IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Lyon

Université Claude Bernard (Lyon 1)

Activity Report 2013

Project-Team BAMBOO

An algorithmic view on genomes, cells, and environments

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

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Computational Biology

Table of contents

Project-Team BAMBOO

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

Creation of the Team: 2009 January 01*, updated into Project-Team:* 2012 January 01.

1. Members

Research Scientists

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

Faculty Members

Hubert Charles [Insa-Lyon, Full Professor, HdR]
Christian Gautier [University Lyon 1, Professor, HdR]
Vincent Lacroix [University Lyon 1, Associate Professor]
Cristina Vieira [University Lyon 1, Professor, IUF, HdR]

Engineers

Alice Julien-Laferrière [Inria, ADT, since December 1st, 2012 and until November 30, 2013]
Camille Marchet [Inria, ADT, since October 1st, 2013]

PhD Students

Beatrice Donati [Inria, scholarship ERC, supervisors: Marie-France Sagot and Pierluigi Crescenzi]
Pierre-Antoine Farnier [Inria, scholarship ERC, supervisors: Marie-France Sagot, Laurence Mouton and Fabrice Vavre]
Mariana Ferrarini [Inria, scholarship ERC, supervisors: Marie-France Sagot and Arnaldo Zaha]
Susan Higashi [Inria, scholarship ERC, supervisors: Christian Gautier, Stefano Colella and Marie-France Sagot]
Alice Julien-Laferrière [Inria, scholarship BachBerry, started December 1st, 2013, supervisors: Susana Vinga and Marie-France Sagot]
Hélène Lopez [University Lyon 1, scholarship MENRT, started September 1st, 2013, supervisors: Cristina Vieira and Vincent Lacroix]
Cecília Coimbra Klein [Inria, scholarship ERC, supervisors: Marie-France Sagot and Ana Tereza Vasconcelos, defended in November 2013]
Paulo Vieira Milreu [Cnrs, scholarship ANR, supervisors: Vincent Lacroix, Christian Gautier and Marie-France Sagot, PhD defended December 19, 2012, postdoc until April 30, 2013]
Gustavo Akio Tominaga Sacomoto [Inria, scholarship ERC, supervisors: Marie-France Sagot, Vincent Lacroix and Pierluigi Crescenzi]
Martin Wannagat [Inria, scholarship ERC, supervisors: Marie-France Sagot, Leen Stougie and Alberto Marchetti-Spaccamela, started Jan. 2, 2013]

Post-Doctoral Fellows

Lilia Brinza [Inria, CDD ERC, until May 31st, 2013]
Christian Baudet [Cnrs, CDD ANR]
Matteo Brilli [Inria, CDD ERC, until September 30, 2012]
Emilie Chautard [Inria, granted by INSERM, started March 1st, 2013]
Ricardo Luiz de Andrade Abrantes [scholarship Science without Borders, Ministry of Science, Technology and Innovation and Ministry of Education, Brazil, started in April 1st, 2013]
Blerina Sinaimeri [Inria, CDD ERC, started February 1st, 2012]
Cecília Coimbra Klein [Inria, CDD ERC, since December 1st, 2013]

Visiting Scientists

Carlos Norberto Fischer [São Paulo State University, Rio Claro, Brazil, visit 3 months]
Maria Cristina Motta [Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, visit 15 days]
Susana Vinga [INESC-ID, IST Lisbon, Portugal, Professor, visit of 1 week]
Arnaldo Zaha [Federal University of Rio Grande do Sul, Porto Alegre, Brazil, visit 15 days]

Administrative Assistant

Florence Bouheddi [Inria]

Others

Pierluigi Crescenzi [University of Florence, Italy, Professor, external collaborator]
Manolo Gouy [Cnrs, Senior Researcher (DR), external collaborator, HdR]
Roberto Grossi [University of Pisa, Italy, Professor, external collaborator]
Gunnar Klau [CWI, Amsterdam, the Netherlands, Researcher, external collaborator]
Alberto Marchetti-Spaccamela [University La Sapienza Rome, Italy, Professor, external collaborator]
Anne Morgat [Swiss Institute of Bioinformatics, Scientist, external collaborator]
Nadia Pisanti [University of Pisa, Italy, Associate Professor, external collaborator]
Leen Stougie [Free University Amsterdam and CWI, Amsterdam, the Netherlands, Professor, external collaborator]
Ana Tereza Vasconcelos [Laboratório Nacional de Computação Científica, Petrópolis, Brazil, Senior Researcher, external collaborator]
Cyril Fournier [INSA, 4th year INSA-Lyon, from May 2013 until July 2013]
Hélène Lopez [University Lyon 1, Master, until August 2013]
Evgueni Jacob [University Lyon 1, L3, from February 2013 until Aug 2013]
Caio Padoan de Sa Godinho [University Lyon 1, Master, from May 2013 until July 2013]
Alexandre Rodolfo Da Silva [Inria, Master, from Feb 2013 until Aug 2013]

2. Overall Objectives

2.1. Highlights of the Year

BAMBOO is proposing the creation of a new Inria project team, ERABLE, that would replace BAMBOO. ERABLE would be a European Inria project team gathering the current members of BAMBOO, together with four researchers in Italy under the banner of the University of Rome La Sapienza (Alberto Marchetti-Spaccamela from La Sapienza, Pierluigi Crescenzi from the University of Florence, Roberto Grossi and Nadia Pisanti from the University of Pisa), and two researchers in the Netherlands under the banner of the CWI (Leen Stougie from the Free University of Amsterdam and the CWI, Gunnar Klau from the CWI). This proposal is currently been evaluated.

