

Inria

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 2019

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

1. Team, Visitors, External Collaborators	1
2. Overall Objectives	2
3. Research Program	3
3.1. Two main goals	3
3.2. Different research axes	4
4. Application Domains	6
5. New Software and Platforms	6
5.1. C3Part/Isofun	6
5.2. Cassis	7
5.3. Coala	7
5.4. CSC	7
5.5. Cycads	7
5.6. DBGWAS	8
5.7. Eucalypt	8
5.8. Fast-SG	8
5.9. Gobbolino-Touché	8
5.10. HapCol	8
5.11. HgLib	9
5.12. KissDE	9
5.13. KisSplice	9
5.14. KisSplice2RefGenome	10
5.15. KisSplice2RefTranscriptome	10
5.16. MetExplore	10
5.17. Mirinho	10
5.18. Momo	11
5.19. Moomin	11
5.20. MultiPus	11
5.21. Pitufolandia	11
5.22. Sasita	12
5.23. Savage	12
5.24. Smile	12
5.25. Rime	12
5.26. Totoro & Kotoura	12
5.27. VG-Flow	13
5.28. Virus-VG	13
5.29. Wengan	13
5.30. WhatsHap	13
6. New Results	14
6.1. General comments	14
6.2. Axis 1: Genomics	14
6.3. Axis 2: Metabolism and post-transcriptional regulation	16
6.4. Axis 3: (Co)Evolution	17
6.5. Axis 4: Human and animal health	18
7. Bilateral Contracts and Grants with Industry	19
8. Partnerships and Cooperations	19
8.1. Regional Initiatives	19
8.2. National Initiatives	19
8.2.1. ANR	19
8.2.1.1. Aster	19

8.2.1.2.	GraphEn	19
8.2.1.3.	GrR	20
8.2.1.4.	Green	20
8.2.1.5.	Hmicmac	20
8.2.1.6.	Networks	20
8.2.1.7.	Resist	20
8.2.1.8.	Swing	20
8.2.1.9.	U4atac-brain	21
8.2.2.	Idex	21
8.2.3.	Others	21
8.2.3.1.	AHeAD	21
8.2.3.2.	CMACBioSeq	21
8.2.3.3.	MyOwnResearch	21
8.2.3.4.	Open Innovation: Digital Innovation for Driving	21
8.3.	European Initiatives	22
8.3.1.	Collaborations in European Programs, Except FP7 & H2020	22
8.3.2.	Collaborations with Major European Organizations	22
8.4.	International Initiatives	22
8.4.1.	Inria Associate Teams Not Involved in an Inria International Lab	22
8.4.2.	Participation in Other International Programs	22
8.5.	International Research Visitors	23
8.5.1.	Visits of International Scientists	23
8.5.2.	Visits to International Teams	23
8.5.2.1.	Sabbatical programme	23
8.5.2.2.	Research Stays Abroad	23
9.	Dissemination	23
9.1.	Promoting Scientific Activities	23
9.1.1.	Scientific Events: Organisation	23
9.1.1.1.	General Chair, Scientific Chair	23
9.1.1.2.	Member of the Organizing Committees	24
9.1.2.	Scientific Events: Selection	24
9.1.2.1.	Member of the Conference Program Committees	24
9.1.2.2.	Reviewer	24
9.1.3.	Journal	24
9.1.3.1.	Member of the Editorial Boards	24
9.1.3.2.	Reviewer - Reviewing Activities	24
9.1.4.	Invited Talks	25
9.1.5.	Scientific Expertise	25
9.1.6.	Research Administration	25
9.2.	Teaching - Supervision - Juries	26
9.2.1.	Teaching	26
9.2.1.1.	France	26
9.2.1.2.	Italy & The Netherlands	26
9.2.2.	Supervision	26
9.2.3.	Juries	27
9.3.	Popularization	27
10.	Bibliography	27

Project-Team ERABLE

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

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. - Evolutionary 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

Solon Pissis [CWI, The Netherlands, Senior Researcher, since Mar 2019]

Marie-France Sagot [Inria, Team leader, Senior Researcher, HDR]

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

Blerina Sinimeri [Inria, Researcher]

Fabrice Vavre [CNRS, Researcher, HDR]

Alain Viari [Inria, Senior Researcher]

