IN PARTNERSHIP WITH:
Centrum Wiskunde & Informatica

Institut national des sciences appliquées de Lyon

Université Claude Bernard (Lyon 1)

Université de Rome la Sapienza

Activity Report 2018

Project-Team ERABLE

European Research team in Algorithms and Biology, formaL and Experimental

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

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Computational Biology

Table of contents

Project-Team ERABLE

Creation of the Team: 2015 January 01, updated into Project-Team: 2015 July 01

ERABLE is a European Inria team gathering French researchers together with researchers in Italy under the banner of the Sapienza University of Rome and researchers in the Netherlands under the banner of the CWI.

Keywords:

Computer Science and Digital Science:

A3. - Data and knowledge
A3.1. - Data
A3.1.1. - Modeling, representation
A3.1.4. - Uncertain data
A3.3. - Data and knowledge analysis
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A7. - Theory of computation
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A8.7. - Graph theory
A8.8. - Network science
A8.9. - Performance evaluation

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1. - Biology
B1.1.1. - Structural biology
B1.1.2. - Molecular and cellular biology
B1.1.4. - Genetics and genomics
B1.1.6. - Evolutionnary biology
B1.1.7. - Bioinformatics
B1.1.10. - Systems and synthetic biology
B2. - Health
B2.2. - Physiology and diseases
B2.2.3. - Cancer
B2.2.4. - Infectious diseases, Virology
B2.3. - Epidemiology

1. Team, Visitors, External Collaborators

Research Scientists

Marie-France Sagot [Inria, Team leader, Senior Researcher, HDR]
Blerina Sinaimeri [Inria, Researcher]
Fabrice Vavre [CNRS, Researcher, HDR]
Alain Viari [Inria, Senior Researcher]

Alexander Schönhuth [CWI, The Netherlands, Senior Researcher, also since Oct 2017 part-time Professor at Univ of Utrecht]

Faculty Members

Hubert Charles [INSA Lyon, Associate Professor, HDR]
Roberto Grossi [Univ Pisa, Italy, Professor]
Pierluigi Crescenzi [Univ Florence, Florence, Italy, Professor]
Giuseppe Francesco Italiano [Univ Tor Vergata, Rome, Italy, Professor]
Vincent Lacroix [Univ de Claude Bernard, Associate Professor]
Alberto Marchetti Spaccamela [Sapienza Univ Rome, Italy, Professor]
Arnaud Mary [Univ de Claude Bernard, Associate Professor]
Nadia Pisanti [Univ Pisa, Italy, Associate Professor]
Andrea Marino [Univ Pisa, Italy, Assistant Professor]
Leen Stougie [CWI & Free Univ Amsterdam, The Netherlands, Professor]
Cristina Vieira [Univ de Claude Bernard, Professor, HDR]

Post-Doctoral Fellows

Ricardo de Andrade Abrantes [Univ São Paulo, Brazil, and Univ Claude Bernard Lyon]
Alex Di Genova [Inria]
Mariana Galvão Ferrarini [Insa Lyon & Univ Claude Bernard Lyon, Mar 2018]
Clara Benoît-Pilven [Univ Claude Bernard Lyon, from Oct 2018]

PhD Students

Marianne Borderes [MaatPharma and Claude Bernard Lyon, since Feb 2018]
Audric Cologne [Inria]
Mattia Gastaldello [Sapienza Univ of Rome and Univ Claude Bernard Lyon, until Feb 2018]
Leandro Ishi Soares de Lima [CNPq Brazil & Univ Claude Bernard Lyon]
Carol Moraga Quinteros [Conicyt & Univ Claude Bernard Lyon]
Henri Pusa [Inria]
Camille Sessegolo [Univ Claude Bernard Lyon]
Yishu Wang [Univ Claude Bernard Lyon]
Irene Ziska [Inria]

Technical staff

Clara Benoît-Pilven [INSERM, until Sep 2018]
Martin Wannagat [Inria, until Feb 2018]

Interns

Louis Duchemin [Inria, from Apr 2018 until Jul 2018]
Nicolas Homberg [Univ de Claude Bernard, from Oct 2018]

Administrative Assistant

Claire Sauer [Inria]

External Collaborators

Laurent Jacob [LBBE UMR5558, Researcher, external collaborator]
Susana Vinga [IST Lisbon, Researcher, external collaborator]

2. Overall Objectives

2.1. Overall Objectives

Cells are seen as the basic structural, functional and biological units of all living systems. They represent the smallest units of life that can replicate independently, and are often referred to as the building blocks of life. Living organisms are then classified into unicellular ones – this is the case of most bacteria and archea – or multicellular – this is the case of animals and plants. Actually, multicellular organisms, such as for instance

human, may be seen as composed of native (human) cells, but also of extraneous cells represented by the diverse bacteria living inside the organism. The proportion in the number of the latter in relation to the number of native cells is believed to be high: this is for example of 90% in humans. Multicellular organisms have thus been described also as "superorganisms with an internal ecosystem of diverse symbiotic microbiota and parasites" (Nicholson *et al.*, Nat Biotechnol, 22(10):1268-1274, 2004)) where symbiotic means that the extraneous unicellular organisms (cells) live a close, and in this case, long-term relation both with the multicellular organisms they inhabit and among themselves. On the other hand, bacteria sometimes group into colonies of genetically identical individuals which may acquire both the ability to adhere together and to become specialised for different tasks. An example of this is the cyanobacterium *Anabaena sphaerica* who may group to form filaments of differentiated cells, some – the heterocysts – specialised for nitrogen fixation while the others are capable of photosynthesis. Such filaments have been seen as first examples of multicellular patterning.

At its extreme, one could then see life as one collection, or a collection of collections of genetically identical or distinct self-replicating cells who interact, sometimes closely and for long periods of evolutionary time, with same or distinct functional objectives. The interaction may be at equilibrium, meaning that it is beneficial or neutral to all, or it may be unstable meaning that the interaction may be or become at some time beneficial only to some and detrimental to other cells or collections of cells. The interaction may involve other living systems, or systems that have been described as being at the edge of life such as viruses, or else genetic or inorganic material such as, respectively, transposable elements and chemical compounds.

The application goal of ERABLE is, through the use of mathematical models and algorithms, to better understand such close and often persistent interactions, with a longer term objective of becoming able in some cases to suggest the means of controlling for or of re-establishing equilibrium in an interacting community by acting on its environment or on its players, how they play and who plays. This goal requires to identify who are the partners in a closely interacting community, who is interacting with whom, how and by which means. Any model is a simplification of reality, but once selected, the algorithms to explore such model should address questions that are precisely defined and, whenever possible, be exact in the answer as well as exhaustive when more than one exists in order to guarantee an accurate interpretation of the results within the given model. This fits well the mathematical and computational expertise of the team, and drives the methodological goal of ERABLE which is to substantially and systematically contribute to the field of exact enumeration algorithms for problems that most often will be hard in terms of their complexity, and as such to also contribute to the field of combinatorics in as much as this may help in enlarging the scope of application of exact methods.

The key objective is, by constantly crossing ideas from different models and types of approaches, to look for and to infer "patterns", as simple and general as possible, either at the level of the biological application or in terms of methodology. This objective drives which biological systems are considered, and also which models and in which order, going from simple discrete ones first on to more complex continuous models later if necessary and possible.

3. Research Program

3.1. Two main goals

ERABLE has two main goals, one related to biology and the other to methodology (algorithms, combinatorics, statistics). In relation to biology, the main goal of ERABLE is to contribute, through the use of mathematical models and algorithms, to a better understanding of close and often persistent interactions between "collections of genetically identical or distinct self-replicating cells" which will correspond to organisms/species or to actual cells. The first will cover the case of what has been called symbiosis, meaning when the interaction involves different species, while the second will cover the case of a (cancerous) tumour which may be seen as a collection of cells which suddenly disrupts its interaction with the other (collections of) cells in an organism by starting to grow uncontrollably.

Such interactions are being explored initially at the molecular level. Although we rely as much as possible on already available data, we intend to also continue contributing to the identification and analysis of the main genomic and systemic (regulatory, metabolic, signalling) elements involved or impacted by an interaction, and how they are impacted. We started going to the population and ecological levels by modelling and analysing the way such interactions influence, and are or can be influenced by the ecosystem of which the "collections of cells" are a part. The key steps are:

- identifying the molecular elements based on so-called omics data (genomics, transcriptomics, metabolomics, proteomics, etc.): such elements may be gene/proteins, genetic variations, (DNA/RNA/protein) binding sites, (small and long non coding) RNAs, etc.
- simultaneously inferring and analysing the network that models how these molecular elements are physically and functionally linked together for a given goal, or find themselves associated in a response to some change in the environment;
- modelling and analysing the population and ecological network formed by the "collections of cells in interaction", meaning modelling a network of networks (previously inferred or as already available in the literature).

One important longer term goal of the above is to analyse how the behaviour and dynamics of such a network of networks might be controlled by modifying it, including by subtracting some of its components from the network or by adding new ones.

In relation to methodology, the main goal is to provide those enabling to address our main biological objective as stated above that lead to the best possible interpretation of the results within a given pre-established model and a well defined question. Ideally, given such a model and question, the method is exact and also exhaustive if more than one answer is possible. Three aspects are thus involved here: establishing the model within which questions can and will be put; clearly defining such questions; exactly answering to them or providing some guarantee on the proximity of the answer given to the "correct" one. We intend to continue contributing to these three aspects:

- at the modelling level, by exploring better models that at a same time are richer in terms of the information they contain (as an example, in the case of metabolism, using hypergraphs as models for it instead of graphs) and are susceptible to an easier treatment:
  - these two objectives (rich models that are at the same time easy to treat) might in many cases be contradictory and our intention is then to contribute to a fuller characterisation of the frontiers between the two;
  - even when feasible, the richer models may lack a full formal characterisation (this is for instance the case of hypergraphs) and our intention is then to contribute to such a characterisation;
- at the question level, by providing clear formalisations of those that will be raised by our biological concerns;
- at the answer level:
  - to extend the area of application of exact algorithms by: (i) a better exploration of the combinatorial properties of the models, (ii) the development of more efficient data structures, (iii) a smarter traversal of the space of solutions when more than one solution exists;
  - when exact algorithms are not possible, or when there is uncertainty in the input data to an algorithm, to improve the quality of the results given by a deeper exploration of the links between different algorithmic approaches: combinatorial, randomised, stochastic.

3.2. Different research axes

The goals of the team are biological and methodological, the two being intrinsically linked. Any division into axes along one or the other aspect or a combination of both is thus somewhat artificial. Following the evaluation of the team at the end of 2017, four main axes were identified, with the last one being the more recently added one. This axis is specifically oriented towards health in general, human or animal. The first three axes are: genomics, metabolism and post-transcriptional regulation, and (c)evolution.

Notice that the division itself is based on the biological level (genomic, metabolic/regulatory, evolutionary) or main current Life Science purpose (health) rather than on the mathematical or computational methodology involved. Any choice has its part of arbitrariness. Through the one we made, we wished to emphasise the fact that the area of application of ERABLE is important for us. *It does not mean that the mathematical and computational objectives are not equally important*, but only that those are, most often, motivated by problems coming from or associated to the general Life Science goal. Notice that such arbitrariness also means that some Life Science topics will be artificially split into two different Axes. One example of this is genomics and the two main health areas currently addressed that are intrinsically inter-related for now.

Axis 1: Genomics

Intra and inter-cellular interactions involve molecular elements whose identification is crucial to understand what governs, and also what might enable to control such interactions. For the sake of clarity, the elements may be classified in two main classes, one corresponding to the elements that allow the interactions to happen by moving around or across the cells, and another that are the genomic regions where contact is established. Examples of the first are non coding RNAs, proteins, and mobile genetic elements such as (DNA) transposons, retro-transposons, insertion sequences, etc. Examples of the second are DNA/RNA/protein binding sites and targets. Furthermore, both types (effectors and targets) are subject to variation across individuals of a population, or even within a single (diploid) individual. Identification of these variations is yet another topic that we wish to cover. Variations are understood in the broad sense and cover single nucleotide polymorphisms (SNPs), copy-number variants (CNVs), repeats other than mobile elements, genomic rearrangements (deletions, duplications, insertions, inversions, translocations) and alternative splicings (ASs). All three classes of identification problems (effectors, targets, variations) may be put under the general umbrella of genomic functional annotation.

Axis 2: Metabolism and post-transcriptional regulation

As increasingly more data about the interaction of molecular elements (among which those described above) becomes available, these should then be modelled in a subsequent step in the form of networks. This raises two main classes of problems. The first is to accurately infer such networks. Assuming such a network, integrated or "simple", has been inferred for a given organism or set of organisms, the second problem is then to develop the appropriate mathematical models and methods to extract further biological information from such networks.

The team has so far concentrated its efforts on two main aspects concerning such interactions: metabolism and post-transcriptional regulation by small RNAs. The more special niche we have been exploring in relation to metabolism concerns the fact that the latter may be seen as an organism's immediate window into its environment. Finely understanding how species communicate through those windows, or what impact they may have on each other through them is thus important when the ultimate goal is to be able to model communities of organisms, for understanding them and possibly, on a longer term, for control. While such communication has been explored in a number of papers, most do so at a too high level or only considered couples of interacting organisms, not larger communities. The idea of investigating consortia, and in the case of synthetic biology, of using them, has thus started being developed in the last decade only, and was motivated by the fact that such consortia may perform more complicated functions than could single populations, as well as be more robust to environmental fluctuations. Another originality of the work that the team has been doing in the last decade has also been to fully explore the combinatorial aspects of the structures used (graphs or directed hypergraphs) and of the associated algorithms. As concerns post-transcriptional regulation, the team has essentially been exploring the idea that small RNAs may have an important role in the dialog between different species.