3. Research Program

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. Experimentally, 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 practically, (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. Domain

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

5. Software and Platforms

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. Coala

Participants: Christian Baudet [EPI, Contact, christian.baudet@univ-lyon1.fr], Pielrluigi Crescenzi, Bea Donati [EPI, Contact, bea.donati@inria.fr], Christian Gautier [EPI], Catherine Matias, Blerina Sinaimeri [EPI, Contact, blerina.sinaimeri@inria.fr], Marie-France Sagot [EPI, Contact, marie-france.sagot@inria.fr].

COALA stands for "CO-evolution Assessment by a Likelihood-free Approach". It is thus a likelihood-free method for the co-phylogeny reconstruction problem which is based on an Approximative Bayesian Computation (ABC).
http://coala.gforge.inria.fr/

5.5. C3Part & Isofun

Participants: Frédéric Boyer, Yves-Pol Deniélou, Anne Morgat [EPI, ext. member], Marie-France Sagot [EPI], Alain Viari [EPI, Contact, alain.viari@inria.fr].

The C3Part / Isofun package implements a generic approach to the local alignment of two or more graphs representing biological data, such as genomes, metabolic pathways or protein-protein interactions, in order to infer a functional coupling between them. It is based on the notion of "common connected components" between graphs.http://www.inrialpes.fr/helix/people/viari/lxgraph/index.html

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. Eucalypt

Participants: Christian Baudet [EPI, Contact, christian.baudet@univ-lyon1.fr], Pielrluigi Crescenzi, Bea Donati [Contact, bea.donati@inria.fr], Blerina Sinaimeri, Marie-France Sagot [EPI].

Algorithm for enumerating all optimal (possibly time-unfeasible) mappings of a parasite tree unto a host tree.
http://eucalypt.gforge.inria.fr/

5.8. Gobbolino & Touché

Participants: Vicente Acuña [EPI], Etienne Birmelé, 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, Contact, marie-france.sagot@inria.fr], Leen Stougie [EPI, ext. member].

Designed to solve the metabolic stories problem, which consists in finding all maximal directed acyclic subgraphs of a directed graph G whose sources and targets belong to a subset of the nodes of G, called the black nodes. Biologically, stories correspond to alternative metabolic pathways that may explain some stress that affected the metabolites corresponding to the black nodes by changing their concentration (measured by metabolomics experiments).
http://gforge.inria.fr/projects/gobbolino

5.9. 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/

5.10. kisSplice & KisSplice2igv

Participants: Lilia Brinza [EPI], Rayan Chikhi, Alice Julien-Lafferière [EPI], Janice Kielbassa, Vincent Lacroix [Contact, EPI], Camille Marchet [EPI], Claire Lemaitre, Pierre Peterlongo, Gustavo Sacomoto [EPI], Marie-France Sagot [EPI], Raluca Uricaru.

Enables to analyse RNA-seq data with or without a reference genome. It is an exact local transcriptome assembler, which can identify SNPs, indels and alternative splicing events. It can deal with an arbitrary number of biological conditions, and will quantify each variant in each condition. KISSPLICE2IGV is a pipeline that combines the outputs of KISSPLICE to a reference transcriptome (obtained with a full-length transcriptome assembler or a reference database). It provides a visualisation of the events found by KISSPLICE in a longer context using a genome browser (IGV).
http://kissplice.prabi.fr/

5.11. 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/?page_id=324

5.12. 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/

5.13. 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

5.14. Mirinho

Participants: Cyril Fournier [EPI], Susan Higashi [EPI, Contact, susan.higashi@inria.fr], Christian Gautier [EPI], Christine Gaspin, Marie-France Sagot [EPI].

Predicts, at a genome-wide scale, microRNA candidates.
http://mirinho.gforge.inria.fr/

5.15. 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/

5.16. 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 (downloadable version - offers different functionalities from the web-based version).
http://pbil.univ-lyon1.fr/software/motus/

5.17. 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/

5.18. PepLine

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

Pipeline for the high-throughput analysis of proteomic data.

5.19. 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/

5.20. 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/

5.21. 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.

5.22. 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

5.23. 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. Symbiont genome evolution and dynamics

The objective of this part of our work was to analyse genome rearrangements and dynamics. The results obtained were both algorithmic and biological.

In terms of algorithms, we developed a new method for repeat identification (RIME) [12], as well as an algorithm for finding the minimum number of three constrained versions of inversions that transform one given genome into another [25]. The constrained versions concerned symmetric, almost-symmetric and unitary inversions. The genome rearrangement algorithm is not exact: it is based on a greedy randomized search procedure to find such minimum number of constrained inversions.