Faculty Members

Hubert Charles [INSA Lyon, Full Professor, HDR]
Roberto Grossi [Univ Pisa, Italy, Full Professor]
Giuseppe Francesco Italiano [LUISS Univ, Rome, Italy, Professor]
Vincent Lacroix [Univ de Claude Bernard, Associate Professor]
Alberto Marchetti Spaccamela [Sapienza Univ Rome, Italy, Full Professor]
Arnaud Mary [Univ de Claude Bernard, Associate Professor]
Nadia Pisanti [Univ Pisa, Italy, Associate 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. of São Paulo, Brazil, since Nov 2019]
Audric Cologne [CNRS, since Nov 2019]
Alex Di Genova [Inria, until Nov 2019]
Mariana Galvão Ferrarini [Insa Lyon & Univ Claude Bernard Lyon]
Scheila Mucha [Inria, from Nov 2019]
Henri Pusa [Inria, until May 2019]

PhD Students

Marianne Borderes [MaatPharma and Claude Bernard Lyon]
Audric Cologne [Inria, until Sep 2019]
Nicolas Homberg [Inra and Inria, from Nov 2019]
Leandro Ishi Soares de Lima [CNPq Brazil & Univ Claude Bernard Lyon, until Apr 2019]
Carol Moraga Quinteros [Conicyt & Univ Claude Bernard Lyon]
Henri Pusa [Inria, until Feb 2019]
Camille Sessegolo [Univ Claude Bernard Lyon]
Yishu Wang [Univ Claude Bernard Lyon]
Irene Ziska [Inria]

Technical staff

Eric Cumunel [Univ Claude Bernard Lyon, from Nov 2019]

Interns and Apprentices

Eric Cumunel [Univ Claude Bernard Lyon, from Jan to Jun 2019]
Thibaut Dayde [Inria, from Apr 2019 until Jun 2019]
Nicolas Homberg [Univ de Claude Bernard, from Oct 2018 to Mar 2019]

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 archaea – 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 in 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 (co)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 main health areas currently addressed that are intrinsically inter-related.

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 on 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/Isosfun

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 Deniérou
- 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

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 Sinimeri, Catherine Matias, Christian Baudet, Christian Gautier, Marie-France Sagot and Pierluigi Crescenzi
- Contact: Blerina Sinimeri
- 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 (<https://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 Sinimeri, Christian Baudet, Marie-France Sagot and Pierluigi Crescenzi
- Contact: Blerina Sinimeri
- 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.

- Participants: Alex Di Genova, Marie-France Sagot, Alejandro Maass and Gonzalo Ruz Heredia
- 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 P. 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://kisssplice.prabi.fr/tools/kissDE/>

5.13. KisSplice

KEYWORDS: Bioinformatics - Bioinformatics 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 : The 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+.

KISSPLICE was applied to a new application field, virology, through a collaboration with the group of Nadia Naffakh at Institut Pasteur. The goal is to understand how a virus (in this case influenza) manipulates the splicing of its host. This led to new developments in KISSPLICE. Taking into account the strandedness of the reads was required, in order not to mis-interpret transcriptional readthrough. We now use BCALM instead of DBG-V4 for the de Bruijn graph construction and this led to major improvements in memory and time requirements of the pipeline. We still cannot scale to very large datasets like in cancer, the time limiting step being the quantification of bubbles.

- 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

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.

- Participants: Ricardo Luiz de Andrade Abrantes, Nuno Mira, Susana Vinga and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: <http://momo-sysbio.gforge.inria.fr>

5.19. Moomin

Mathematical explORation of Omics data on a Metabolic Network

KEYWORDS: Metabolic networks - Transcriptomics

FUNCTIONAL DESCRIPTION: MOOMIN is a tool for analysing differential expression data. It takes as its input a metabolic network and the results of a DE analysis: a posterior probability of differential expression and a (logarithm of a) fold change for a list of genes. It then forms a hypothesis of a metabolic shift, determining for each reaction its status as "increased flux", "decreased flux", or "no change". These are expressed as colours: red for an increase, blue for a decrease, and grey for no change. See the paper for full details: <https://doi.org/10.1093/bioinformatics/btz584>

- Participants: Henri Taneli Pusa, Mariana Ferrarini, Ricardo Luiz de Andrade Abrantes, Arnaud Mary, Alberto Marchetti-Spaccamela, Leen Stougie and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: <https://github.com/htpusa/moomin>

5.20. MultiPus

KEYWORDS: Systems Biology - Algorithm - Graph algorithmics - Metabolic networks - Computational biology

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/>

5.21. Pitufolandia

KEYWORDS: Bioinformatics - Graph algorithmics - Systems Biology

FUNCTIONAL DESCRIPTION: The algorithms in PITUFOLANDIA (PITUFO / PITUFINA / PAPA PITUFO) 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.