Axis 3: (Co)Evolution

Understanding how species that live in a close relationship with others may (co)evolve requires understanding for how long symbiotic relationships are maintained or how they change through time. This may have deep implications in some cases also for understanding how to control such relationships, which may be a way of controlling the impact of symbionts on the host, or the impact of the host on the symbionts and on the environment (by acting on its symbiotic partner(s)). These relationships, also called *symbiotic associations*, have however not yet been very widely studied, at least not at a large scale.

One of the problems is getting the data, meaning the trees for hosts and symbionts but even prior to that, determining with which symbionts the present-day hosts are associated (or are "infected" by as may be the term used in some contexts) which is a big enterprise in itself. The other problem is measuring the stability of the association. This has generally been done by concomitantly studying the phylogenies of hosts and symbionts, that is by doing what is called a *cophylogeny* analysis, which itself is often realised by performing what is called a *reconciliation* of two phylogenetic trees (in theory, it could be more than two but this is a problem that has not yet been addressed by the team), one for the symbionts and one for the hosts with which the symbionts are associated. This consists in mapping one of the trees (usually, the symbiont tree) to the other. Cophylogeny inherits all the difficulties of phylogeny, among which the fact that it is not possible to check the result against the "truth" as this is now lost in the past. Cophylogeny however also brings new problems of its own which are to estimate the frequency of the different types of events that could lead to discrepant evolutionary histories, and to estimate the duration of the associations such events may create.

Axis 4: Human, animal and plant health
As indicated above, this is a recent axis in the team and concerns various applications to human and animal health. In some ways, it overlaps with the three previous axes as well as with Axis 5 on the methodological aspects, but since it gained more importance in the past few years, we decided to develop more these particular applications. Most of them started through collaborations with clinicians. Such applications are currently focused on three different topics: (i) Infectiology, (ii) Rare diseases, and (iii) Cancer.

Infectiology is the oldest one. It started by a collaboration with Arnaldo Zaha from the Federal University of Rio Grande do Sul in Brazil that focused on pathogenic bacteria living inside the respiratory tract of swines. Since our participation in the H2020 ITN MicroWine, we started interested in infections affecting plants this time, and more particularly vine plants. Rare Diseases on the other hand started by a collaboration with clinicians from the Centre de Recherche en Neurosciences of Lyon (CNRL) and is focused the Taybi-Linder Syndrome (TALS) and on abnormal splicing of U12 introns, while Cancer rests on a collaboration with the Centre Léon Bérard (CLB) and Centre de Recherche en Cancérologie of Lyon (CRCL) which is focused on Breast and Prostate carcinomas and Gynaecological carcinosarcomas.

The latter collaboration was initiated through a relationship between a member of ERABLE (Alain Viari) and Dr. Gilles Thomas who had been friends since many years. G. Thomas was one of the pioneers of Cancer Genomics in France. After his death in 2014, Alain Viari took the (part time) responsibility of his team at CLB and pursued the main projects he had started.

Within Inria and beyond, the first two applications (Infectiology and Rare Diseases) may be seen as unique because of their specific focus (resp. respiratory tract of swines / vine plants on one hand, and TALS on the other). In the first case, such uniqueness is also related to the fact that the work done involves a strong computational part but also experiments *performed within ERABLE itself*.

4. Application Domains

4.1. Biology and Health

The main areas of application of ERABLE are: (1) biology understood in its more general sense, with a special focus on symbiosis and on intracellular interactions, and (2) health with a special emphasis for now on infectious diseases, rare diseases, and cancer.

5. New Software and Platforms

5.1. C3Part/Isofun

KEYWORDS: Bioinformatics - Genomics

FUNCTIONAL DESCRIPTION: 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.

- Participants: Alain Viari, Anne Morgat, Frédéric Boyer, Marie-France Sagot and Yves-Pol Deniélou
- Contact: Alain Viari
- URL: http://www.inrialpes.fr/helix/people/viari/lxgraph/index.html

5.2. Cassis

KEYWORDS: Bioinformatics - Genomics
FUNCTIONAL DESCRIPTION: Implements methods for the precise detection of genomic rearrangement breakpoints.

- Participants: Christian Baudet, Christian Gautier, Claire Lemaitre, Eric Tannier and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: http://pbil.univ-lyon1.fr/software/Cassis/

5.3. Coala

CO-evolution Assessment by a Likelihood-free Approach
KEYWORDS: Bioinformatics - Evolution
SCIENTIFIC DESCRIPTION: Despite an increasingly vaster literature on cophylogenetic reconstructions for studying host-parasite associations, understanding the common evolutionary history of such systems remains a problem that is far from being solved. Many of the most used algorithms do the host-parasite reconciliation analysis using an event-based model, where the events include in general (a subset of) cospeciation, duplication, loss, and host-switch. All known event-based methods then assign a cost to each type of event in order to nd a reconstruction of minimum cost. The main problem with this approach is that the cost of the events strongly inuence the reconciliation obtained.

To deal with this problem, we developed an algorithm, called Coala, for estimating the frequency of the events based on an approximate Bayesian computation approach.
FUNCTIONAL DESCRIPTION:COALA stands for "COevolution Assessment by a Likelihood-free Approach". It is thus a likelihood-free method for the co-phylogeny reconstruction problem which is based on an Approximate Bayesian Computation (ABC) approach.

- Participants: Beatrice Donati, Blerina Sinaimeri, Catherine Matias, Christian Baudet, Christian Gautier, Marie-France Sagot and Pierluigi Crescenzi
- Contact: Blerina Sinaimeri
- URL: http://coala.gforge.inria.fr/

5.4. CSC

KEYWORDS: Genomics - Algorithm
FUNCTIONAL DESCRIPTION: Given two sequences x and y, CSC (which stands for Circular Sequence Comparison) finds the cyclic rotation of x (or an approximation of it) that minimises the blockwise q-gram distance from y.

- Contact: Nadia Pisanti
- URL: https://github.com/solonas13/csc

5.5. Cycads

KEYWORDS: Systems Biology - Bioinformatics

FUNCTIONAL DESCRIPTION: Annotation database system to ease the development and update of enriched BIOCYC databases. CYCADS allows the integration of the latest sequence information and functional annotation data from various methods into a metabolic network reconstruction. Functionalities will be added in future to automate a bridge to metabolic network analysis tools, such as METEXPLORE. CYCADS was used to produce a collection of more than 22 arthropod metabolism databases, available at ACYPICYC (http://acypicyc.cycadsys.org) and ARTHROPODACYC (http://arthropodacyc.cycadsys.org). It will continue to be used to create other databases (newly sequenced organisms, Aphid biotypes and symbionts...).

- Participants: Augusto Vellozo, Hubert Charles, Marie-France Sagot and Stefano Colella
- Contact: Hubert Charles
- URL: http://www.cycadsys.org/

5.6. DBGWAS

KEYWORDS: Graph algorithmics - Genomics
FUNCTIONAL DESCRIPTION:DBGWAS is a tool for quick and efficient bacterial GWAS. It uses a compacted De Bruijn Graph (cDBG) structure to represent the variability within all bacterial genome assemblies given as input. Then cDBG nodes are tested for association with a phenotype of interest and the resulting associated nodes are then re-mapped on the cDBG. The output of DBGWAS consists of regions of the cDBG around statistically significant nodes with several informations related to the phenotypes, offering a representation helping in the interpretation. The output can be viewed with any modern web browser, and thus easily shared.

- Contact: Leandro Ishi Soares De Lima
- URL: https://gitlab.com/leoisl/dbgwas

5.7. Eucalypt

KEYWORDS: Bioinformatics - Evolution
FUNCTIONAL DESCRIPTION:EUCALYPT stands for "EnUmerator of Coevolutionary Associations in PoLYnomial-Time delay". It is an algorithm for enumerating all optimal (possibly time-unfeasible) mappings of a symbiont tree unto a host tree.

- Participants: Beatrice Donati, Blerina Sinaimeri, Christian Baudet, Marie-France Sagot and Pierluigi Crescenzi
- Contact: Blerina Sinaimeri
- URL: http://eucalypt.gforge.inria.fr/

5.8. Fast-SG

KEYWORDS: Genomics - Algorithm - NGS
FUNCTIONAL DESCRIPTION:FAST-SG enables the optimal hybrid assembly of large genomes by combining short and long read technologies.

- Contact: Alex Di Genova
- URL: https://github.com/adigenova/fast-sg

5.9. Gobbolino-Touché

KEYWORDS: Bioinformatics - Graph algorithmics - Systems Biology
FUNCTIONAL DESCRIPTION: 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.

- Participants: Etienne Birmelé, Fabien Jourdan, Ludovic Cottret, Marie-France Sagot, Paulo Vieira Milreu, Pierluigi Crescenzi, Vicente Acuna Aguayo and Vincent Lacroix
- Contact: Marie-France Sagot
- URL: http://gforge.inria.fr/projects/gobbolino

5.10. HapCol

KEYWORDS: Bioinformatics - Genomics
FUNCTIONAL DESCRIPTION: A fast and memory-efficient DP approach for haplotype assembly from long reads that works until 25x coverage and solves a constrained minimum error correction problem exactly.

- Contact: Nadia Pisanti
- URL: http://hapcol.algolab.eu/

5.11. HgLib

HyperGraph Library
KEYWORDS: Graph algorithmics - Hypergraphs
FUNCTIONAL DESCRIPTION: The open-source library hglib is dedicated to model hypergraphs, which are a generalisation of graphs. In an *undirected* hypergraph, an hyperedge contains any number of vertices. A *directed* hypergraph has hyperarcs which connect several tail and head vertices. This library, which is written in C++, allows to associate user defined properties to vertices, to hyperedges/hyperarcs and to the hypergraph itself. It can thus be used for a wide range of problems arising in operations research, computer science, and computational biology.

RELEASE FUNCTIONAL DESCRIPTION: Initial version

- Participants: Martin Wannagat, David Parsons, Arnaud Mary and Irene Ziska
- Contact: Arnaud Mary
- URL: https://gitlab.inria.fr/kirikomics/hglib

5.12. KissDE

KEYWORDS: Bioinformatics - NGS
FUNCTIONAL DESCRIPTION:KISSDE is an R Package enabling to test if a variant (genomic variant or splice variant) is enriched in a condition. It takes as input a table of read counts obtained from an NGS data pre-processing and gives as output a list of condition-specific variants.

RELEASE FUNCTIONAL DESCRIPTION: This new version improved the recall and made more precise the size of the effect computation.

- Participants: Camille Marchet, Aurélie Siberchicot, Audric Cologne, Clara Benoît-Pilven, Janice Kielbassa, Lilia Brinza and Vincent Lacroix
- Contact: Vincent Lacroix
- URL: http://kissplice.prabi.fr/tools/kissDE/

5.13. KisSplice

KEYWORDS: Bioinformatics - Bioinfirmatics search sequence - Genomics - NGS
FUNCTIONAL DESCRIPTION: 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.

RELEASE FUNCTIONAL DESCRIPTION: Improvements : KissReads module has been modified and sped up, with a significant impact on run times. Parameters : –timeout default now at 10000: in big datasets, recall can be increased while run time is a bit longer. Bugs fixed : Reads containing only 'N': the graph construction was stopped if the file contained a read composed only of 'N's. This is was a silence bug, no error message was produced. Problems compiling with new versions of MAC OSX (10.8+): KisSplice is now compiling with the new default C++ compiler of OSX 10.8+.

- Participants: Alice Julien-Laferrière, Leandro Ishi Soares De Lima, Vincent Miele, Rayan Chikhi, Pierre Peterlongo, Camille Marchet, Gustavo Akio Tominaga Sacomoto, Marie-France Sagot and Vincent Lacroix
- Contact: Vincent Lacroix
- URL: http://kissplice.prabi.fr/

5.14. KisSplice2RefGenome

KEYWORDS: Bioinformatics - NGS - Transcriptomics
FUNCTIONAL DESCRIPTION:KISSPLICE identifies variations in RNA-seq data, without a reference genome. In many applications however, a reference genome is available. KISSPLICE2REFGENOME enables to facilitate the interpretation of the results of KISSPLICE after mapping them to a reference genome.

- Participants: Audric Cologne, Camille Marchet, Camille Sessegolo, Alice Julien-Laferrière and Vincent Lacroix
- Contact: Vincent Lacroix
- URL: http://kissplice.prabi.fr/tools/kiss2refgenome/

5.15. KisSplice2RefTranscriptome

KEYWORDS: Bioinformatics - NGS - Transcriptomics
FUNCTIONAL DESCRIPTION:KISSPLICE2REFTRANSCRIPTOME enables to combine the output of KISSPLICE with the output of a full length transcriptome assembler, thus allowing to predict a functional impact for the positioned SNPs, and to intersect these results with condition-specific SNPs. Overall, starting from RNA-seq data only, we obtain a list of condition-specific SNPs stratified by functional impact.