The main set of biological results [4], [14] concerned trypanosomatids of the genera *Angomonas* and *Strigomonas* that live in a mutualistic association characterised by extensive metabolic cooperation with obligate endosymbiotic Betaproteobacteria. In contrast to their counterparts lacking symbionts, such trypanosomatids exhibit lower nutritional requirements and are autotrophic for essential amino acids and vitamins. Phylogenetic analyses showed that the cooperation in the first case is complemented by multiple horizontal gene transfers, from bacterial lineages to trypanosomatids, that appear to have occurred several times in the course of evolution. In contrast, but for three exceptions, such transfers are absent as concerns vitamin biosynthesis.

The above work was made possible in part because of the sequencing and annotation of the genomes whose metabolic pathways could then be inferred. We participated in these for some of the genomes involved in the above study [17].

6.2. Host-symbiont metabolic dialog

The methodological work done has covered one main question concerning what we called metabolic stories. Given a subset of metabolites representing those monitored as being under- or over-produced in some condition (e.g., interaction with a parasite) and a metabolic network represented as a compound graph, metabolic stories are maximal directed acyclic graphs (DAGs) that cover all the metabolites in the subset of interest, and have all sources and targets among these metabolites. One exact algorithm (TOUCHÉ, [24]) was developed to enumerate all metabolic stories that improved on our previous method (GOBBOLINO).

The algorithm above was validated on biological data [16] in a study of the response of yeast to cadmium exposure. We used this system as a proof of concept for our method and we showed that we are able to find a story that reproduces very well the current knowledge about the yeast response to cadmium. We further showed that this response is mostly based on enzyme activation. We also provided a framework for exploring the alternative pathways or side effects this local response is expected to have in the rest of the network. Finally, we discussed several interpretations for the changes we see and we suggest hypotheses which could in principle be experimentally tested.

6.3. Host-symbiont genetic dialog

Two sets of problems were addressed: (i) the development of algorithms for analysing NGS data especially RNAseq, and (ii) the development of algorithms for identifying small RNAs, notably microRNAs, and their targets.

The computational work on NGS is described in another section.

Computational work on small RNAs, initially miRNAs, led to the development of a new algorithmic method. This builds upon previously developed approaches, one which was applied to *Anopheles darlingi* for inferring miRNAs that however had a high rate of false positives, and a second that provided a way for navigating among all the candidates found. Recently however, we arrived at a better model for such inference in the double sense that the rate of false positives is smaller without losing in sensitivity, while the method is much faster. The paper presenting this work and the algorithm (MIRINHO) was submitted and is currently in revision.

6.4. Symbiont-host co-cladogenesis and co-evolution at the sequence and network levels

The problem here was to: (i) study the co-evolution of a set of hosts and their symbionts, and (ii) to understand the genetic architecture of a parasitic invasion by investigating the different phenotypes such invasion produces in the host.

Work on the first point took longer than initially planned but two papers are now submitted. In the first, titled “Co-phylogeny Reconstruction via an Approximate Bayesian Computation", we describe an algorithm (COALA) for estimating the frequency of co-evolutionary events based on a likelihood-free approach. The benefits of this method are twofold: (1) it provides more confidence in the set of costs to be used in a reconciliation, and (2) it allows to estimate the frequency of the events in cases where the dataset consists of trees with a large number of taxa. We evaluate our method on simulated and on real datasets. We show that in both cases, for a same pair of host and parasite trees, different sets of frequencies for the events constitute equally probable solutions. Moreover, sometimes these sets lead to different parsimonious optimal reconciliations, in the sense of presenting a different number of the events. For this reason, it appears crucial to take this into account before attempting any further biological interpretation of such reconciliations. More generally, we also show that the set of frequencies can vary widely depending on the input host and parasite trees. Indiscriminately applying a standard vector of costs may thus not be a good strategy.

In the second submitted paper related to the study of co-evolution and titled “EUCALYPT: Efficient tree reconciliation enumerator", we present a polynomial-delay algorithm for enumerating all optimal reconciliations. We show that in general many optimal solutions exist. We give an example where, for two pairs of host-parasite trees having each less than 40 leaves, the number of solutions is 2309, even when only time feasible solutions are kept. To facilitate their interpretation, those solutions are also classified in terms of the number of each event that they contain. This often enables to reduce considerably the number of different classes of solutions to examine further, but the number may remain high enough (16 for the same example). Depending on the cost vector, both numbers may increase considerably (for the same instance, to respectively 4080384 and 275).

Concerning the second question (genetic architecture of a parasitic invasion), one such phenotype is called “cytoplasmic incompatibility” (CI). Briefly, when a parasite invades a male host, it induces the death of the host’s offspring unless the female is also infected. This has been explained by a toxin/antitoxin model that involves a toxin deposited by the parasites in the male’s sperm inducing the death of the zygote unless neutralised by an antidote produced by the parasites in the egg. One toxin/antitoxin pair is usually linked to one genetic factor. Given a set of observed CIs, the question is how many genetic factors explain it. In its simplest form, this mathematically translates into, given a bipartite graph, finding its minimum biclique edge cover. One biclique corresponds to one factor. We had previously analysed the complexity of the problem and proposed an algorithm that was this year applied to a set of CI data from *Culex pipiens* [18].