- Participants: Vicente Acuna Aguayo, Paulo Vieira Milreu, Alberto Marchetti-Spaccamela, Leen Stougie, Martin Wannagat and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: <http://gforge.inria.fr/projects/pitufo/>

5.22. Sasita

KEYWORDS: Bioinformatics - Graph algorithmics - Systems Biology

FUNCTIONAL DESCRIPTION: SASITA is a software for the exhaustive enumeration of minimal precursor sets in metabolic networks.

- Participants: Vicente Acuna Aguayo, Ricardo Luiz de Andrade Abrantes, Paulo Vieira Milreu, Alberto Marchetti-Spaccamela, Leen Stougie, Martin Wannagat and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: <http://sasita.gforge.inria.fr/>

5.23. 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>

5.24. 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

5.25. Rime

KEYWORDS: Bioinformatics - Genomics - Sequence alignment

FUNCTIONAL DESCRIPTION: Detects long similar fragments occurring at least twice in a set of biological sequences.

- Participants: Nadia Pisanti and Marie-France Sagot
- Contact: Nadia Pisanti

5.26. 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.

- Participants: Alice Julien-Laferrière, Ricardo Luiz de Andrade Abrantes, Arnaud Mary, Mariana Ferrarini, Susana Vinga, Irene Ziska and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: <http://hyperstories.gforge.inria.fr/>

5.27. VG-Flow

Viral haplotype reconstruction from contigs using variation graphs

KEYWORD: Haplotyping

FUNCTIONAL DESCRIPTION: The goal of haplotype-aware genome assembly is to reconstruct all individual haplotypes from a mixed sample and to provide corresponding abundance estimates. VG-FLOW provides a reference-genome-independent solution based on the construction of a variation graph, capturing all quasispecies diversity present in the sample. We solve the contig abundance estimation problem and propose a greedy algorithm to efficiently build full-length haplotypes. Finally, we obtain accurate frequency estimates for the reconstructed haplotypes through linear programming techniques.

- Contact: Alexander Schonhuth
- URL: <https://bitbucket.org/jbaaijens/vg-flow>

5.28. Virus-VG

Viral haplotype reconstruction from contigs using variation graphs

KEYWORD: Haplotyping

FUNCTIONAL DESCRIPTION: 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. VIRUS-VG aims to reconstruct full-length haplotypes together with their abundances from such contigs, represented as a genome variation graph.

- Contact: Alexander Schonhuth
- URL: <https://bitbucket.org/jbaaijens/virus-vg>

5.29. Wengan

Making the path

KEYWORD: Genome assembly

FUNCTIONAL DESCRIPTION: WENGAN is a new genome assembler that unlike most of the current long-reads assemblers avoids entirely the all-vs-all read comparison. The key idea behind WENGAN is that long-read alignments can be inferred by building paths on a sequence graph. To achieve this, WENGAN builds a new sequence graph called the Synthetic Scaffolding Graph. The SSG is built from a spectrum of synthetic mate-pair libraries extracted from raw long-reads. Longer alignments are then built by performing a transitive reduction of the edges. Another distinct feature of WENGAN is that it performs self-validation by following the read information. WENGAN identifies miss-assemblies at different steps of the assembly process.

- Participants: Alex Di Genova and Marie-France Sagot
- Contact: Marie-France Sagot
- URL: <https://github.com/adigenova/wengan>

5.30. 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/whatschap/whatschap>

6. New Results

6.1. General comments

We present in this section the main results obtained in 2019.

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 chose not to detail here the results on more theoretical aspects of computer science when these are initially addressed in contexts not directly related to computational biology even though those on string [11], [36], [40], [41], [23], [45] and graph algorithms in general [35], [39], [38], [17], [43] are relevant for life sciences, such as for instance pan-genome analysis, or could become more specifically so in a near future. One important example of the latter concerns enumeration algorithms that has always been at the heart of the computer science and mathematics interests of the team. In such context, the so-called reconfiguration problem which asks whether one solution can be transformed into the other in a step-by-step fashion such that each intermediate solution is also feasible is of particular relevance. This was explored in the context of a perfect matching problem [37].

A few other results of 2019 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised [8], [9], [12], [44]. In the same way, also for space reasons, we chose not to detail the results presented in some biological papers of the team when these did not require a mathematical or algorithmic input [16], [22].

