, O. DURON, M.-F. SAGOT. *Mod/Resc Parsimony Inference: Theory and application*, in "Information and Computation", 2012, vol. 213, p. 23-32 [*DOI :* 10.1016/J.IC.2011.03.008], http://hal.inria.fr/hal-00763405.

[24] G. A. T. SACOMOTO, J. KIELBASSA, R. CHIKHI, R. URICARU, P. ANTONIOU, M.-F. SAGOT, P. PETER-LONGO, V. LACROIX. *KisSplice: de-novo calling alternative splicing events from RNA-seq data*, in "BMC Bioinformatics", March 2012, http://hal.inria.fr/hal-00681995.

[25] D. SANTOS-GARCIA, P.-A. FARNIER, F. BEITIA, E. ZCHORI-FEIN, F. VAVRE, L. MOUTON, A. MOYA, A. LATORRE, F. J. SILVA. *Complete Genome Sequence of "Candidatus Portiera aleyrodidarum" BT-QVLC, an Obligate Symbiont That Supplies Amino Acids and Carotenoids to Bemisia tabaci.*, in "J Bacteriol", December 2012, vol. 194, n^{o} 23, p. 6654-6655 [*DOI :* 10.1128/JB.01793-12], http://hal.inria.fr/hal-00763436.

[26] A. A. T. SMITH, E. BELDA, A. VIARI, C. MEDIGUE, D. VALLENET. *The CanOE strategy: integrating genomic and metabolic contexts across multiple prokaryote genomes to find candidate genes for orphan enzymes.*, in "PLoS Computational Biology", May 2012, vol. 8, n^{o} 5, e1002540 [*DOI :* 10.1371/JOURNAL.PCBI.1002540], http://hal.inria.fr/hal-00763496.

[27] S. SORN, T. SOK, S. LY, S. RITH, N. TUNG, A. VIARI, L. GAVOTTE, D. HOLL, H. SENG, N. ASGARI, B. RICHNER, D. LAURENT, N. CHEA, V. DUONG, T. TOYODA, C. Y. YASUDA, P. KITSUTANI, P. ZHOU, S. BING, V. DEUBEL, R. DONIS, R. FRUTOS, P. BUCHY. *Dynamic of H5N1 virus in Cambodia and emergence of a novel endemic sub-clade.*, in "Infection, Genetics and Evolution", June 2012 [*DOI :* 10.1016/J.MEEGID.2012.05.013], http://hal.inria.fr/hal-00763497.

[28] F. VAVRE, S. CHARLAT. *Making (good) use of Wolbachia: what the models say.*, in "Curr Opin Microbiol", June 2012, vol. 15, n^{o} 3, p. 263-8 [*DOI :* 10.1016/J.MIB.2012.03.005], http://hal.inria.fr/hal-00763517.

[29] C. VIEIRA, M. FABLET, E. LERAT, M. BOULESTEIX, R. REBOLLO, N. BURLET, A. AKKOUCHE, B. HUBERT, H. MORTADA, C. BIÉMONT. *A comparative analysis of the amounts and dynamics of transposable elements in natural populations of Drosophila melanogaster and Drosophila simulans.*, in "Journal of Environmental Radioactivity", November 2012, vol. 113, p. 83-6 [*DOI :* 10.1016/J.JENVRAD.2012.04.001], http://hal.inria.fr/hal-00763500.

International Conferences with Proceedings

[30] H. BAUMANN, P. CRESCENZI, P. FRAIGNIAUD. *Brief Announcement: Flooding in Dynamic Graphs with Arbitrary Degree Sequence*, in "26th International Symposium on Distributed Computing", Salvador, Brésil, 2012, p. 417-419, http://hal.inria.fr/hal-00763446.

[31] E. BIRMELÉ, P. CRESCENZI, R. FERREIRA, R. GROSSI, V. LACROIX, A. MARINO, N. PISANTI, G. A. T. SACOMOTO, M.-F. SAGOT. *Efficient Bubble Enumeration in Directed Graphs*, in "SPIRE", Cartagena, Colombie, Lecture Notes in Computer Science, 2012, vol. 7608, p. 118-129 [*DOI :* 10.1007/978-3-642-34109-0_13], http://hal.inria.fr/hal-00738927.

[32] S. CANZAR, M. EL-KEBIR, R. POOL, K. M. ELBASSIONI, A. K. MALDE, A. E. MARK, D. P. GEERKE, L. STOUGIE, G. W. KLAU. *Charge Group Partitioning in Biomolecular Simulation*, in "RECOMB", Lecture Notes in Computer Science, Springer, 2012, vol. 7262, p. 29-43.

[33] P. CRESCENZI, R. GROSSI, L. LANZI, A. MARINO. *On Computing the Diameter of Real-World Directed (Weighted) Graphs*, in "11th International Symposium on Experimental Algorithms", Bordeaux, France, 2012, p. 99-110, http://hal.inria.fr/hal-00763452.

[34] A. D. S. FREIRE, V. ACUÑA, P. CRESCENZI, C. E. FERREIRA, V. LACROIX, P. V. MILREU, E. MORENO, M.-F. SAGOT. *Minimum ratio cover of matrix columns by extreme rays of its induced cone*, in "Proceedings of the Second international conference on Combinatorial Optimization", Athens, Grèce, Lecture Notes in Computer Science, Springer-Verlag, 2012, vol. 7422, p. 165–177 [*DOI :* 10.1007/978-3-642-32147-4_16], http://hal.inria.fr/hal-00763455.

Scientific Books (or Scientific Book chapters)

[35] P. RIBECA, V. LACROIX, M. SAMMETH, R. GUIGÓ. *Analysis of RNA Transcripts by High-Throughput RNA Sequencing*, Wiley-VCH Verlag GmbH & Co. KGaA, 2012, p. 544–554 [*DOI :* 10.1002/9783527636778.CH50], http://hal.inria.fr/hal-00750967.