6.5. NGS for biodiversity

In collaboration with the Laboratoire d’Écologie Alpine (LECA) at Grenoble where there is a strong expertise on DNA meta-barcoding, we had devised several tools for barcode design and analysis. ECOPRIMERS thus identified new barcode markers and their associated PCR primers within a DNA meta-barcoding approach. The algorithm was optimised two quality indexes measuring taxonomical range and discrimination to select the most efficient markers from a set of reference sequences, according to some experimental constraints such as marker length or specifically targeted taxa. We had also devised assembler algorithms directed to organelles (mitochondria or chloroplasts). This year, in collaboration with the Inria project-team MISTIS, we developed a statistical modelling approach to investigate the spatial cross-correlations between different taxa identified by meta-barcoding of soil sample from French Guiana (this was selected as a conference paper at the “45ème Journées de Statistiques” 2013 that took place at Toulouse and is organised by the Société Française de Statistique. This approach allows to visualise the co-occurrence pattern as a “species interaction graph”, and to study the mutual exclusion (competition) or inclusion (symbiosis) of different plant species.

6.6. NGS for genotypic variation detection

The computational work on NGS data concerned both algorithmic design and complexity analysis.

Based on the idea that each genotypic variation will correspond to a recognisable pattern in a de Bruijn graph constructed from a set of sequence reads, we had proposed a generic model for SNPs in DNA data, and then generalised it to the analysis of RNA. In this case, not only SNPs are present but also alternative splicing (AS) events, which, once again, generate a recognisable pattern in the de Bruijn Graph. We had therefore proposed a general model for all these variations (SNPs, indels and AS events) and introduced an exact algorithm (KISSPLICE) to extract all alternative splicing events. The algorithm also outputs candidate SNPs and indels. This year, we improved the algorithm [26]. As the problem relates to an old one in algorithmics (cycle enumeration), we also revisited it from a theoretical point of view [23].

The improved version of KISSPLICE [26] was used to analyse RNAseq data from two lines of *Asobara tabida* exhibiting different ovarian phenotypes in the absence of its endosymbiont *Wolbachia*. Although infected individuals of the two lines have similar phenotypes, numerous genes are differentially expressed between the two infected conditions. This could mean that two divergent strategies of tolerance have evolved. Preliminary results on the analysis of polymorphisms between these two lines suggest that differentially expressed genes tend to accumulate more variation. We are currently, via experiments done by the biologists in our team, testing the hypothesis that such genes are under strong selection pressure and may evolve through mutation accumulation, a process that could be related to assimilation.

A preliminary analysis of human data from the ENCODE project performed with KISSPLICE showed that an assembly-based method (without reference genome) is able to recover AS events that are missed by mapping-based methods (with a reference genome). Some of these events were experimentally validated, which represents the best type of proof we can provide to the biologists. The experimental part is made by our collaborator from the Inserm, Didier Auboeuf, in his team at the Centre National de Cancérologie of Lyon (CNCL), with whom we had an Inserm project, EXOMIC, funded for three years starting from 2012.

The identification of SNPs is also getting renewed interest even in the presence of a reference genome thanks to the possibility of re-sequencing many times the genome of a same or of very closely related species. The difficulty in the case of SNPs is to distinguish them from sequencing errors and from inexact repeats. We proposed a statistical test enabling to identify variations that are condition-specific, which enables to greatly enrich the list of potential SNP candidates. The paper on this test is in preparation. Its results as applied to the RNAseq data from two lines of *Asobara tabida* (see above) and to Drosophila species having diverged very recently were validated by, respectively, Fabrice Vavre and Cristina Vieira, both members of BAMBOO.

We also started addressing the problem that repeats (such as transposable elements for instance but not only) represent more in general for both local and global assemblers. We are thus developing a method that would enable to identify, in a de Bruijn graph built from RNAseq data, the vertices potentially corresponding to the borders of a repeated sequence. Preliminary results on simulated and real data show that the approach is promising (paper in preparation).

7. Partnerships and Cooperations

7.1. Regional Initiatives

- 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. Colib'read

- Title: Methods for efficient detection and visualization of biological information from non assembled NGS data
- Coordinator: P. Peterlomgo
- BAMBOO participant(s): V. Lacroix, A. Julien-Lafferière, C. Marchet, G. Sacomoto, M.-F. Sagot, B. Sinaimeri
- Type: ANR (2013-2016)
- Web page: http://colibread.inria.fr/

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. 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. BacHBerry

- Title: BACterial Hosts for production of Bioactive phenolics from bERRY fruits
- Coordinator: Jochen Förster (Novo Nordisk Foundation Center for Biosustainability (CFB), Copenhagen, Danemark)
- BAMBOO participant(s): V. Lacroix, Alice J. Lafferière, M.-F. Sagot, A. Viari, M. Wannagat
- Type: KBBE (2013-2016)
- Web page: Not yet available.

7.3.1.2. DroParCon

- Title: Drosophila parasitoid consortium
- Coordinator: Jochen Förster (Novo Nordisk Foundation Center for Biosustainability (CFB), Copenhagen, Danemark)
- BAMBOO participant(s): F. Vavre
- Type: PHC (2012-2014)
- Web page: http://www.droparcon.org.