On the other hand, we do mention a couple of works that were in preparation or about to be submitted towards the end of 2018.

6.2. Axis 1: Genomics

Transcriptome profiling using Nanopore sequencing Our vision of DNA transcription and splicing has changed dramatically with the introduction of short-read sequencing. These high-throughput sequencing technologies promised to unravel the complexity of any transcriptome. Generally gene expression levels are well-captured using these technologies, but there are still remaining caveats due to the limited read length and the fact that RNA molecules had to be reverse transcribed before sequencing. Oxford Nanopore Technologies has recently launched a portable sequencer which offers the possibility of sequencing long reads and most importantly RNA molecules. In [28], we generated a full mouse transcriptome from brain and liver using such Oxford Nanopore device. As a comparison, we sequenced RNA (RNA-Seq) and cDNA (cDNA-Seq) molecules using both long and short reads technologies and tested the TeloPrime preparation kit, dedicated to the enrichment of full-length transcripts. Using spike-in data, we confirmed in [28] that expression levels are efficiently captured by cDNA-Seq using short reads. More importantly, Oxford Nanopore RNA-Seq tends to be more efficient, while cDNA-Seq appears to be more biased. We further showed that the cDNA library preparation of the Nanopore protocol induces read truncation for transcripts containing internal runs of T's. This bias is marked for runs of at least 15 T's, but is already detectable for runs of at least 9 T's and therefore concerns more than 20% of the expressed transcripts in mouse brain and liver. Finally, we outlined that bioinformatic challenges remain ahead for quantifying at the transcript level, especially when reads are not full-length. Accurate quantification of repeat-associated genes such as processed pseudogenes also remains difficult, and we show in the paper that current mapping protocols which map reads to the genome largely over-estimate their expression, at the expense of their parent gene.

Genotyping and variant detection The amount of genetic variation discovered and characterised in human populations is huge, and is growing rapidly with the widespread availability of modern sequencing technologies. Such a great deal of variation data, that accounts for human diversity, leads to various challenging computational tasks, including variant calling and genotyping of newly sequenced individuals. The standard pipelines for addressing these problems include read mapping, which is a computationally expensive procedure. A few mapping-free tools were proposed in recent years to speed up the genotyping process. While such tools have highly efficient run-times, they focus on isolated, bi-allelic SNPs, providing limited support for multi-allelic SNPs, indels, and genomic regions with high variant density. To address these issues, we introduced MALVA, a fast and lightweight mapping-free method to genotype an individual directly from a sample of reads [10]. MALVA is the first mapping-free tool that is able to genotype multi-allelic SNPs and indels, even in high density genomic regions, and to effectively handle a huge number of variants such as those provided by the 1000 Genome Project. An experimental evaluation on whole-genome data shows that MALVA requires one order of magnitude less time to genotype a donor than alignment-based pipelines, providing similar accuracy. Remarkably, on indels, MALVA provides even better results than the most widely adopted variant discovery tools.

Still on the issue of SNP detection, in [25], we developed the positional clustering theory that (i) describes how the extended Burrows–Wheeler Transform (eBWT) of a collection of reads tends to cluster together bases that cover the same genome position, (ii) predicts the size of such clusters, and (iii) exhibits an elegant and precise LCP array based procedure to locate such clusters in the eBWT. Based on this theory, we designed and implemented an alignment-free and reference-free SNP calling method, and we devised a SNP calling pipeline. Experiments on both synthetic and real data show that SNPs can be detected with a simple scan of the eBWT and LCP arrays as, in agreement with our theoretical framework, they are within clusters in the eBWT of the reads. Finally, our tool intrinsically performs a reference-free evaluation of its accuracy by returning the coverage of each SNP. Based on the results of the experiments on synthetic and real data, we conclude that the positional clustering framework can be effectively used for the problem of identifying SNPs, and it appears to be a promising approach for calling other types of variants directly on raw sequencing data.

Finally, variant detection and various related algorithmic problems were extensively explored in the PhD of Leandro I. S. de Lima [2] defended in April 2019.

Bubble generator Bubbles are pairs of internally vertex-disjoint (s, t) -paths in a directed graph, which have many applications in the processing of DNA and RNA data such as variant calling as presented above. Listing and analysing all bubbles in a given graph is usually unfeasible in practice, due to the exponential number of bubbles present in real data graphs. In [4], we proposed a notion of bubble generator set, *i.e.*, a polynomial-sized subset of bubbles from which all the other bubbles can be obtained through a suitable application of a specific symmetric difference operator. This set provides a compact representation of the bubble space of a graph. A bubble generator can be useful in practice, since some pertinent information about all the bubbles can be more conveniently extracted from this compact set. We provided a polynomial-time algorithm to decompose any bubble of a graph into the bubbles of such a generator in a tree-like fashion. Finally, we presented two applications of the bubble generator on a real RNA-seq dataset.