- Participants: Helene Lopez Maestre, Mathilde Boutigny and Vincent Lacroix
- Contact: Vincent Lacroix
- URL: http://kissplice.prabi.fr/tools/kiss2rt/

5.16. MetExplore

KEYWORDS: Systems Biology - Bioinformatics
SCIENTIFIC DESCRIPTION: MetExplore stores metabolic networks of 160 organisms into a relational database. Information about metabolic networks mainly come from BioCyc-like databases. Two BioCyc-like databases contain information about several organisms: PlantCyc and MetaCyc. MetExplore contains also the information about metabolites stored in Metabolome.jp. Note that there is no information about reactions in this database and is only useful to identify compounds from masses. Several genome-scale models designed for Flux Balance Analysis have also been imported into MetExplore. The table below gives details about the sources of the metabolic networks present in MetExplore.
FUNCTIONAL DESCRIPTION: Web-server that allows to build, curate and analyse genome-scale metabolic networks. METEXPLORE is also able to deal with data from metabolomics experiments by mapping a list of masses or identifiers onto filtered metabolic networks. Finally, it proposes several functions to perform Flux Balance Analysis (FBA). The web-server is mature, it was developed in PHP, JAVA, Javascript and Mysql. METEXPLORE was started under another name during Ludovic Cottret's PhD in Bamboo, and is now maintained by the METEXPLORE group at the Inra of Toulouse.

- Participants: Fabien Jourdan, Hubert Charles, Ludovic Cottret and Marie-France Sagot
- Contact: Fabien Jourdan
- URL: https://metexplore.toulouse.inra.fr/index.html/

5.17. Mirinho

KEYWORDS: Bioinformatics - Computational biology - Genomics - Structural Biology
FUNCTIONAL DESCRIPTION: Predicts, at a genome-wide scale, microRNA candidates.

- Participants: Christian Gautier, Christine Gaspin, Cyril Fournier, Marie-France Sagot and Susan Higashi
- Contact: Marie-France Sagot
- URL: http://mirinho.gforge.inria.fr/

5.18. Momo

Multi-Objective Metabolic mixed integer Optimization
KEYWORDS: Metabolism - Metabolic networks - Multi-objective optimisation
FUNCTIONAL DESCRIPTION:MOMO is a multi-objective mixed integer optimisation approach for enumerating knockout reactions leading to the overproduction and/or inhibition of specific compounds in a metabolic network.

- Contact: Marie-France Sagot
- URL: http://momo-sysbio.gforge.inria.fr

5.19. MultiPus

KEYWORDS: Systems Biology - Algorithm - Graph algorithmics - Metabolic networks - Computational biology
SCIENTIFIC DESCRIPTION: Synthetic biology has boomed since the early 2000s when it started being shown that it was possible to efficiently synthetise compounds of interest in a much more rapid and effective way by using other organisms than those naturally producing them. However, to thus engineer a single organism, often a microbe, to optimise one or a collection of metabolic tasks may lead to difficulties when attempting to obtain a production system that is efficient, or to avoid toxic effects for the recruited microorganism. The idea of using instead a microbial consortium has thus started being developed in the last decade. This was motivated by the fact that such consortia may perform more complicated functions than could single populations and be more robust to environmental fluctuations. Success is however not always guaranteed. In particular, establishing which consortium is best for the production of a given compound or set thereof remains a great challenge. The algorithm MultiPus is based on an initial model that enables to propose a consortium to synthetically produce compounds that are either exogenous to it, or are endogenous but where interaction among the species in the consortium could improve the production line.
FUNCTIONAL DESCRIPTION:MULTIPUS (for "MULTIple species for the synthetic Production of Useful biochemical Substances") is an algorithm that, given a microbial consortium as input, identifies all optimal sub-consortia to synthetically produce compounds that are either exogenous to it, or are endogenous but where interaction among the species in the sub-consortia could improve the production line.

- Participants: Alberto Marchetti-Spaccamela, Alice Julien-Laferrière, Arnaud Mary, Delphine Parrot, Laurent Bulteau, Leen Stougie, Marie-France Sagot and Susana Vinga
- Contact: Marie-France Sagot
- URL: http://multipus.gforge.inria.fr/

5.20. Pitufolandia

KEYWORDS: Bioinformatics - Graph algorithmics - Systems Biology
FUNCTIONAL DESCRIPTION: The algorithms in PITUFOLANDIA (PITUFO / PITUFINA / PAPAPITUFO) are designed to solve the minimal precursor set problem, which consists in finding all minimal sets of precursors (usually, nutrients) in a metabolic network that are able to produce a set of target metabolites.

- Contact: Marie-France Sagot
- URL: http://gforge.inria.fr/projects/pitufo/

5.21. Sasita

KEYWORDS: Bioinformatics - Graph algorithmics - Systems Biology
FUNCTIONAL DESCRIPTION:SASITA is a software for the exhaustive enumeration of minimal precursor sets in metabolic networks.

- Contact: Marie-France Sagot
- URL: http://sasita.gforge.inria.fr/

5.22. Savage

KEYWORDS: Algorithm - Genomics
FUNCTIONAL DESCRIPTION: Reconstruction of viral quasi species without using a reference genome.

- Contact: Alexander Schonhuth
- URL: https://bitbucket.org/jbaaijens/savage

5.23. Smile

KEYWORDS: Bioinformatics - Genomic sequence
FUNCTIONAL DESCRIPTION: Motif inference algorithm taking as input a set of biological sequences.

- Participant: Marie-France Sagot
- Contact: Marie-France Sagot

5.24. Rime

KEYWORDS: Bioinformatics - Genomics - Sequence alignment
FUNCTIONAL DESCRIPTION: Detects long similar fragments occurring at least twice in a set of biological sequences.

- Contact: Nadia Pisanti

5.25. Totoro & Kotoura

KEYWORDS: Bioinformatics - Graph algorithmics - Systems Biology
FUNCTIONAL DESCRIPTION: Both TOTORO and KOTOURA decipher the reaction changes during a metabolic transient state, using measurements of metabolic concentrations. These are called metabolic hyperstories. TOTORO (for TOpological analysis of Transient metabOlic RespOnse) is based on a qualitative measurement of the concentrations in two steady-states to infer the reaction changes that lead to the observed differences in metabolite pools in both conditions. In the currently available release, a pre-processing and a post-processing steps are included. After the post-processing step, the solutions can be visualised using DINGHY (http://dinghy.gforge.inria.fr). KOTOURA (for Kantitative analysis Of Transient metabOlic and regUlatory Response And control) infers quantitative changes of the reactions using information on measurement of the metabolite concentrations in two steady-states.

- Contact: Marie-France Sagot
- URL: http://hyperstories.gforge.inria.fr/

5.26. WhatsHap

KEYWORDS: Bioinformatics - Genomics
FUNCTIONAL DESCRIPTION:WHATSHAP is a DP approach for haplotype assembly from long reads that works until 20x coverage and solves the minimum error correction problem exactly. PWHATSHAP is a parallelisation of the core dynamic programming algorithm of WHATSHAP.

- Contact: Nadia Pisanti
- URL: https://bitbucket.org/whatshap/whatshap

6. New Results

6.1. General comments

We present in this section the main results obtained in 2018.

We tried to organise these along the four axes as presented above. Clearly, in some cases, a result obtained overlaps more than one axis. In such case, we chose the one that could be seen as the main one concerned by such results.

We did not indicate here the results on more theoretical aspects of computer science if it did not seem for now that they could be relevant in contexts related to computational biology. Actually, those on string [32], [33], [36], [11] and graph algorithms in general [2], [35], [38], [37], [39], [41], [54], [42], [44], [43], [40], [47], [5], [45], [48], [49], [53], [23], [24], or on more general algorithmic problems notably related to data structures are already relevant for life sciences (biology or ecology) or in the future could become more specifically so. We do in particular believe that dynamic graph approaches could be of great interest in the future for some of the enumeration problems we constantly meet in biology.

A few other results of 2018 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised [52], or the work represented a survey *e.g.* [21]). Likewise, also for space reasons, we do not detail the results presented in some biological papers of the team when these did not require a mathematical or algorithmical input or are surveys [1], [4], [7], [8], [12], [19], [20], [22], [25], [31].

On the other hand, we do mention a couple of works that were submitted towards the end of 2018.

6.2. Axis 1: Genomics

Genome hybrid assembly
Long read sequencing technologies are considered to be the solution for handling genome repeats, allowing near reference-level reconstructions of large genomes. However, long read de novo assembly pipelines are computationally intense and require a considerable amount of coverage, thereby hindering their broad application to the assembly of large genomes. Alternatively, hybrid assembly methods that combine short and long read sequencing technologies can reduce the time and cost required to produce de novo assemblies of large genomes. In [10], we proposed a new method, called FAST-SG, that uses a new ultrafast alignment-free algorithm specifically designed for constructing a scaffolding graph using lightweight data structures. FAST-SG can construct the graph from either short or long reads. This allows the reuse of efficient algorithms designed for short read data and permits the definition of novel modular hybrid assembly pipelines. Using comprehensive standard datasets and benchmarks, we showed how FAST-SG outperforms the state-of-the-art short read aligners when building the scaffolding graph and can be used to extract linking information from either raw or error-corrected long reads. We also showed how a hybrid assembly approach using FAST-SG with shallow long-read coverage (5X) and moderate computational resources can produce long-range and accurate reconstructions of the genomes of *Arabidopsis thaliana* (Ler-0) and human (NA12878). We are currently working on the assembly process itself, using the scaffolding graphs obtained with FAST-SG. The results obtained so far are extremely promising and a paper is currently in preparation. This is part of the work done by Alex di Genova, postdoc in ERABLE.

Variant annotation

Genome-wide analyses estimate that more than 90% of multi exonic human genes produce at least two transcripts through a genomic variant called alternative splicing (AS). Various bioinformatics methods are available to analyse AS from RNAseq data. Most methods start by mapping the reads to an annotated reference genome, but some start by a *de novo* assembly of the reads. In [3], we presented a systematic comparison of a mapping-first approach (FARLINE) and an assembly-first approach (scKisSplice). We applied these methods to two independent RNAseq datasets and found that the predictions of the two pipelines overlapped (70% of exon skipping events were common), but with noticeable differences. The assembly-first approach allowed to find more novel variants, including novel unannotated exons and splice sites. It also predicted AS in recently duplicated genes. The mapping-first approach allowed to find more lowly expressed splicing variants, and splice variants overlapping repeats. This work demonstrated that annotating AS with a single approach leads to missing out a large number of candidates, many of which are differentially regulated across conditions and can be validated experimentally. We therefore advocate for the combined use of both mapping-first and assembly-first approaches for the annotation and differential analysis of AS from RNAseq datasets. This was part of the work of Clara Benoît-Pilven, postdoc at Inserm and in ERABLE, to which also participated other current or ex-members of ERABLE, namely Camille Marchet (during her stay as ADT engineer with ERABLE), Emilie Chautard (when she was postdoc Inserm and in ERABLE), Gustavo Sacomoto (when he was PhD and then for one year postdoc in ERABLE), and Leandro Lima (current PhD student of ERABLE).

Another type of variant, namely SNPs was also considered in [51]. In this paper, mutations are detected by eBWT (extended Burrows-Wheeler Transform). Indeed, we notices that eBWT of a collection of DNA fragments tend to cluster together the copies of nucleotides sequenced from a genome. We showed that it is thus possible to accurately predict how many copies of any nucleotide are expected inside each such cluster, and that a precise LCP array based procedure can locate these clusters in the eBWT. These theoretical insights were validated in practice with SNPs being clustered in the eBWT of a reads collection. We developed a tool for finding SNPs with a simple scan of the eBWT and LCP arrays. Preliminary results show that our method requires much less coverage than the state-of-the-art tools while drastically improving precision and sensitivity.

Both types of variants correspond to special types of st-paths in graphs, a topic that was also explored from a more purely theoretical point of view in two papers, one already accepted [46] and on that is about to be submitted and extends the results obtained in 2017 on bubble (as st-paths are also called in bioinformatics) generators in directed graphs.

Full-length *de novo* viral quasispecies assembly through variation graph construction
Viruses populate their hosts as a viral quasispecies: a collection of genetically related mutant strains. Viral quasispecies assembly refers to reconstructing the strain-specific haplotypes from read data, and predicting their relative abundances within the mix of strains, an important step for various treatment-related reasons. Reference-genome-independent ("de novo") approaches have yielded benefits over reference-guided approaches, because reference-induced biases can become overwhelming when dealing with divergent strains. While being very accurate, extant *de novo* methods only yield rather short contigs. It remains to reconstruct full-length haplotypes together with their abundances from such contigs. In [34], we first constructed a variation graph, a recently popular, suitable structure for arranging and integrating several related genomes, from the short input contigs, without making use of a reference genome. To obtain paths through the variation graph that reflect the original haplotypes, we solved a minimisation problem that yields a selection of maximal-length paths that is optimal in terms of being compatible with the read coverages computed for the nodes of the variation graph. We output the resulting selection of maximal length paths as the haplotypes, together with their abundances. Benchmarking experiments on challenging simulated data sets showed significant improvements in assembly contiguity compared to the input contigs, while preserving low error rates. As a consequence, our method outperforms all state-of-the-art viral quasispecies assemblers that aim at the construction of full-length haplotypes, in terms of various relevant assembly measures. The tool, called VIRUS-VG, is availabe at https://bitbucket.org/jbaaijens/virus-vg.

A member of ERABLE was also involved in the Second Annual Meeting of the European Virus Bioinformatics Center (EVBC), held in Utrecht, Netherlands, and whose focus was on computational approaches in virology, with topics including (but not limited to) virus discovery, diagnostics, (meta-)genomics, modeling, epidemiology, molecular structure, evolution, and viral ecology. Approximately 120 researchers from around the world attended the meeting this year. An overview of new developments and novel research findings that emerged during the meeting was published in the journal *Viruses* [16].