7.3.1.3. 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

7.3.1.4. 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

7.3.1.5. 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: http://www.arimnet.net/index.php?p=fp_swipe

7.3.1.6. 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: http://www.2020-horizon.com/SYMBIOX-Role-of-the-oxidative-environment-in-the-stability-of-symbiotic-associations%28SYMBIOX%29-s4673.html

7.4. International Initiatives

7.4.1. Inria International Partners

Bamboo has an Inria International Partnership, called AMICI (see http://team.inria.fr/bamboo/amici/), with three partners in Italy (Universities of Rome “La Sapienza", Florence, and Pisa) and one in the Netherlands (Free University of Amsterdam / CWI). There are two unifying interests to all the projects of AMICI: algorithmics, and biology. At the present time, mostly because the current work of BAMBOO is centered on the ERC project SISYPHE (“Species Identity and SYmbiosis Formally and Experimentally explored"), the biology is very oriented to the general study, at the molecular level, of the symbiotic relation (genomics and other associated “omics", evolution, biochemical and interaction networks). This should evolve in future to extend the symbiotic study to either the ecological or a more health-oriented level, or to address new biology-related problems using mathematical modelling and techniques, and algorithmics.

7.4.2. Inria International Labs

BAMBOO participates in a project within the Inria-Chile CIRIC (Communication and Information Research and Innovation Center) titled “Omics Integrative Sciences”. The main objectives of the project are the development and implementation of mathematical and computational methods and the associated computational platforms for the exploration and integration of large sets of heterogeneous omics data and their application to the production of biomarkers and bioidentification systems for important Chilean productive sectors. The project started in 2011 and is coordinated in Chile by Alejandro Maass, Mathomics, University of Chile, Santiago.

7.4.3. Participation In other International Programs

BAMBOO is member 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 bIOinformatique") and is coordinated by Ana Tereza Vasconcelos from the LNCC and Marie-France Sagot from BAMBOO. The LIA was created in January 2012 for 4 years, renewable once. A preliminary web page for the LIA LIRIO is available at this address: http://team.inria.fr/bamboo/en/cnrs-lia-laboratoire-international-associe-lirio/.

BAMBOO has two other projects with Brazil. One is the Inria-Faperj project RAMPA (“Bioinformatics for the Reconstruction and Analysis of the Metabolism of PArasites") whose coordinators are M.-F. Sagot (France) and A. T. Vasconcelos (LNCC, Brazil). This project will finish at the end of 2013. Its main objective was to computationally and experimentally study the dialog between the trypanosomatids *Angomonas deanei* and *Strigomonas culicis* and their respective endosymbiont mainly at the metabolic level. The second project is the CAPES-COFECUB project titled: “Multidisciplinary Approach to the Study of the Biodiversity, Interactions and Metabolism of the Microbial Ecosystem of Swines". The coordinators are M.-F. Sagot (France) and A. T. Vasconcelos (LNCC, Brazil) with also the participation of Arnaldo Zaha (Federal University of Rio Grande do Sul. The project started in 2013 for 2 years, renewable once. The main objective of this project is to experimentally and mathematically explore the biodiversity of the bacterial organisms living in the respiratory tract of swines, many of which are pathogenic.

7.5. International Research Visitors

7.5.1. Visits of International Scientists

During 2013, the team had 4 international scientists visiting our group for at least one week. These included:

- Carlos Norberto Fischer, São Paulo State University, Rio Claro, Brazil, visit 3 months;
- Maria Cristina Motta, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, visit 15 days;
- Susana Vinga, INESC-ID, IST Lisbon, Portugal, visit of 1 week;
- Arnaldo Zaha, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, visit 15 days.

The above does not count the frequent visits of our external collaborators, members of the Inria International Partnership AMICI or of the LIA LIRIO.

7.5.2. Visits to International Teams

The visits to international teams were done mostly in the context of the Inria International Partnership AMICI, the LIA LIRIO, or the CIRIC project with Chile. Besides those, there were also visits of at least one week to Susana Vinga, INESC-ID, IST Lisbon, Portugal.

8. Dissemination

8.1. Scientific Animation

- Hubert Charles: He is director of studies of the "Bioinformatique et Modélisation (BIM)" track at the Insa-Lyon, co-director of the Biosciences Department of the Insa-Lyon, and co-director of the Doctoral School E2M2 of the Université Lyon 1.
- Marie-France Sagot: She is, since 2010, member of the Board of Directors of the ACM Special Interest Group on Bioinformatics, Computational Biology, and Biomedical Informatics (SIGBio). She is, since 2011, 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 (SIF). She is, since 2012, in the Steering Committee of the Labex Ecofect of the University of Lyon.
- Fabrice Vavre: He was elected in 2012 member of the Section 29 of the CoNRS. He is, since 2012, in the Steering Committee of the Labex Ecofect of the University of Lyon.
- Alain Viari: Since 2011, he is a member of the Scientific Advisory Board ("Conseil Scientifique (COS)") for the Inria Grenoble Rhône-Alpes Research Center. Since 2012, he is Deputy Scientific Director at Inria responsible for the ICST for Life and Environmental Sciences. He thus represents Inria at several national instances related to Life Sciences and Health and is member of a number of scientific advisory boards (IMMI (Institut de Microbiologie et Maladies Infectieuses / Aviesan); IRT (Institut de Recherche Technologique) BioAster; Genopôle-Evry). Since 2013 he is also the French coordinator of the Bioinformatics working group in the U.S.-France Joint Commission on Science and Technology Cooperation.
- Cristina Vieira: She is director of the GDRE "Comparative genomics" since the GDRE was renewed in 2010.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