Genome assembly The continuous improvement of long-read sequencing technologies along with the development of ad-doc algorithms has launched a new *de novo* assembly era that promises high-quality genomes. However, it has proven difficult to use only long reads to generate accurate genome assemblies of large, repeat-rich human genomes. To date, most of the human genomes assembled from long error-prone reads add accurate short reads to further improve the consensus quality (polishing). In a paper to be submitted before the end of 2019 (with as main authors A. di Genova and M.-F. Sagot), we report the development of an algorithm for hybrid assembly, WENGAN, and its application to hybrid sequence datasets from four human samples. WENGAN implements efficient algorithms that exploit the sequence information of short and long reads to tackle assembly contiguity as well as consensus quality. We show that the resulting genome assemblies have high contiguity (contig NG50:16.67-62.06 Mb), few assembly errors (contig NGA50:10.9-45.91 Mb), good consensus quality (QV:27.79-33.61), high gene completeness (BUSCO complete: 94.6-95.1%), and consume few computational resources (CPU hours:153-1027). In particular, the WENGAN assembly of the

haploid CHM13 sample achieved a contig NG50 of 62.06 Mb (NGA50:45.91 Mb), which surpasses the contiguity of the current human reference genome (GRCh38 contig NG50:57.88 Mb). Because of its lower cost, WENGAN is an important step towards the democratisation of the *de novo* assembly of human genomes. WENGAN is available at <https://github.com/adigenova/wengan>.

On assembly still, although haplotype-aware genome assembly plays an important role in genetics, medicine and various other disciplines, the generation of haplotype-resolved *de novo* assemblies remains a major challenge. Beyond distinguishing between errors and true sequential variants, one needs to assign the true variants to the different genome copies. Recent work has pointed out that the enormous quantities of traditional NGS read data have been greatly underexploited in terms of haplotig computation so far, which reflects the fact that the methodology for reference independent haplotig computation has not yet reached maturity. We presented in [7] a new approach, called POLYploid genome fitTER (POLYTE) for a *de novo* generation of haplotigs for diploid and polyploid genomes of known ploidy. Our method follows an iterative scheme where in each iteration reads or contigs are joined, based on their interplay in terms of an underlying haplotype-aware overlap graph. Along the iterations, contigs grow while preserving their haplotype identity. Benchmarking experiments on both real and simulated data demonstrate that POLYTE establishes new standards in terms of error-free reconstruction of haplotype-specific sequences. As a consequence, POLYTE outperforms state-of-the-art approaches in various relevant aspects, notably in polyploid settings.

Others Besides the above, we have also explored a proteogenomics workflow for the expert annotation of eukaryotic genomes [18], as well as a technology- and species-independent simulator of sequencing data and genomic variants [42].

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, who was until the end of 2018 postdoc at University 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 Analysis of differential expression of genes is often performed to understand how the metabolic activity of an organism is impacted by a perturbation. However, because the system of metabolic regulation is complex and all changes are not directly reflected in the expression levels, interpreting these data can be difficult. In [26], we presented a new algorithm and computational tool that uses a genome-scale metabolic reconstruction to infer metabolic changes from differential expression data. Using the framework of constraint-based analysis, our method produces a qualitative hypothesis of a change in metabolic activity. In other words, each reaction of the network is inferred to have increased, decreased, or remained unchanged in flux. In contrast to similar previous approaches, our method does not require a biological objective function and does not assign on/off activity states to genes. An implementation is provided and is available online at the address <https://github.com/htpusa/moomin>. We applied the method to three published datasets to show that it successfully accomplishes its two main goals: confirming or rejecting metabolic changes suggested by differentially expressed genes based on how well they fit in as parts of a coordinated metabolic change, as well

as inferring changes in reactions whose genes did not undergo differential expression. The above work was also part of the PhD of Taneli Pusa [3] defended in February 2019.