Bacterial genome-wide association studies (GWAS)
Genome-wide association study (GWAS) methods applied to bacterial genomes have shown promising results for genetic marker discovery or detailed assessment of marker effect. Recently, alignment-free methods based on k-mer composition have proven their ability to explore the accessory genome. However, they lead to redundant descriptions and results which are sometimes hard to interpret. In [17], we introduced DBGWAS, an extended k-mer-based GWAS method producing interpretable genetic variants associated with distinct phenotypes. Relying on compacted de Bruijn graphs (cDBG), our method gathers cDBG nodes, identified by the association model, into subgraphs defined from their neighbourhood in the initial cDBG. DBGWAS is alignment-free and only requires a set of contigs and phenotypes. In particular, it does not require prior annotation or reference genomes. It produces subgraphs representing phenotype-associated genetic variants such as local polymorphisms and mobile genetic elements (MGE). It offers a graphical framework which helps interpret GWAS results. Importantly, it is also computationally efficient (the experiments took one hour and a half on average). We validated our method using antibiotic resistance phenotypes for three bacterial species. DBGWAS recovered known resistance determinants such as mutations in core genes in *Mycobacterium tuberculosis*, and genes acquired by horizontal transfer in *Staphylococcus aureus* and *Pseudomonas aeruginosa* along with their MGE context. It also enabled us to formulate new hypotheses involving genetic variants not yet described in the antibiotic resistance literature. This is part of the work of Magali Jaillard, PhD student of Laurent Jacob who is an external collaborator of ERABLE, and of Leandro I. S. de Lima, PhD student co-supervised by three members of ERABLE.

6.3. Axis 2: Metabolism and post-transcriptional regulation

Multi-objective metabolic mixed integer optimisation: with an application to yeast strain engineering
In a paper submitted and already available in bioRxiv (https://www.biorxiv.org/content/early/2018/11/22/476689), we explored the concept of multi-objective optimisation in the field of metabolic engineering when both continuous and integer decision variables are involved in the model. In particular, we proposed a multi-objective model which may be used to suggest reaction deletions that maximise and/or minimise several functions simultaneously. The applications may include, among others, the concurrent maximisation of a bioproduct and of biomass, or maximisation of a bioproduct while minimising the formation of a given by-product, two common requirements in microbial metabolic engineering. Production of ethanol by the widely used cell factory *Saccharomyces cerevisiae* was adopted as a case study to demonstrate the usefulness of the proposed approach in identifying genetic manipulations that improve productivity and yield of this economically highly relevant bioproduct. We did an *in vivo* validation and we could show that some of the predicted deletions exhibit increased ethanol levels in comparison with the wild-type strain. The multi-objective programming framework we developed, called MOMO, is open-source and uses POLYSCIP as underlying multi-objective solver. This is part of the work of Ricardo de Andrade, postdoc at Unversity of São Paulo with Roberto Marcondes, and in ERABLE. It is joint work with Susana Vinga, external collaborator of ERABLE and partner of the Inria Associated Team Compasso.

Metabolic shifts

With the increasing availability of so-called 'omics data – transcriptomics, proteomics, and metabolics – there has been growing interest in various ways of integrating them with the metabolic network. When the network is represented by a graph, 'omics data can guide the extraction of subnetworks of interest to find metabolic pathways or sets of related genes. Within the framework of constraint-based modelling, 'omics data can be used to improve the prediction of metabolic behaviour and to build context-specific metabolic models. One interesting application of metabolic reconstructions in conjunction with 'omics data is to use the two to understand metabolic shifts. When an organism encounters a change in environmental conditions, often a re-organisation of metabolism follows. Comparative measurements of gene expression and metabolite concentrations can be used to gain insight into these changes but this data is "structureless", meaning it lacks the information about how the metabolic components relate to each other. A metabolic network on the other hand contains this information, and can thus greatly benefit such an analysis. We developed a new method, called MOOMIN, that combines the results of a differential expression analysis comparing the gene expression levels in two different conditions with a metabolic network to produce a hypothesis of a metabolic shift. The idea is to use the network structure to define feasible global changes in metabolism. These changes are then scored based on the gene expression data with the goal of finding the change that best agrees with the observations. Finding the best-scoring change is formulated into an optimisation problem that can be solved using Mixed-Integer Linear Programming. This is part of the work of Henri Taneli Pusa, co-supervised by 3 members of ERABLE, whose manuscript was submitted to the reviewers and who should be defending his PhD in early February 2019. The paper on MOOMIN will be submitted soon, and the software then made available. Participated also in this work Mariana G. Ferrarini, postdoc at Insa and in ERABLE, and Ricardo Andrade, postdoc at University of São Paulo with Roberto Marcondes and in ERABLE.

Metabolic games
The PhD of Taneli also investigated game theory in the context of metabolism. Game theory is a branch of applied mathematics that deals with interacting rational agents with conflicting goals. When rationality is replaced with natural selection, *evolutionary* game theory can be used to explain the "decisions" taken by even microscopic organisms. The PhD manuscript presents the idea of a *metabolic game*, a game theoretical model for the prediction of metabolic behaviour. In contrast to Flux Balance Analysis, where the metabolic state is predicted using simple optimisation, a metabolic game takes into account the fact that optimality is influenced by the surrounding members of a microbial community. By changing the availability of nutrients, or secreting beneficial or harmful molecules, microbes essentially create their own environment and make optimal behaviour context-specific. A paper is submitted that reviews the literature that has applied game theory to the study of microbes, with a focus on metabolism and especially games derived using metabolic networks and constraint-based modelling. In the PhD manuscript, Taneli further explains the idea behind a metabolic game and discusses different aspects of defining such games: the choice of players, actions, and payoffs.

6.4. Axis 3: (Co)Evolution

Exploring the robustness of the parsimonious reconciliation method in host-symbiont cophylogeny
Following our previous work on reconciliation methods for cophylogeny, in [29], we explored the robustness of the parsimonious host-symbiont tree reconciliation method under editing or small perturbations of the input. The editing involved making different choices of unique symbiont mapping to a host in the case where multiple associations exist. This is made necessary by the fact that the tree reconciliation model is currently unable to handle such associations. The analysis performed could however also address the problem of errors. The perturbations were re-rootings of the symbiont tree to deal with a possibly wrong placement of the root specially in the case of fast-evolving species. In order to do this robustness analysis, we introduced a simulation scheme specifically designed for the host-symbiont cophylogeny context, as well as a measure to compare sets of tree reconciliations, both of which are of interest by themselves. This work was also part of the PhD of a previous student of ERABLE, Laura Urbini.

Geometric medians in reconciliation spaces

Recently, there has been much interest in studying spaces of tree reconciliations (as used in cophylogenetic studies), which arise by defining some metric d on the set $\mathcal{R}(P, H, \phi)$ of all possible reconciliations between two trees P and H where ϕ represents the map between the leaf-sets of P and H (corresponding to present-day associations). In [14], we studied the following question: how do we compute a *geometric median* for a given subset Ψ of $\mathcal{R}(P, H, \phi)$ relative to d, *i.e.* an element $\psi_{med} \in \mathcal{R}(P, H, \phi)$ such that

$$\sum_{\psi' \in \Psi} d(\psi_{med}, \psi') \leq \sum_{\psi' \in \Psi} d(\psi, \psi')$$

holds for all $\psi \in \mathcal{R}(P, H, \phi)$? For a model where so-called host-switches or transfers are not allowed, and for a commonly used metric d called the *edit-distance*, we showed that although the cardinality of $\mathcal{R}(P, H, \phi)$ can be super-exponential, it is still possible to compute a geometric median for a set Ψ in $\mathcal{R}(P, H, \phi)$ in polynomial time. We expect that this result could be useful for computing a summary or consensus for a set of reconciliations (*e.g.* for a set of suboptimal reconciliations). The collaboration with Katharina Huber and Vincent Moulton from the School of Computing Sciences at the University of New Anglia was made possible by a Royal Society Grant obtained by the two partners (UNA and ERABLE).

Exploring and Visualising Spaces of Tree Reconciliations
A common approach to tree reconciliation involves specifying a model that assigns costs to certain events, such as cospeciation, and then tries to find a mapping between two specified phylogenetic trees which minimises the total cost of the implied events. For such models, it has been shown, including by the ERABLE members in previous papers, that there may be a huge number of optimal solutions, or at least solutions that are close to optimal. It is therefore of interest to be able to systematically compare and visualise whole collections of reconciliations between a specified pair of trees. In [13] , we considered various metrics on the set of all possible reconciliations between a pair of trees, some that have been defined before but also new metrics that we proposed. We showed that the diameter for the resulting spaces of reconciliations can in some cases be determined theoretically, information that we used to normalise and compare properties of the metrics. We also implemented the metrics and compared their behaviour on several host parasite datasets, including the shapes of their distributions. In addition, we showed that in combination with multidimensional scaling, the metrics can be useful for visualising large collections of reconciliations, much in the same way as phylogenetic tree metrics can be used to explore collections of phylogenetic trees. Implementations of the metrics can be downloaded from https://team.inria.fr/erable/en/team-members/blerina-sinaimeri/reconciliation-distances/. This work was also funded by a Royal Society Grant obtained by the two partners (at University of New Anglia and ERABLE).

Variants of phylogenetic network problems
Although not falling within the general topic of coevolution, phylogenetic networks are of great interest as another way of representing the evolution of a set of species. In the context of such representations, unrooted and root-uncertain variants of several well-known phylogenetic network problems were explored. The hybridisation number problem requires to embed a set of binary rooted phylogenetic trees into a binary rooted phylogenetic network such that the number of nodes with indegree two is minimised. However, from a biological point of view accurately inferring the root location in a phylogenetic tree is notoriously difficult and poor root placement can artificially inflate the hybridisation number. To this end, we studied in [30] a number of relaxed variants of this problem. We started by showing that the fundamental problem of determining whether an unrooted phylogenetic network displays (*i.e.* embeds) an unrooted phylogenetic tree, is NP-hard. On the positive side, we show that this problem is FPT in the reticulation number. In the rooted case, the corresponding FPT result is trivial, but here we required more subtle argumentation. Next we showed that the hybridisation number problem for unrooted networks (when given two unrooted trees) is equivalent to the problem of computing the tree bisection and reconnect distance of the two unrooted trees. In the third part of the paper, we considered the "root uncertain" variant of hybridisation number. Here we were free to choose the root location in each of a set of unrooted input trees such that the hybridisation number of the resulting rooted trees is minimised. On the negative side, we showed that this problem is APX-hard. On the positive side, we showed that the problem is FPT in the hybridisation number, via kernelisation, for any number of input trees.

6.5. Axis 4: Human, animal and plant health

Hydrogen peroxide production and myo-inositol metabolism as important traits for virulence of Mycoplasma hyopneumoniae

Mycoplasma hyopneumoniae is the causative agent of enzootic pneumonia. In a previous work, we had reconstructed the metabolic models of this species along with two other mycoplasmas from the respiratory tract of swine: *Mycoplasma hyorhinis*, considered less pathogenic but which nonetheless causes disease and *Mycoplasma flocculare*, a commensal bacterium. We had identified metabolic differences that partially explained their different levels of pathogenicity. One important trait was the production of hydrogen peroxide from the glycerol metabolism only in the pathogenic species. Another important feature was a pathway for the metabolism of myo-inositol in *M. hyopneumoniae*. In the paper accepted this year [9], we tested these traits to understand their relation to the different levels of pathogenicity, comparing not only the species but also pathogenic and attenuated strains of *M. hyopneumoniae*. Regarding the myo-inositol metabolism, we showed that only *M. hyopneumoniae* assimilated this carbohydrate and remained viable when myo-inositol was the primary energy source. Strikingly, only the two pathogenic strains of *M. hyopneumoniae* produced hydrogen peroxide in complex medium. We also showed that this production was dependent on the presence of glycerol. Although further functional tests are needed, this work enabled to identify two interesting metabolic traits of *M. hyopneumoniae* that might be directly related to its enhanced virulence. This is part of the work of Mariana G. Ferrarini, currently postdoc at Insa and in ERABLE, and of Scheila G. Mucha whose PhD (defended in Sept. 2018) was co-supervised by Arnaldo Zaha and by a member of ERABLE.

Cancer

A member of ERABLE continues deeply involved with the Centre Léon Bérard in Lyon, and in that context, a number of works are running, all related to cancer genomics. In the first [28], an integrated genomic study was performed of 25 tumour tissues from radical prostatectomy of aggressive (defined by the International Society of Urological Pathology) prostate cancer patients (10 African Caribbean and 15 French Caucasian) using single nucleotide polymorphism arrays, whole-genome sequencing, and RNA sequencing. The results showed that African Caribbean tumours are characterised by a more frequent deletion at 1q41-43 encompassing the DNA repair gene PARP1, and a higher proportion of intra-chromosomal rearrangements including duplications associated with CDK12 truncating mutations. Transcriptome analyses showed an over-expression of genes related to androgen receptor activity in African Caribbean tumours, and of PVT1, a long non-coding RNA located at 8q24 that confirms the strong involvement of this region in prostate tumours from men of African ancestry. In a second study [15], gene-expression profiling data was used to build and validate a predictive model of outcome for patients with follicular lymphoma. A robust 23-gene expression-based predictor of progression-free survival that is applicable to routinely available formalin-fixed, paraffin-embedded tumour biopsies from such patients was thus developed and validated. Applying this score could allow individualised therapy for patients according to their risk category. In a third study, an integrated analysis highlighted APC11 protein expression as a likely new independent predictive marker for colorectal cancer [6].