The members of BAMBOO teach both at the Department of Biology of the University of Lyon (in particular within the MIV (Mathematics and Computer Science for the Life Sciences) specialty) and at the department of Bioinformatics of the Insa (National Institute of Applied Sciences). Cristina Vieira is responsible for the Evolutionary Genetics and Genomics academic career of the Master Ecosciences-Microbiology. She was awarded an IUF (Institut Universitaire de France) distinction and teaches genetics 64 hours per year at the University and École Normale Supérieure. Hubert Charles is responsible for the Master of Modelling and Bioinformatics (BIM) at the Insa of Lyon. He teaches 192 hours per year in statistics and biology. Vincent Lacroix is responsible for several courses both at the University (L2 Bioinformatics, L3 Advanced Bioinformatics) and at the Insa (M1: Gene Expression, M2:Introduction to Bioinformatics for Biochemists). He teaches 192 hours per year, except in 2012-2013 where he taught 96 hours as he had a partial sabbatical funded by Inria (through the ERC AdG Sisyphe grant) for one semester. He teaches bioinformatics and statistics.

Two PhD students in BAMBOO were able to teach by applying to a monitorat (mentoring program – 64 hours per year). Cecilia Klein taught 64 hours in 2012-2013 at the Department of Biology (L2: Bioinformatics). Two other PhD students taught more occasionally in 2012-2013: Gustavo Sacomoto taught 12 hours of graph algorithms at the Department of Biology (M1: Bioinformatics), and Mariana Ferrarini taught 23 hours at the Department of Biology (L3: Bioinformatics). The postdocs are also involved in teaching. Blerina Sinaimeri thus taught 8 hours of graph algorithms, Lilia Brinza 40 hours of bioinformatics at the Insa, and Christian Baudet taught 12 hours of bioinformatics at the University.

All members of the BAMBOO team are affiliated to the doctoral school E2M2 (Ecology-Evolution-Microbiology-Modelling). Hubert Charles is the vice-director of the school.

8.2.2. Supervision

The following is the PhD defended in BAMBOO in 2013.

PhD: Cecilia Klein, University of Lyon 1, November 12, supervisors A. T. Vasconcelos and M.-F. Sagot.

8.2.3. Juries

M.-F. Sagot: Reviewer of the HDR of Andrei Zinovyev (ENS Paris), and of the PhD of Kalle Karhu (University of Helsinki, Finland).

C. Vieira: Reviewer of the PhD of Jonathan Grandaubert (University Paris 11), member of the PhD committee of Eugenia Pessia (University of Lyon), reviewer of the HDR of Karine Alix (AgroParisTech).

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] C. C. KLEIN. , *Une étude bioinformatique du dialogue métabolique entre un trypanosome non pathogène et son endosymbiote à des buts évolutifs et fonctionnels*, Université Claude Bernard - Lyon I, November 2013, http://hal.inria.fr/tel-00923295

Articles in International Peer-Reviewed Journals

[2] A. AKKOUCHE, T. GRENTZINGER, M. FABLET, C. ARMENISE, N. BURLET, V. BRAMAN, S. CHAMBEYRON, C. VIEIRA. *Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells*, in "EMBO reports", May 2013, vol. 14, n^{o} 5, pp. 458-64 [*DOI :* 10.1038/EMBOR.2013.38], http://hal.inria.fr/hal-00850319

[3] R. ALIKIITEAGA GUTIÉRREZ, A. VIARI, B. GODELLE, R. FRUTOS, P. BUCHY. *Biased mutational pattern and quasispecies hypothesis in H5N1 virus*, in "Infection, Genetics and Evolution", 2013, vol. 15, pp. 69-76 [*DOI :* 10.1016/J.MEEGID.2011.10.019], http://hal.inria.fr/pasteur-00853145

[4] J. ALVES, C. KLEIN, F. DA SILVA, A. COSTA-MARTINS, M. SERRANO, G. BUCK, A. T. VASCONCELOS, M.-F. SAGOT, M. TEIXEIRA, M. C. MOTTA, E. CAMARGO. *Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers*, in "BMC Evolutionary Biology", 2013, vol. 13, n^{o} 1, 190 p. [*DOI :* 10.1186/1471-2148-13-190], http://hal.inria.fr/hal-00868763

[5] S. BERTHOUMIEUX, M. BRILLI, D. KAHN, H. DE JONG, E. CINQUEMANI. *On the identifiability of metabolic network models*, in "Journal of Mathematical Biology", 2013, vol. 67, n° 6-7, pp. 1795-1832 [*DOI :* 10.1007/s00285-012-0614-x], http://hal.inria.fr/hal-00762620

[6] M. BRILLI, P. LIÒ, V. LACROIX, M.-F. SAGOT. *Short and long-term genome stability analysis of prokaryotic genomes*, in "BMC Genomics", 2013, vol. 14, n° 1, 309 p. [*DOI :* 10.1186/1471-2164-14-309], http://hal.inria.fr/hal-00834426

[7] L. BRINZA, F. CALEVRO, H. CHARLES. *Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of Buchnera*, in "BMC Genomics", 2013, vol. 14, n° 1, 73 p. [*DOI :* 10.1186/1471-2164-14-73], http://hal.inria.fr/hal-00788161