In a parallel work by another member of ERABLE [27], it was proposed that cancer Is not (only) a senescence problem. Age is indeed one of the strongest predictors of cancer and risk of death from cancer. Cancer is therefore generally viewed as a senescence-related malady. However, cancer also exists at subclinical levels in humans and other animals, but its earlier effects on the body are poorly known by comparison. What was argued in [27] is that cancer is a significant but ignored burden on the body and is likely to be a strong selective force from early during the lifetime of an organism. It was thus proposed that time has come to adopt this novel view of malignant pathologies to improve our understanding of the ways in which oncogenic phenomena influence the ecology and evolution of animals long before their negative impacts become evident and fatal.

Xylella fastidiosa epidemiological model

Xylella fastidiosa is a notorious plant pathogenic bacterium that represents a threat to crops worldwide. Its subspecies, *Xylella fastidiosa* subsp. *fastidiosa* is the causal agent of Pierce's disease of grapevines. Pierce's disease has presented a serious challenge for the grapevine industry in the United States and turned into an epidemic in Southern California due to the invasion of the insect vector *Homalodisca vitripennis*. In an attempt to minimize the effects of *Xylella fastidiosa* subsp. *fastidiosa* in vineyards, various studies have been developing and testing strategies to prevent the occurrence of Pierce's disease, *i.e.*, prophylactic strategies. Research has also been undertaken to investigate therapeutic strategies to cure vines infected by *Xylella fastidiosa* subsp. *fastidiosa*. In [18], we explicitly review all the strategies published to date and specifies their current status. Furthermore, an epidemiological model of *Xylella fastidiosa* subsp. *fastidiosa* is proposed and key parameters for the spread of Pierce's disease deciphered in a sensitivity analysis of all model parameters. Based on these results, it is concluded that future studies should prioritise therapeutic strategies, while investments should only be made in prophylactic strategies that have demonstrated promising results in vineyards. This is part of the PhD of Henri Taneli Pusa in the context of the H2020 ITN MicroWine, together with another PhD student of the ITN, Ifigeneia Kyrkou. Ifigeneia was the first author of the paper [18] but the mathematical model is the work of Taneli.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Grants with Industry

7.1.1. Spock

- Title: characterization of hoSt-gut microbiota interactions and identification of key Players based on a unified reference for standardized quantitative metagenOmics and metaboliC analysis frameworK
- Industrial Partner: MaatPharma (Person responsible: Lilia Boucinha).
- ERABLE participants: Marie-France Sagot (ERABLE coordinator and PhD main supervisor with Susana Vinga from IST, Lisbon, Portugal, as PhD co-supervisor), Marianne Borderes (beneficiary of the PhD scholarship in MaatPharma).
- Type: ANR Technology (2018-2021).
- Web page: http://team.inria.fr/erable/en/projects/#anr-technology-spock.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Muse

- Title: Multi-Omics and Metabolic models iNtegration to study growth Transition in *Escherichia coli*
- Coordinators: Delphine Ropers (EPI Ibis) and Marie-France Sagot
- ERABLE participants: Marie-France Sagot and Arnaud Mary.
- Type: IXXI Project (2018-2019).
- Web page: none for now.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. Aster

- Title: Algorithms and Software for Third gEneration Rna sequencing

- Coordinator: Hélène Touzet, University of Lille and CNRS.
- ERABLE participants: Vincent Lacroix (ERABLE coordinator), Clara Benoît-Pilven, Audric Cologne, Alex di Genova, Leandro I. S. de Lima, Arnaud Mary, Marie-France Sagot, Camille Sessegolo, Blerina Sinaimeri.
- Type: ANR (2016-2020).
- Web page: http://bioinfo.cristal.univ-lille.fr/aster/.

8.2.1.2. ExHyb

- Title: Exploring genomic stability in hybrids
- Coordinator: C. Vieira
- ERABLE participant(s): C. Vieira
- Type: ANR (2014-2018)
- Web page: Not available

8.2.1.3. GraphEn

- Title: Enumération dans les graphes et les hypergraphes : Algorithmes et complexité
- Coordinator: D. Kratsch
- ERABLE participant(s): A. Mary
- Type: ANR (2015-2019)
- Web page: http://graphen.isima.fr/

8.2.1.4. GrR

- Title: Graph Reconfiguration
- Coordinator: N. Bousquet
- ERABLE participant(s): A. Mary
- Type: ANR JCJC (2019-2021)
- Web page: Not available

8.2.1.5. Green

- Title: Deciphering host immune gene regulation and function to target symbiosis disturbance and endosymbiont control in insect pests
- Coordinator: A. Heddi
- ERABLE participant(s): M.-F. Sagot, C. Vieira
- Type: ANR (2018-2021)
- Web page: Not yet available

8.2.1.6. Hmicmac

- Title: Host-microbiota co-adaptations: mechanisms and consequences
- Coordinator: F. Vavre
- ERABLE participant(s): F. Vavre
- Type: ANR PRC (2017-2020)
- Web page: Not available

8.2.1.7. IMetSym

- Title: Immune and Metabolic Control in Intracellular Symbiosis of Insects
- Coordinator: A. Heddi
- ERABLE participant(s): H. Charles, S. Colella
- Type: ANR Blanc (2014-2017)

- Web page: Not available

8.2.1.8. Resist

- Title: Rapid Evolution of Symbiotic Interactions in response to STress: processes and mechanisms
- Coordinator: N. Kremer
- ERABLE participant(s): F. Vavre
- Type: ANR JCJC (2017-2020)
- Web page: Not available

8.2.1.9. Suzukill

- Title: Managing cold tolerance and quality of mass-produced *Drosophila suzukii* flies to facilitate the application of biocontrol through incompatible and sterile insect techniques
- Coordinator: H. Colinet
- ERABLE participant(s): F. Vavre
- Type: ANR PCRI (2015-2018)
- Web page: Not available

8.2.1.10. Swing

- Title: Worldwide invasion of the Spotted WING Drosophila: Genetics, plasticity and evolutionary potential
- Coordinator: P. Gibert
- ERABLE participant(s): C. Vieira
- Type: ANR PCR (2016-2020)
- Web page: Not available

8.2.1.11. U4atac-brain

- Title: Rôle de l'épissage mineur dans le développement cérébral
- Coordinator: Patrick Edery, Centre de Recherche en Neurosciences de Lyon.
- ERABLE participants: Vincent Lacroix (ERABLE coordinator), Clara Benoît-Pilven, Audric Cologne.
- Type: ANR (2018-2021).
- Web page: Not available.

8.2.2. Idex

8.2.2.1. Micro-be-have

- Title: Microbial Impact on insect behaviour: from niche and partner selection to the development of new control methods for pests and disease vectors
- Coordinator: F. Vavre
- ERABLE participant(s): F. Vavre
- Type: AO Scientific Breakthrough (2018-2021)
- Web page: Not available

8.2.3. ADT Inria

8.2.3.1. ADT Inria Kirikomics

- Main objective: Development of a portal to increase the visibility of the tools and resources elaborated by ERABLE around the analysis – using omics data – of metabolic networks modelled by hypergraphs, and enable to visualise the results. (the web page is for now private, it will be made public later in the project).
- Duration: 2016-2017, renewable one more year.
- Person responsible for ADT: Arnaud Mary with David Parsons (Inria).
- Beneficiary of ADT: Martin Wannagat.
- Funds received: Salary for engineer.

8.2.4. Others

Notice that were included here national projects of our members from Italy and the Netherlands when these have no other partners than researchers from the same country.

8.2.4.1. Advanced computational methodologies for the analysis of biomedical data

- Title: Advanced computational methodologies for the analysis of biomedical data
- Coordinator: P. Milazzo
- ERABLE participant(s): R. Grossi, N. Pisanti
- Type: PRA, MIUR PRIN, Italian Ministery of Research National Projects (2017-2018)
- Web page: Not available

8.2.4.2. Advanced Tools and Techniques for the analysis of criminal networks

- Title: Advanced Tools and Techniques for the analysis of criminal networks
- Coordinator: G. Italiano
- ERABLE participant(s): G. Italiano
- Type: LEONARDO SpA (2015-2018)
- Web page: Not available

8.2.4.3. Open Innovation: Digital Innovation for Driving

- Title: Open Innovation: Digital Innovation for Driving
- Coordinator: G. Italiano
- ERABLE participant(s): G. Italiano
- Type: Bridgestone (2018-2019)
- Web page: Not available

8.2.4.4. CMACBioSeq

- Title: Combinatorial Methods for analysis and compression of biological sequences
- Coordinator: G. Rosone
- ERABLE participant(s): N. Pisanti
- Type: SIR, MIUR PRIN, Italian Ministery of Research National Projects (2015-2019)
- Web page: http://pages.di.unipi.it/rosone/CMACBioSeq.html

8.2.4.5. Statistical Models for Structural Genetic Variants in the Genome of the Netherlands

- Title: Statistical Models for Structural Genetic Variants in the Genome of the Netherlands
- Coordinator: A. Schönhuth
- ERABLE participant(s): A. Schönhuth
- Type: Nederlandse Wetenschappelijke Organisatie (NWO) (2013-2018)
- Web page: Not available

8.2.4.6. TALS and splicing

- Title: Development of bioinformatic methods for the analysis of splicing events in patients with the Taybi-Linder Syndrome (TALS)
- Coordinator: P. Edery
- ERABLE participant(s): C. Benoît-Pilven, Audric Cologne, V. Lacroix
- Type: INSERM
- Web page: Not available

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. MicroWine

- Title: Microbial metagenomics and the modern wine industry

- Duration: January 2015 - January 2019
- Coordinator: Lars Hestbjerg Hansen, University of Copenhagen
- ERABLE participant(s): A. Marchetti-Spaccamela, A. Mary, H. T. Pusa, M.-F. Sagot, L. Stougie
- Type: H2020-MSCA-ETN-2014
- Web page: https://team.inria.fr/erable/en/microwine/ and http://www.microwine.eu/

8.3.2. Collaborations in European Programs, Except FP7 & H2020

8.3.2.1. Combinatorics of co-evolution

- Title: The combinatorics of co-evolution
- Duration: 2015 - 2018
- Coordinator: Katharina Huber, University of Warwick, UK
- ERABLE participant(s): M.-F. Sagot, B. Sinaimeri
- Type: The Royal Society
- Web page: not available

8.3.3. Collaborations with Major European Organizations

By itself, ERABLE is built from what initially were collaborations with some major European Organisations (CWI, Sapienza University of Rome, Universities of Florence and Pisa, Free University of Amsterdam) and now has become a European Inria Team.

8.4. International Initiatives

8.4.1. Inria Associate Teams Not Involved in an Inria International Labs

Compasso

- Title: COMmunity Perspective in the health sciences: Algorithms and Statistical approacheS for explOring it
- Duration: 2018, renewable from 2 to 5 years more
- Coordinator: On the Portuguese side, Susana Vinga, IST, Lisbon, Portugal; on the French side, Marie-France Sagot
- ERABLE participant(s): R. Andrade, M. Ferrarini, G. Italiano, A. Marchetti-Spaccamela, A. Mary, H. T. Pusa, M.-F. Sagot, B. Sinaimeri, L. Stougie, A. Viari, I. Ziska
- Web page: http://team.inria.fr/erable/en/projects/inria-associated-team-compasso/

8.4.2. Participation in International Programs

ERABLE is coordinator 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 BAOBAB-ERABLE. The LIA was created in January 2012 for 4 years, renewable once for 4 more years. A web page for the LIA LIRIO is available at this address: http://team.inria.fr/erable/en/cnrs-lia-laboratoire-international-associe-lirio/.

ERABLE also participated to the BASIS project. This was funded by the European Community Seventh Framework Programme (Grant 242006 - 2010-2015). It was led by Dr. Mike Stratton and involved six European countries. It was primarily focused on ER+/HER2- breast cancers, but during the course of the project, was merged with the HER2+ French-ICGC and triple negative UK-ICGC projects, resulting in the analysis of the whole spectrum of breast cancers. The French group was initiated by Dr. Gilles Thomas and was pursued by Alain Viari after the loss of Dr. Thomas in 2014. The project resulted in the sequencing and thorough analysis of 560 breast cancer whole genomes (Nik-Zainai *et al.*, *Nature*, 534:47-54, 2016), including 75 HER2+ performed by the French working group (Ferrari *et al.*, *Nature Communications*, 7, 2016) and funded by the Institut National du Cancer and by Inserm.

Finally, Marie-France Sagot participates in a Portuguese FCT project, Perseids for "Personalizing cancer therapy through integrated modeling and decision" (2016-2019), with Susana Vinga and a number of other Portuguese researchers. The budget of Perseids is managed exclusively by the other Portuguese partner.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

In 2018, ERABLE greeted the following International scientists:

- In France: Alexander Stuart Ralph (University of Melbourne), Katharina Huber and Vincent Moulton (University of Warwick, UK), Ifigeneia Kyrkou (Aarhus University, Danemark), Ana Tereza Vasconcelos from the LNCC, Brazil), Nuno Mira (IST Portugal), May Alzamel and Costas Iliopoulos (King's College, London, UK), Simona Rombo (University of Palermo, Italy).
- In Italy: Loukas Georgiadis (University of Ioannina, Greece), Matthias Mnich (University of Bonn, Germany), Adam Karczmarcz (University of Warsaw, Poland).
- In the Netherlands: Solon Pissis (King's College, UK).

8.5.1.1. Internships

In 2018, ERABLE in France greeted the following Internships:

- Gabriela Paludo, Federal University of Mato Grosso do Sul, Brazil, from Dec 1st 2017 to June 30 2018, funds for 6 months from Capes, Brazil, and for the last month from ERABLE;
- Rafael Nahat, University of São Paulo, Brazil, from June 1st 2018 to Dec 15, 2018.