[8] C. M. A. CARARETO, E. H. HERNANDEZ, C. VIEIRA. *Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling Drosophila species*, in "Gene", December 2013, pp. 93-99 [*DOI :* 10.1016/j.gene.2013.11.080], http://hal.inria.fr/hal-00922703

[9] D. COSTECHAREYRE, J.-F. CHICH, J.-M. STRUB, Y. RAHBÉ, G. CONDEMINE. *Transcriptome of Dickeya dadantii infecting Acyrthosiphon pisum reveals a strong defense against antimicrobial peptides*, in "PLoS ONE", 2013, vol. 8, n° 1 [*DOI :* 10.1371/journal.pone.0054118], http://hal.inria.fr/hal-00850312

[10] V. EYRAUD, L. KARAKI, I. RAHIOUI, C. SIVIGNON, P. DA SILVA, Y. RAHBÉ, C. ROYER, F. GRESSENT. *Expression and Biological Activity of the Cystine Knot Bioinsecticide PA1b (Pea Albumin 1 Subunit b).*, in "PLoS ONE", 2013, vol. 8, n° 12 [*DOI :* 10.1371/journal.pone.0081619], http://hal.inria.fr/hal-00921788

[11] M. FABLET, A. AKKOUCHE, V. BRAMAN, C. VIEIRA. *Variable expression levels detected in the Drosophila effectors of piRNA biogenesis*, in "Gene", December 2013, pp. 149-153 [*DOI :* 10.1016/j.gene.2013.11.095], http://hal.inria.fr/hal-00922704

[12] M. FEDERICO, P. PETERLONGO, N. PISANTI, M.-F. SAGOT. *Rime: Repeat identification*, in "Discrete Applied Mathematics", March 2013 [*DOI :* 10.1016/j.dam.2013.02.016], http://hal.inria.fr/hal-00802023

[13] M. GIACOFCI, S. LAMBERT-LACROIX, G. MAROT, F. PICARD. *Wavelet-based clustering for mixed-effects functional models in high dimension*, in "Biometrics", March 2013, vol. 69, n° 1, pp. 31-40 [*DOI :* 10.1111/j.1541-0420.2012.01828.x], http://hal.inria.fr/hal-00782458

[14] C. C. KLEIN, J. M. P. ALVES, M. G. SERRANO, G. A. BUCK, A. T. VASCONCELOS, M.-F. SAGOT, M. TEIXEIRA, E. P. CAMARGO, M. C. M. MOTTA. *Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses*, in "PLoS ONE", 2013, vol. 8, n° 11 [*DOI :* 10.1371/journal.pone.0079786], http://hal.inria.fr/hal-00922587

[15] A. MERVILLE, S. VENNER, H. HENRI, A. VALLIER, F. MENU, F. VAVRE, A. HEDDI, M.-C. BEL-VENNER. *Endosymbiont diversity among sibling weevil species competing for the same resource*, in "BMC Evolutionary Biology", 2013, vol. 13, 28 p. [*DOI :* 10.1186/1471-2148-13-28], http://hal.inria.fr/hal-00850306

[16] P. V. MILREU, C. C. KLEIN, L. COTTRET, V. ACUÑA, E. BIRMELÉ, M. BORASSI, C. JUNOT, A. MARCHETTI-SPACCAMELA, A. MARINO, L. STOUGIE, F. JOURDAN, P. CRESCENZI, V.

LACROIX, M.-F. SAGOT. *Telling metabolic stories to explore metabolomics data: a case study on the yeast response to cadmium exposure*, in "Bioinformatics", January 2014, vol. 30, n° 1, pp. 61-70 [*DOI :* 10.1093/BIOINFORMATICS/BTT597], http://hal.inria.fr/hal-00922567

[17] M. C. M. MOTTA, A. C. D. A. MARTINS, S. S. DE SOUZA, C. M. C. CATTA-PRETA, R. SILVA, C. C. KLEIN, L. G. P. DE ALMEIDA, O. DE LIMA CUNHA, L. P. CIAPINA, M. BROCCHI, A. C. COLABARDINI, B. DE ARAUJO LIMA, C. R. MACHADO, C. M. DE ALMEIDA SOARES, C. M. PROBST, C. B. A. DE MENEZES, C. E. THOMPSON, D. C. BARTHOLOMEU, D. F. GRADIA, D. P. PAVONI, E. C. GRISARD, F. FANTINATTI-GARBOGGINI, F. K. MARCHINI, G. F. RODRIGUES-LUIZ, G. WAGNER, G. H. GOLDMAN, J. L. R. FIETTO, M. C. ELIAS, M. H. S. GOLDMAN, M.-F. SAGOT, M. PEREIRA, P. H. STOCO, R. P. DE MENDONÇA-NETO, S. M. R. TEIXEIRA, T. E. F. MACIEL, T. A. DE OLIVEIRA MENDES, T. P. ÜRMÉNYI, W. DE SOUZA, S. SCHENKMAN, A. T. R. DE VASCONCELOS. *Predicting the proteins of Angomonas deanei, Strigomonas culicis and their respective endosymbionts reveals new aspects of the trypanosomatidae family*, in "PLoS ONE", 2013, vol. 8, n° 4 [*DOI :* 10.1371/JOURNAL.PONE.0060209], http://hal.inria.fr/hal-00846833