In the Netherlands, ERABLE greeted the following Internship: Luca Denti, University Bicocca of Milano, Italy, from October 1 to January 2019.

8.5.2. Visits to International Teams

8.5.2.1. Research Stays Abroad

In 2018, two members of ERABLE from France did research stays at Sapienza University of Rome. These were Marie-France Sagot and Blerina Sinaimeri, both funded by Sapienza, M.-F. Sagot as senior scientist for a visit of one month, and B. Sinaimeri as a junior scientist for a visit of three months. The visits took place at the beginning of 2018, Jan-Feb for M.-F. Sagot, and Jan-Apr for B. Sinaimeri. In the context of her visit to Sapienza, Blerina furthermore gave a mini-course (9h) at the University.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

- Giuseppe Italiano is member of the Steering Committee of the *Workshop on Algorithm Engineering and Experimentation (ALENEX)*, of the *International Colloquium on Automata, Languages and Programming (ICALP)*, and the Workshop/Symposium on Experimental Algorithms (SEA).
- Alberto Marchetti-Spaccamela is a member of the Steering committee of *Workshop on Graph Theoretic Concepts in Computer Science (WG)*), and of *Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS)*.
- Arnaud Mary is member of the Steering Committee of *Workshop on Enumeration Problems and Applications (WEPA)*.

- Marie-France Sagot is member of the Steering Committee of *European Conference on Computational Biology (ECCB)*, *International Symposium on Bioinformatics Research and Applications (ISBRA)*, and *Workshop on Enumeration Problems and Applications (WEPA)*.
- Alexander Schönhuth is member of the Steering committee of the *Research in Computational Molecular Biology, satellite conference on massively parallel sequencing (RECOMB-seq)*.

9.1.1.2. Member of the Organizing Committees

- Marie-France Sagot was co-organiser of the Dagstuhl Seminar on "Algorithmic Enumeration: Output-sensitive, Input-Sensitive, Parameterized, Approximative", Dagstuhl Schloss, Oct 14 to 19, 2018.
- Alexander Schönhuth was main organiser of main organizer of the Lorentz workshop on "Single Cell Data Science", Lorentz center, Leiden, June 4-8, 2018.
- Leen Stougie was member of the Organizing Committee of the *3rd Highlights of Algorithms (HALG 2018) conference* in the Free University of Amsterdam, June 4-6, 2018.

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

- Giuseppe Italiano was a member of the Program Committee of *ISAAC*,*STOC*, *ICALP*, and *WALCOM* in 2018.
- Arnaud Mary was a member of the Program Committee of *CIAC* in 2018.
- Nadia Pisanti was a member of the Program Committee of *BIOINFORMATICS*, *FUN*, *ISBRA*, *BIOKDD*, *CIBB*, *Hi BI BI* in 2018.
- Marie-France Sagot was a member of the Program Committee of *SEA*, *CIAC*, *WABI*, *WEPA*, in 2018.
- Alexander Schönhuth was a member of the Program Committee of *RECOMB*, *ISMB*, *GCB* (German bioinformatics conference), and *ISMCO* (International Symposium on Mathematical and Computational Oncology) in 2018.

9.1.2.2. Reviewer

Members of ERABLE have reviewed papers for a number of workshops and conferences including: *CPM*, *ISMB*, *RECOMB*, *WEPA*, *WABI*.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Pierluigi Crescenzi is member of the Editorial Board of *Journal of Computer and Systems Science and Electronic Notes on Theoretical Computer Science*.
- Roberto Grossi is member of the Editorial Board of *Theory of Computing Systems (TOCS)* and pf *RAIRO – Theoretical Informatics and Applications*.
- Giuseppe Italiano is member of the Editorial Board of *Algorithmica* and *Theoretical Computer Science*.
- Alberto Marchetti-Spaccamela is member of the Editorial Board of *Theoretical Computer Science*.
- Arnaud Mary is Editor-in-Chief of a special issue of *Discrete Applied Mathematics* dedicated to *WEPA* 2016.
- Nadia Pisanti is since 2012 member of Editorial Board of *International Journal of Computer Science and Application (IJCSA)* and since 2017 of *Network Modeling Analysis in Health Informatics and Bioinformatics*.
- Marie-France Sagot is member of the Editorial Board of *BMC Bioinformatics*, *Algorithms for Molecular Biology*, and *Lecture Notes in BioInformatics*.

- Leen Stougie is member of the Editorial Board of *AIMS Journal of Industrial and Management Optimization*.
- Cristina Vieira is Executive Editor of *Gene*, and since 2014 member of the Editorial Board of *Mobile DNA*.

9.1.3.2. Reviewer - Reviewing Activities

Members of ERABLE have reviewed papers for a number of journals including: *Theoretical Computer Science*, *Algorithmica*, *Algorithms for Molecular Biology*, *Bioinformatics*, *BMC Bioinformatics*, *Genome Biology*, *Genome Research*, *IEEE/ACM Transactions in Computational Biology and Bioinformatics (TCBB)*, *Molecular Biology and Evolution*, *Nucleic Acid Research*.

9.1.4. Invited Talks

Giuseppe Italiano gave an Invited Talk on "2-Connectivity in Directed Graphs" at the *17th International Symposium on Experimental Algorithms (SEA 2018)* (http://cs.gssi.it/sea2018/, in L'Aquila, Italy, June 27–29, 2018.

Vincent Lacroix gave an Invited Presentation on "Traitement des données -omics" at the "9ème réunion annuelle Institut Thématique Multi-Organismes Technologies pour la Santé", https://its.aviesan.fr/index.php?pagendx=1046.

Alexander Schönhuth gave Invited Talks at the CONTRA workshop, Warszaw (organisers: Jens Lagergren, Ewa Szczurek, Niko Beerenwinkel), in September and at Harvard Medical School, Boston, in October.

Leen Stougie gave an Invited Talk on "A decomposition theory for vertex enumeration of convex polyhedra" at the *Martin Dyer Day and Queen Mary Algorithms Day*, Queen Mary University, London, UK, July 16-17, 2018. He gave Invited Lectures on "Full-length de novo viral quasispecies assembly through variation graph construction" at the *4th Dutch Bioinformatics & Systems Biology Conference*, De Werelt, Lunteren, Netherlands, May 15; on "Polynomial time vertex enumeration of convex polytopes of bounded branch-width" at the Department of Mathematics, University of Bremen, Germany, Sept 11; and on "Algorithmic Problems in Biological Networks" at the *9th Networks Day*, Hortus Botanicus, Leiden, Netherlands, Sept 19, 2018.

Fabrice Vavre gave Invited Talks on "Impact du microbiote sur l'évolution et la diversité des insectes" at the *Colloque de la Société Francophone de Microbiologie*, Paris, Oct 1-3, 2018; and on "Evolution of host-microbiota interactions: what insects tell us about selection levels?" at the Journées de l'IXXI, Lyon, Oct 16, 2018.

9.1.5. Scientific Expertise

Giuseppe F. Italiano is member of the Council of the European Association for Theoretical Computer Science. Leen Stougie is member of the General Board of the Dutch Network on the Mathematics of Operations Research (Landelijk Netwerk Mathematische Besliskunde (LNMB)).

9.1.6. Research Administration

Hubert Charles is director of the Biosciences Department of the Insa-Lyon and co-director of studies of the "Bioinformatique et Modélisation (BIM)" track.

Giuseppe Italiano is member of the Advisory Board of MADALGO - Center for MAssive Data ALGOrithmics, Aarhus, Denmark.

Alberto Marchetti-Spaccamela was Director of the Department of Computer Engineering and Management Antonio Ruberti at Sapienza University from 2013 until end of May 2018.

Nadia Pisanti is since November 1st 2017 member of the Board of the PhD School in Data Science (University of Pisa jointly with Scuola Normale Superiore Pisa, Scuola S. Anna Pisa, IMT Lucca).

Marie-France Sagot is member of the Advisory Board of CWI, Amsterdam, the Netherlands, and chair of the CSS for MBIO at Inra.

Alexander Schönhuth is member of the Scientific Board of BioSB (the Dutch organisation for bioinformatics) since May 2017.

Leen Stougie is since April 2017 Leader of the Life Science Group at CWI.

Alain Viari was until Feb 2018 Deputy Scientific Director at Inria responsible for the ICST for Life and Environmental Sciences. He thus represented Inria at several national instances related to Life Sciences and Health (Allenvi, Aviesan, Ibisa, etc.). He is member of a number of scientific advisory boards (IRT (Institut de Recherche Technologique) BioAster; Centre Léon Bérard). He also coordinates together with J.-F. Deleuze (CNRGH-Evry) the Research & Development part (CRefIX) of the "Plan France Médecine Génomique 2025".

Fabrice Vavre is President of the Section 29 of the CoNRS and participated to the HCERES Evaluation Committee of the IAME Laboratory (UMR1137, Dir. E. Denamur), Paris, Mar 1-2, 2018.

Cristina Vieira is member of the "Conseil National des Universités" (CNU) 67 ("Biologie des Populations et Écologie"), and since 2017 member of the "Conseil de la Faculté des Sciences et Technologies (FST)" of the University Lyon 1.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

9.2.1.1. France

The members of ERABLE teach both at the Department of Biology of the University of Lyon (in particular within the BISM (BioInformatics, Statistics and Modelling) specialty, and at the department of Bioinformatics of the Insa (National Institute of Applied Sciences). Cristina Vieira is responsible for the Master Biodiversity, Ecology and Evolution (https://www.bee-lyon-univ.fr/). She teaches genetics 192 hours per year at the University and at the ENS-Lyon. Hubert Charles is responsible for the Master of Modelling and Bioinformatics (BIM) at the Insa of Lyon (http://biosciences.insa-lyon.fr/). He teaches 192 hours per year in statistics and biology. Vincent Lacroix is responsible for several courses of the Master in Bioinformatics (https://www.bioinfo-lyon.fr/) (L3: Advanced Bioinformatics, M1: Methods for Data Analysis in Genomics, M1: Methods for Data Analysis in Transcriptomics, M1: Bioinformatics Project, M2: Ethics). He teaches 192 hours per year in bioinformatics. Arnaud Mary is responsible for two courses of the Bioinformatics Curriculum at the University (L2: Introduction to Bioinformatics and Biostatistics, M1: Object Oriented Programming) and one at Insa (Discrete Mathematics). He taught 198 hours in 2018. Blerina Sinaimeri taught 36 hours in 2018 on graph algorithms for the M1 students of the Master in Bioinformatics, and on Discrete Mathematics at Insa. Fabrice Vavre taught 20h at the Master level.

The ERABLE team regularly welcomes M1 and M2 interns from the bioinformatics Master.

Vincent Lacroix was an instructor in NGS data analysis training for the CNRS Formation in the last 4 years since 2015, a course coordinated by Eric Rivals from the LIRMM, Montpellier (https://www.france-bioinformatique.fr/fr/formations/bioinformatique-pour-traitement-donn%C3%A9es-s%C3%A9quen%C3%A7age-ngs.

All French members of the ERABLE team are affiliated to the doctoral school E2M2 (Ecology-Evolution-Microbiology-Modelling, http://e2m2.universite-lyon.fr/).

9.2.1.2. Italy & The Netherlands

Italian researchers teach between 90 and 140 hours per year, at both the undergraduate and at the Master levels. The teaching involves pure computer science courses (such as Programming foundations, Programming in C or in Java, Computing Models, Distributed Algorithms) and computational biology (such as Algorithms for Bioinformatics).

Dutch researchers teach between 40 and 270 hours per year, again at the undergraduate and Master levels, in pure computer science (*e.g.* Algorithm Engineering, Randomised Algorithms), applied mathematics (*e.g.* Operational Research, Advanced Linear Programming) and computational biology (*e.g.* Biological Network Analysis).

9.2.2. Supervision

The following PhDs were defended in ERABLE in 2018:

- Damiano di Francesco Maesa, University of Pisa, co-supervisors: Laura Ricci and Andrea Marino, Oct 2018.
- Mattia Gastaldello, Sapienza University of Rome and University of Lyon 1 (funded by "Vinci Program-Université Franco-Italienne", co-supervisors: Tiziana Calamoneri, Sapienza University of Rome; Marie-France Sagot), Feb 2018.
- Scheila Gabriele Mucha, Federal University of Rio Grande do Sul, Brazil (PhD scholarship from Brazil with one year spent as "sandwich PhD" in France funded by Capes, and 6 more months funded by a grant from ERABLE, co-supervisors: Arnaldo Zaha, Federal University of Rio Grande do Sul; Marie-France Sagot), Sept 2018.

The following are the PhDs in progress:

- Marianne Borderes, University Lyon 1 (funded by ANR Technology Spock, co-supervisors: Susana Vinga – Instituto Superior Técnico at Lisbon; Marie-France Sagot)
- Audric Cologne, University of Lyon 1 (funded by Inserm and Inria, co-supervisors: Patrick Edery – Federation of Health Research of Lyon-Est, Vincent Lacroix)
- Leandro Ishi Soares de Lima, University of Lyon 1 (funded by the Brazilian "Science without Borders" program, co-supervisors: Giuseppe Italiano, Vincent Lacroix, Marie-France Sagot)
- Carol Moraga Quinteros, University of Lyon 1 (funded by Conicyt Chile, co-supervisors: Rodrigo Gutierrez – Catholic University of Chile, Marie-France Sagot)
- Henri Taneli Pusa, University of Lyon 1 (funded by H2020-MSCA-ETN-2014 project MicroWine, co-supervisors: Alberto Marchetti-Spaccamela, Arnaud Mary, Marie-France Sagot)
- Camille Sessegolo, University of Lyon 1 (funded by ANR Aster; co-supervisors: Vincent Lacroix, Arnaud Mary)
- Yishu Wang, University Lyon 1 (funded by Ministère de l?Enseignement supérieur, de la Recherche et de l?Innovation, co-supervisors: Mário Figueiredo – Instituto Superior Técnico at Lisbon; Marie-France Sagot; Blerina Sinaimeri)
- Irene Ziska, University Lyon 1 (funded by Inria Cordi-S, co-supervisors: Susana Vinga – Instituto Superior Técnico at Lisbon; Marie-France Sagot)

Besides the PhD students indicated above, who are physically located within one of the premises of ERABLE, the project-team has PhD students in co-supervision who spend the majority or the whole of their time in the premises of other teams. These include: Rita Ramos (funded by Portuguese FCT, co-supervisors: Claudia Nunes dos Santos – ITQB Lisbon, Marie-France Sagot), and André Veríssimo (funded by Portuguese FCT, co-supervisors: Susana Vinga – Instituto Superior Técnico in Lisbon, Marie-France Sagot).

9.2.3. Juries

The following are the PhD or HDR juries to which members of ERABLE participated in 2018.

- Giuseppe Italiano: External Examiner of the PhD of Daniel Wolleb-Graf, ETH, Zurich, Switzerland, Dec 3, 2018.
- Marie-France Sagot: External Reviewer of the PhD of Clémence Frioux, University of Rennes 1, France, Nov 19, 2018.
- Leen Stougie: Member of the Committee of the PhD of Bart Kamphorst, Technical University Eindhoven, Netherlands, May 2018; Member of the Committee of Gregorios Koumoutsos, Technical University Eindhoven, Netherlands, Sept 2018.
- Fabrice Vavre: External Reviewer of the HDR Committee of Thierry Lefevre, Univ of Montpellier, France, March 3, 2018.

9.3. Popularization

9.3.1. Interventions

Blerina Sinaimeri and Ricardo Andrade participated to the Inria Fête de la Science in October 2018. Blerina Sinaimeri and Mariana G. Ferrarini presented the research topics of the team in an event for welcoming international scientists at the Espace Ulys of the University of Lyon. The presentation was entitled “ERABLE : A bio-info symbiosis” and discussed mainly about the multidisciplinarity of the ERABLE Team. Finally, Mariana Galvao Ferrarini participated in the Fête de la Science INRA at the Mediathéque de Bron in October 2018.

Fabrice Vavre participated to the Forum des métiers, Collège Clément Marot, on “La recherche au CNRS”.

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

[1] U. ASHRAF, C. BENOIT-PILVEN, V. LACROIX, V. NAVRATIL, N. NAFFAKH. *Advances in Analyzing Virus-Induced Alterations of Host Cell Splicing*, in "Trends in Microbiology", December 2018, https://hal.inria.fr/hal-01964983

[2] L. BEAUDOU, P. DANKELMANN, F. FOUCAUD, M. A. HENNING, A. MARY, A. PARREAU. *Bounding the Order of a Graph Using Its Diameter and Metric Dimension: A Study Through Tree Decompositions and VC Dimension*, in "Siam Journal on Discrete Mathematics", April 2018, vol. 32, n^{o} 2, pp. 902 - 918 [*DOI :* 10.1137/16M1097833], https://hal.archives-ouvertes.fr/hal-01848648

[3] C. BENOIT-PILVEN, C. MARCHET, E. CHAUTARD, L. LIMA, M.-P. LAMBERT, G. SACOMOTO, A. REY, A. COLOGNE, S. TERRONE, L. DULAURIER, J.-B. CLAUDE, C. BOURGEOIS, D. AUBOEUF, V. LACROIX. *Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data*, in "Scientific Reports", December 2018, vol. 8, n^{o} 1 [*DOI :* 10.1038/s41598-018-21770-7], https://hal.inria.fr/hal-01924204

[4] S. COLELLA, N. PARISOT, P. SIMONET, K. GAGET, G. DUPORT, P. BAA-PUYOULET, Y. RAHBÉ, H. CHARLES, G. FEBVAY, P. CALLAERTS, F. CALEVRO. *Bacteriocyte Reprogramming to Cope With Nutritional Stress in a Phloem Sap Feeding Hemipteran, the Pea Aphid Acyrthosiphon pisum*, in "Frontiers in Physiology", 2018, vol. 9, 1498 p. [*DOI :* 10.3389/FPHYS.2018.01498], https://hal.archives-ouvertes.fr/hal-01905508

[5] D. DI FRANCESCO MAESA, A. MARINO, L. RICCI. *Data-driven analysis of Bitcoin properties: exploiting the users graph*, in "International Journal of Data Science and Analytics", August 2018, vol. 6, n^{o} 1, pp. 63-80, https://hal.inria.fr/hal-01964718

[6] Y. DROUET, I. TREILLEUX, A. VIARI, S. LÉON, M. DEVOUASSOUX-SHISHEBORAN, N. VOIRIN, C. DE LA FOUCHARDIERE, B. MANSHIP, A. PUISIEUX, C. LASSET, C. MOYRET-LALLE. *Integrated analysis highlights APC11 protein expression as a likely new independent predictive marker for colorectal cancer*, in "Scientific Reports", December 2018, vol. 8, n^{o} 1 [*DOI :* 10.1038/s41598-018-25631-1], https://hal.inria.fr/hal-01920572

[7] O. DURON, P. DOUBLET, F. VAVRE, D. BOUCHON. *The Importance of Revisiting Legionellales Diversity*, in "Trends in Parasitology", December 2018, vol. 34, n° 12, pp. 1027 - 1037 [*DOI :* 10.1016/J.PT.2018.09.008], https://hal.archives-ouvertes.fr/hal-01933206

[8] O. DURON, O. MOREL, V. NOËL, M. BUYSSE, F. BINETRUY, R. LANCELOT, E. LOIRE, C. MÉNARD, O. BOUCHEZ, F. VAVRE, L. VIAL. *Tick-Bacteria Mutualism Depends on B Vitamin Synthesis Pathways*, in "Current Biology - CB", June 2018, vol. 28, n° 12, pp. 1896 - 1902.e5 [*DOI :* 10.1016/J.CUB.2018.04.038], https://hal.archives-ouvertes.fr/hal-01933362

[9] M. G. FERRARINI, S. G. MUCHA, D. PARROT, G. MEIFFREN, J. F. R. BACHEGA, G. COMTE, A. ZAHA, M.-F. SAGOT. *Hydrogen peroxide production and myo-inositol metabolism as important traits for virulence of Mycoplasma hyopneumoniae*, in "Molecular Microbiology", 2018, pp. 1-45 [*DOI :* 10.1111/MMI.13957], https://hal.inria.fr/hal-01765751

[10] A. D. GENOVA, G. A. RUZ, M.-F. SAGOT, A. MAASS. *Fast-SG: an alignment-free algorithm for hybrid assembly*, in "GigaScience", May 2018, vol. 7, n° 5, pp. 1 - 15 [*DOI :* 10.1093/GIGASCIENCE/GIY048], https://hal.inria.fr/hal-01842462

[11] R. P. GROSSI, G. MENCONI, N. P. PISANTI, R. TRANI, S. VIND. *Motif trie: An efficient text index for pattern discovery with don't cares*, in "Theoretical Computer Science", February 2018, vol. 710, pp. 74-87, https://hal.inria.fr/hal-01964717

[12] L. GUIO, C. VIEIRA, J. GONZÁLEZ. *Stress affects the epigenetic marks added by natural transposable element insertions in Drosophila melanogaster*, in "Scientific Reports", December 2018, vol. 8, n° 1, https://hal.inria.fr/hal-01965035

[13] K. HUBER, V. MOULTON, M.-F. SAGOT, B. SINAIMERI. *Exploring and Visualising Spaces of Tree Reconciliations*, in "Systematic Biology", December 2018, https://hal.inria.fr/hal-01964682

[14] K. HUBER, V. MOULTON, M.-F. SAGOT, B. SINAIMERI. *Geometric medians in reconciliation spaces*, in "Information Processing Letters", 2018, vol. 136, pp. 96 - 101 [*DOI :* 10.1016/J.IPL.2018.04.001], https://hal.inria.fr/hal-01842439

[15] S. HUET, B. TESSON, J.-P. JAIS, A. FELDMAN, L. MAGNANO, E. THOMAS, A. TRAVERSE-GLEHEN, B. ALBAUD, M. CARRÈRE, L. XERRI, S. ANSELL, L. BASEGGIO, C. REYES, K. TARTE, S. BOYAULT, C. HAIOUN, B. LINK, P. FEUGIER, A. LOPEZ-GUILLERMO, H. TILLY, P. BRICE, S. HAYETTE, F. JARDIN, F. OFFNER, P. SUJOBERT, D. GENTIEN, A. VIARI, E. CAMPO, J. CERHAN, G. SALLES. *A gene-expression profiling score for prediction of outcome in patients with follicular lymphoma: a retrospective training and validation analysis in three international cohorts*, in "Lancet Oncology", April 2018, vol. 19, n° 4, pp. 549 - 561, Erratum in: Correction to Lancet Oncol 2018; 19: 549-61. [Lancet Oncol. 2018] [*DOI :* 10.1016/S1470-2045(18)30102-5], https://hal.archives-ouvertes.fr/hal-01788996

[16] B. IBRAHIM, K. ARKHIPOVA, A. C. ANDEWEG, S. POSADA-CÉSPEDES, F. ENAULT, A. GRUBER, E. KOONIN, A. KUPCZOK, P. LEMEY, A. C. MCHARDY, D. P. MCMAHON, B. E. PICKETT, D. ROBERTSON, R. H. SCHEUERMANN, A. ZHERNAKOVA, M. P. ZWART, A. SCHÖNHUTH, B. E. DUTILH, M. MARZ. *Bioinformatics Meets Virology: The European Virus Bioinformatics Center's Second Annual Meeting*, in "Viruses", May 2018, vol. 10, n° 5, 256 p. [*DOI :* 10.3390/V10050256], https://hal.inria.fr/hal-01964743

[17] M. JAILLARD, L. LIMA, M. TOURNOUD, P. MAHÉ, A. VAN BELKUM, V. LACROIX, L. JACOB. *A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between k-mers and genetic events*, in "PLoS Genetics", 2018, pp. 1-28 [*DOI :* 10.1371/JOURNAL.PGEN.1007758], https://hal.inria.fr/hal-01920359

[18] I. KYRKOU, T. PUSA, L. ELLEGAARD-JENSEN, M.-F. SAGOT, L. H. HANSEN. *Pierce's Disease of Grapevines: A Review of Control Strategies and an Outline of an Epidemiological Model*, in "Frontiers in Microbiology", September 2018, vol. 9, pp. 1-23 [*DOI :* 10.3389/FMICB.2018.02141], https://hal.inria.fr/hal-01920527

[19] E. LERAT, C. GOUBERT, S. GUIRAO-RICO, M. MERENCIANO, A.-B. DUFOUR, C. VIEIRA, J. GONZÁLEZ. *Population specific dynamics and selection patterns of transposable element insertions in European natural populations*, in "Molecular Ecology", 2018, https://hal.inria.fr/hal-01965027

[20] G. A. MARAIS, J.-M. GAILLARD, C. VIEIRA, I. PLOTTON, D. SANLAVILLE, F. GUEYFFIER, J.-F. LEMAÎTRE. *Sex gap in aging and longevity: can sex chromosomes play a role?*, in "Biology of Sex Differences", December 2018, vol. 9, n^{o} 1, https://hal.inria.fr/hal-01965036

[21] T. MARSCHALL, M. MARZ, T. ABEEL, L. DIJKSTRA, B. E. DUTILH, A. GHAFFAARI, P. KERSEY, W. P. KLOOSTERMAN, V. MAKINEN, A. M. NOVAK, B. PATEN, D. PORUBSKY, E. RIVALS, C. ALKAN, J. A. BAAIJENS, P. I. W. DE BAKKER, V. BOEVA, R. J. P. BONNAL, F. CHIAROMONTE, R. CHIKHI, F. D. CICCARELLI, R. CIJVAT, E. DATEMA, C. M. V. DUIJN, E. E. EICHLER, C. ERNST, E. ESKIN, E. GARRISON, M. EL-KEBIR, G. W. KLAU, J. O. KORBEL, E.-W. LAMEIJER, B. LANGMEAD, M. MARTIN, P. MEDVEDEV, J. C. MU, P. NEERINCX, K. OUWENS, P. PETERLONGO, N. PISANTI, S. RAHMANN, B. RAPHAEL, K. REINERT, D. DE RIDDER, J. DE RIDDER, M. SCHLESNER, O. SCHULZ-TRIEGLAFF, A. D. SANDERS, S. SHEIKHIZADEH, C. SHNEIDER, S. SMIT, D. VALENZUELA, J. WANG, L. WESSELS, Y. ZHANG, V. GURYEV, F. VANDIN, K. YE, A. SCHÖNHUTH. *Computational pan-genomics: status, promises and challenges*, in "Briefings in Bioinformatics", 2018, vol. 19, n^{o} 1, pp. 118-135 [*DOI :* 10.1093/BIB/BBW089], https://hal.inria.fr/hal-01390478

[22] B. MENEZES, J. SALCES-ORTIZ, H. MULLER, N. BURLET, S. MARTINEZ, M. FABLET, C. VIEIRA. *An attempt to select non-genetic variation in resistance to starvation and reduced chill coma recovery time in Drosophila melanogaster*, in "The Journal of Experimental Biology", December 2018, vol. 221, n^{o} 23, jeb186254, https://hal.inria.fr/hal-01965032