[18] I. NOR, J. ENGELSTÄDTER, O. DURON, M. REUTER, M.-F. SAGOT, S. CHARLAT. *On the genetic architecture of cytoplasmic incompatibility: inference from phenotypic data*, in "American Naturalist", July 2013, vol. 182, n° 1, pp. 15-24 [*DOI :* 10.1086/670612], http://hal.inria.fr/hal-00837003

[19] E. PRESTAT, S. R. DE MORAIS, J. A. VENDRELL, A. THOLLET, C. GAUTIER, P. A. COHEN, A. AUSSEM. *Learning the local Bayesian network structure around the ZNF217 oncogene in breast tumours*, in "Computers in Biology and Medicine", May 2013, vol. 43, n° 4, pp. 334-341 [*DOI :* 10.1016/J.COMPBIOMED.2012.12.002], http://hal.inria.fr/hal-00851231

[20] A. RABATEL, G. FEBVAY, K. GAGET, G. DUPORT, P. BAA-PUYOULET, P. SAPOUNTZIS, N. BENDRIDI, M. REY, Y. RAHBÉ, H. CHARLES, F. CALEVRO, S. COLELLA. *Tyrosine pathway regulation is host-mediated in the pea aphid symbiosis during late embryonic and early larval development*, in "BMC Genomics", 2013, vol. 14, n° 1, 235 p. [*DOI :* 10.1186/1471-2164-14-235], http://hal.inria.fr/hal-00824527

[21] O. REY, A. ESTOUP, B. FACON, A. LOISEAU, A. AEBI, O. DURON, F. VAVRE, J. FOUCAUD. *Distribution of endosymbiotic reproductive manipulators reflects invasion process and not reproductive system polymorphism in the little fire ant Wasmannia auropunctata*, in "PLoS ONE", 2013, vol. 8, n° 3 [*DOI :* 10.1371/JOURNAL.PONE.0058467], http://hal.inria.fr/hal-00850303

[22] M. VALIZADEH, C. DERAISON, S. KAZEMITABAR, Y. RAHBÉ, M. A. JONGSMA. *Aphid Resistance in Florist's Chrysanthemum (Chrysanthemum morifolium Ramat.) Induced by Sea Anemone Equistatin Overexpression*, in "African Journal of Biotechnology", December 2013, pp. 6922-6930 [*DOI :* 10.5897/AJB2013.12956], http://hal.inria.fr/hal-00919389

International Conferences with Proceedings

[23] E. BIRMELÉ, R. FERREIRA, R. GROSSI, A. MARINO, N. PISANTI, R. RIZZI, G. SACOMOTO. *Optimal Listing of Cycles and st-Paths in Undirected Graphs*, in "24th Annual ACM-SIAM Symposium on Discrete Algorithms", New Orleans, United States, 2013, http://hal.inria.fr/hal-00751345

[24] M. BORASSI, P. CRESCENZI, V. LACROIX, A. MARINO, M.-F. SAGOT, P. VIEIRA MILREU. *Telling Stories Fast*, in "12th International Symposium Experimental Algorithms (SEA)", Rome, Italy, Lecture Notes in Computer Science, 2013, vol. 7933, pp. 200-211, http://hal.inria.fr/hal-00835201

[25] U. DIAS, C. BAUDET, Z. DIAS. *Greedy Randomized Search Procedure to Sort Genomes using Symmetric, Almost-Symmetric and Unitary Inversions*, in "Proceedings of the 4th ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB 2013)", Maryland, United States, J. GAO (editor), ACM Special Interest Group on Bioinformatics, Computational Biology, and Biomedical Informatics, 2013, pp. 181–190 [*DOI :* 10.1145/2506583.2506614], http://hal.inria.fr/hal-00922670

[26] G. A. T. SACOMOTO, V. LACROIX, M.-F. SAGOT. *A polynomial delay algorithm for the enumeration of bubbles with length constraints in directed graphs and its application to the detection of alternative splicing in RNA-seq data*, in "Workshop on Algorithms in Bioinformatics (WABI)", Nice, France, Lecture Notes in Computer Science, Springer-Verlag, 2013, http://hal.inria.fr/hal-00842982

[27] K. SALIKHOV, G. SACOMOTO, G. KUCHEROV. *Using cascading Bloom filters to improve the memory usage for de Brujin graphs*, in "Workshop on Algorithms in Bioinformatics", Sophia Antipolis, France, Lecture Notes in Computer Science, Springer, 2013, vol. 8126, 13 p. , submitted, http://hal.inria.fr/hal-00824697

National Conferences with Proceedings

[28] A. BRETAUDEAU, O. DAMERON, F. LEGEAI, Y. RAHBÉ. *AphidAtlas : avancées récentes*, in "Proceedings of BAPOA 2013 MOP. INRA, CIRAD Lavalette Campus Montpellier, France", Montpellier, France, M. UZEST (editor), 2013, [In French], http://hal.inria.fr/hal-00913155

[29] A. STUDENY, F. FORBES, E. COISSAC, A. VIARI, C. MERCIER, L. ZINGER, A. BONIN, F. BOYER, P. TABERLET. *Spatial modelling of plant diversity from high-throughput environmental dna sequence data*, in "45èmes Journées de Statistique", Toulouse, France, Société Française de Statistique, May 2013, http://hal.inria.fr/hal-00933699