[23] A. MONTI, B. SINAIMERI. *On variants of Vertex Geography on undirected graphs*, in "Discrete Applied Mathematics", December 2018, vol. 251, pp. 268-275, https://hal.inria.fr/hal-01965197

[24] D. P. MORTON, W. W. ROMEIJNDERS, R. R. SCHULTZ, L. STOUGIE. *The stochastic programming heritage of Maarten van der Vlerk*, in "Computational Management Science", October 2018, vol. 15, n^{o} 3-4, pp. 319 - 323 [*DOI :* 10.1007/S10287-018-0335-8], https://hal.inria.fr/hal-01925979

[25] R. ROMBA, O. GNANKINE, S. F. DRABO, F. TIENDRÉBÉOGO, H. HENRI, L. MOUTON, F. VAVRE. *Abundance of Bemisia tabaci Gennadius (Hemiptera: Aleyrodidae) and its parasitoids on vegetables and cassava plants in Burkina Faso (West Africa)*, in "Ecology and Evolution", 2018, vol. 8, n^{o} 12, pp. 6091-6103 [*DOI :* 10.1002/ECE3.4078], https://hal.archives-ouvertes.fr/hal-01916799

[26] P. SIMONET, K. GAGET, S. BALMAND, M. RIBEIRO LOPES, N. PARISOT, K. BUHLER, G. DUPORT, V. VULSTEKE, G. FEBVAY, A. HEDDI, H. CHARLES, P. CALLAERTS, F. CALEVRO. *Bacteriocyte cell death in*

the pea aphid/ Buchnera symbiotic system, in "Proceedings of the National Academy of Sciences of the United States of America ", February 2018, vol. 115, n^{o} 8, pp. E1819 - E1828 [*DOI :* 10.1073/PNAS.1720237115], https://hal.archives-ouvertes.fr/hal-01768487

[27] F. THOMAS, F. VAVRE, T. TISSOT, M. VITTECOQ, M. GIRAUDEAU, F. F. BERNEX, D. MISSÉ, F. RENAUD, N. RAVEN, C. BECKMANN, R. HAMEDE, P. BIRO, B. UJVARI. *Cancer Is Not (Only) a Senescence Problem*, in "Trends in Cancer", March 2018, vol. 4, n^{o} 3, pp. 169 - 172 [*DOI :* 10.1016/J.TRECAN.2018.01.002], https://hal.archives-ouvertes.fr/hal-01933408

[28] L. TONON, G. FROMONT, S. BOYAULT, E. THOMAS, A. FERRARI, A.-S. SERTIER, J. KIELBASSA, V. LE TEXIER, A. KAMOUN, N. ELAROUCI, J. IRANI, L. MULTIGNER, I. I. GUT, M. GUT, P. BLANCHET, A. DE REYNIÈS, G. CANCEL-TASSIN, A. VIARI, O. CUSSENOT. *Mutational Profile of Aggressive, Localised Prostate Cancer from African Caribbean Men Versus European Ancestry Men*, in "European Urology", September 2018, vol. 75, n^{o} 1, pp. 11-15 [*DOI :* 10.1016/J.EURURO.2018.08.026], https://hal.inria.fr/hal-01921597

[29] L. URBINI, B. SINAIMERI, C. MATIAS, M.-F. SAGOT. *Exploring the Robustness of the Parsimonious Reconciliation Method in Host-Symbiont Cophylogeny*, in "IEEE/ACM Transactions on Computational Biology and Bioinformatics", May 2018, pp. 1-11 [*DOI :* 10.1109/TCBB.2018.2838667], https://hal.inria.fr/hal-01842451

[30] L. VAN IERSEL, S. KELK, G. STAMOULIS, L. STOUGIE, O. BOES. *On Unrooted and Root-Uncertain Variants of Several Well-Known Phylogenetic Network Problems*, in "Algorithmica", 2018, vol. 64, n^{o} 4, pp. 621 - 637 [*DOI :* 10.1093/SYSBIO/SYV020], https://hal.inria.fr/hal-01599716

[31] G. L. WALLAU, C. VIEIRA, É. L. S. LORETO. *Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome*, in "Mobile DNA", December 2018, vol. 9, n^{o} 1, https://hal.inria.fr/hal-01965038

International Conferences with Proceedings

[32] M. ALZAMEL, L. A. K. AYAD, G. BERNARDINI, R. P. GROSSI, C. S. ILIOPOULOS, N. P. PISANTI, S. P. PISSIS, G. ROSONE. *Degenerate String Comparison and Applications*, in "Workshop on Algorithms in Bioinformatics (WABI)", Helsinki, Finland, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018, vol. 113 [*DOI :* 10.4230/LIPICS.WABI.2018.21], https://hal.inria.fr/hal-01964706

[33] L. A. K. AYAD, G. BERNARDINI, R. P. GROSSI, C. S. ILIOPOULOS, N. PISANTI, S. P. PISSIS, G. ROSONE. *Longest Property-Preserved Common Factor*, in "International Symposium on String Processing and Information Retrieval", Lima, Peru, 2018, https://hal.inria.fr/hal-01921603

[34] J. A. BAAIJENS, B. VAN DER ROEST, J. KOESTER, L. STOUGIE, A. SCHOENHUTH. *Full-length de novo viral quasispecies assembly through variation graph construction*, in "RECOMB Satellite on Massively Parallel Sequencing (RECOMB-Seq)", Paris, France, 2018 [*DOI :* 10.1101/287177], https://hal.inria.fr/hal-01964730

[35] S. BHATTACHARYA, M. HENZINGER, G. F. ITALIANO. *Design of Dynamic Algorithms via Primal-Dual Method*, in "International Colloquium on Automata, Languages, and Programming (ICALP)", Kyoto, Japan, Lecture Notes in Computer Science, 2018, vol. 9134, pp. 206-218, https://arxiv.org/abs/1604.05337 , https://hal.inria.fr/hal-01964700

[36] P. BILLE, M. BERGGREEN ETTIENNE, R. P. GROSSI, I. L. GØRTZ, E. ROTENBERG. *Compressed Communication Complexity of Longest Common Prefixes*, in "International Symposium on String Processing and Information Retrieval (SPIRE)", Lima, Peru, Lecture Notes in Computer Science, 2018, vol. 11147, pp. 74-87, https://hal.inria.fr/hal-01964719

[37] T. BOSMAN, M. VAN EE, Y. JIAO, A. MARCHETTI-SPACCAMELA, R. RAVI, L. STOUGIE. *Approximation Algorithms for Replenishment Problems with Fixed Turnover Times*, in "Latin American Symposium on Theoretical Informatics (LATIN)", Buenos Aires, Argentina, Lecture Notes in Computer Science, 2018, vol. 10807, pp. 217-230, https://arxiv.org/abs/1712.05218 , https://hal.inria.fr/hal-01964685

[38] G. BOTTAZZI, G. F. ITALIANO, D. SPERA. *Preventing Ransomware Attacks Through File System Filter Drivers*, in "Second Italian Conference on Cyber Security", Milan, Italy, 2018, https://hal.inria.fr/hal-01925958

[39] N. BOUSQUET, A. MARY. *Reconfiguration of graphs with connectivity constraints*, in "International Workshop on Approximation and Online Algorithms (WAOA)", Helsinki, Finland, Lecture Notes in Computer Science, 2018, vol. 11312, pp. 295-309, https://arxiv.org/abs/1809.05443 , https://hal.inria.fr/hal-01964723

[40] A. CONTE, T. DE MATTEIS, D. DE SENSI, R. P. GROSSI, A. MARINO, L. P. VERSARI. *D2K: Scalable Community Detection in Massive Networks via Small-Diameter k-Plexes*, in "ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)", London, United Kingdom, 2018, pp. 1272-1281, https://hal.inria.fr/hal-01964688

[41] A. CONTE, G. FERRARO, R. P. GROSSI, A. MARINO, K. SADAKANE, T. UNO. *Node Similarity with q -Grams for Real-World Labeled Networks*, in "ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)", London, United Kingdom, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2018, pp. 1282-1291 [*DOI :* 10.1145/3219819.3220085], https://hal.inria.fr/hal-01964716

[42] A. CONTE, R. P. GROSSI, A. MARINO, R. RIZZI, T. UNO, L. P. VERSARI. *Tight Lower Bounds for the Number of Inclusion-Minimal st-Cuts*, in "International Workshop on Graph-Theoretic Concepts in Computer Science (WG)", Cottbus, Germany, Lecture Notes in Computer Science, 2018, vol. 11159, pp. 100-110, https://hal.inria.fr/hal-01964710

[43] A. CONTE, R. P. GROSSI, A. MARINO, R. RIZZI, L. P. VERSARI. *Listing Subgraphs by Cartesian Decomposition*, in "International Symposium on Mathematical Foundations of Computer Science (MFCS)", Dagstuhl, Germany, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018, vol. 117, pp. 1868-8969 [*DOI :* 10.4230/LIPICS.MFCS.2018.84], https://hal.inria.fr/hal-01964703

[44] A. CONTE, R. P. GROSSI, A. MARINO, L. P. VERSARI. *Finding Maximal Common Subgraphs via Time-Space Efficient Reverse Search*, in "International Computing and Combinatorics Conference (COCOON)", Qing Dao, China, Lecture Notes in Computer Science, 2018, vol. 10976, pp. 328-340, https://hal.inria.fr/hal-01964709

[45] L. GEORGIADIS, G. F. ITALIANO, N. PAROTSIDIS. *Incremental Strong Connectivity and 2-Connectivity in Directed Graphs*, in "Latin American Symposium on Theoretical Informatics", Buenos Aires, Argentina, 2018, https://arxiv.org/abs/1802.10189 , https://hal.inria.fr/hal-01925953

[46] R. P. GROSSI, A. MARINO, L. P. VERSARI. *Efficient Algorithms for Listing k Disjoint st-Paths in Graphs*, in "Latin American Symposium on Theoretical Informatics (LATIN)", Buenos Aires, Argentina, Lecture Notes in Computer Science, 2018, vol. 10807, pp. 544-557, https://hal.inria.fr/hal-01964712

[47] R. P. GROSSI, L. P. VERSARI. *Round-Hashing for Data Storage: Distributed Servers and External-Memory Tables*, in "European Symposium on Algorithms (ESA)", Helsinki, Finland, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018, vol. 43, pp. 1 - 43 [*DOI :* 10.4230/LIPICS.ESA.2018.43], https://hal.inria.fr/hal-01964721

[48] J. HOLM, G. F. ITALIANO, A. KARCZMARZ, J. ŁĄCKI, E. ROTENBERG. *Decremental SPQR-trees for Planar Graphs*, in "European Symposium on Algorithms", Helsinki, Finland, 2018, https://arxiv.org/abs/1806.10772 , https://hal.inria.fr/hal-01925961

[49] G. F. ITALIANO, Y. MANOUSSAKIS, K. T. NGUYEN, H.-P. PHAM. *Maximum colorful cliques in vertex-colored graphs*, in "COCOON 2018 - 24th International Computing and Combinatorics Conference", Qingdao, China, LNCS, Springer, July 2018, vol. 10976, pp. 480-491 [*DOI :* 10.1007/978-3-319-94776-1_40], https://hal.archives-ouvertes.fr/hal-01827893

[50] N. OLVER, K. PRUHS, K. SCHEWIOR, R. SITTERS, L. STOUGIE. *The Itinerant List Update Problem*, in "Workshop on Approximation and Online Algorithms (WAOA)", Helsinki, Finland, 2018, pp. 310-326, https://hal.inria.fr/hal-01972562

[51] N. PREZZA, N. P. PISANTI, M. SCIORTINO, G. ROSONE. *Detecting Mutations by eBWT*, in "Workshop on Algorithms in Bioinformatics", Helsinki, Finland, 2018, https://arxiv.org/abs/1805.01876 [*DOI :* 10.4230/LIPICS.CVIT.2016.23], https://hal.inria.fr/hal-01925950

[52] B. YIN, M. BALVERT, D. ZAMBRANO, A. SCHÖNHUTH, S. BOHTE. *An image representation based convolutional network for DNA classification*, in "International Conference on Learning Representations (ICLR)", Vancouver, Canada, 2018, https://arxiv.org/abs/1806.04931 - Published at ICLR 2018, https://openreview.net/pdf?id=HJvvRoe0W, https://hal.inria.fr/hal-01964736

Conferences without Proceedings

[53] G. F. ITALIANO, Y. MANOUSSAKIS, N. KIM THANG, H.-P. PHAM. *Maximum Colorful Cycles in Vertex-Colored Graphs*, in "CSR 2018 - 13th International Computer Science Symposium in Russia", Moscow, Russia, Lecture Notes in Computer Science, Springer, June 2018, vol. 10846, pp. 106-117 [*DOI :* 10.1007/978-3-319-90530-3_10], https://hal.archives-ouvertes.fr/hal-01762581

Other Publications

[54] R. CHIKHI, A. SCHÖNHUTH. *Dualities in Tree Representations*, November 2018, https://arxiv.org/abs/1804.04263 - CPM 2018, extended version, https://hal.archives-ouvertes.fr/hal-01935566

[55] C. MARCHET, L. LIMA. *Simulation of RNA sequencIng with Oxford Nanopore Technologies*, July 2018, pp. 1-2, JOBIM 2018 - Journées Ouvertes Biologie, Informatique et Mathématiques, Poster, https://hal.archives-ouvertes.fr/hal-01929917