Metabolic games Game theory is a branch of applied mathematics originally developed to describe and reason about situations where two or more rational agents, the “homo economicus”, are faced with choices and have potentially conflicting goals. All participants want to maximise their own well-being, but are doing so taking into account that everyone else is doing the same. Thus paradoxical, suboptimal, outcomes are possible and even common. Evolutionary game theory was born out of the realisation that rational choice can be replaced by natural selection: in the course of evolution the strategy (phenotype) that would “win” the game would prevail by simply proliferating more successfully thanks to its success in the “game”. It turns out that phenotype prediction in the context of metabolic networks is exactly the type of problem that evolutionary game theory was meant to answer: given a set of choices (as defined by a metabolic network reconstruction), what will be the actual metabolism observed? In other words, if we culture a set of organisms together in a given medium, which are the phenotype(s) that emerge as winners? In [27], we sought to provide a short introduction to both evolutionary game theory and its use in the context of metabolic modelling. This work was also part of the PhD of Taneli Pusa [3].

6.4. Axis 3: (Co)Evolution

Modelling invasion Nowadays, the most used model in studies of the coevolution of hosts and symbionts is phylogenetic tree reconciliation. A crucial issue in this model is that from a biological point of view, reasonable cost values for an event-based reconciliation are not easily chosen. Different methods have been developed to infer such cost values for a given pair of host and symbiont trees, including one we established in the past. However, a major limitation of these methods is their inability to model the “invasion” of different host species by a same symbiont species (referred to as a spread event), which is often observed in symbiotic relations. Indeed, many symbionts are generalist. For instance, the same species of insect may pollinate different species of plants. In a paper currently in preparation, we propose a method, called AMOCOALA, which for a given pair of host and symbiont trees, estimates the frequency of the cophylogenetic events, in presence of spread events, based on an approximate Bayesian computation (ABC) approach that may be more efficient than a classical likelihood method. The algorithm that we propose on one hand provides more confidence in the set of costs to be used for a given pair of host and symbiont trees, while on the other hand, it allows to estimate the frequency of the events even in the case of large datasets. We evaluated our method on both synthetic and real datasets.

Co-divergence and tree topology In reconstructing the common evolutionary history of hosts and symbionts, the current method of choice is the phylogenetic tree reconciliation. In this model, we are given a host tree H , a symbiont tree S , and a function σ mapping the leaves of S to the leaves of H and the goal is to find, under some biologically motivated constraints, a reconciliation, that is a function from the vertices of S to the vertices of H that respects σ and allows the identification of biological events such as co-speciation, duplication and host switch. The maximum co-divergence problem consists in finding the maximum number of co-speciations in a reconciliation. This problem is NP-hard for arbitrary phylogenetic trees and no approximation algorithm is known. In [14], we considered the influence of tree topology on the maximum co-divergence problem. In particular, we focused on a particular tree structure, namely caterpillar, and showed that in this case the heuristics that are mostly used in the literature provide solutions that can be arbitrarily far from the optimal value. We then proved that finding the max co-divergence is equivalent to computing the maximum length of a subsequence with certain properties of a given permutation. This equivalence leads to two consequences: (i) it shows that we can compute efficiently in polynomial time the optimal time-feasible reconciliation, and (ii) it can be used to understand how much the tree topology influences the value of the maximum number of co-speciations.

6.5. Axis 4: Human and animal health

Rare disease studies Minor intron splicing plays a central role in human embryonic development and survival. Indeed, biallelic mutations in RNU4ATAC, transcribed into the minor spliceosomal U4atac snRNA, are responsible for three rare autosomal recessive multiformation disorders named Taybi-Linder (TALS/MOPD1), Roifman (RFMN), and Lowry-Wood (LWS) syndromes, which associate numerous overlapping signs of varying severity. Although RNA-seq experiments have been conducted on a few RFMN patient cells, none have been performed in TALS, and more generally no in-depth transcriptomic analysis of the 700 human genes containing a minor (U12-type) intron had been published as yet. We thus sequenced RNA from cells derived from five skin, three amniotic fluid, and one blood biosamples obtained from seven unrelated TALS cases and from age- and sex-matched controls. This allowed us to describe for the first time the mRNA expression and splicing profile of genes containing U12-type introns, in the context of a functional minor spliceosome. Concerning RNU4ATAC-mutated patients, we showed in [15] that as expected, they display distinct U12-type intron splicing profiles compared to controls, but that rather unexpectedly the mRNA expression levels are mostly unchanged. Furthermore, although U12-type intron missplicing concerns most of the expressed U12 genes, the level of U12-type intron retention is surprisingly low in fibroblasts and amniocytes, and much more pronounced in blood cells. Interestingly, we found several occurrences of introns that can be spliced using either U2, U12, or a combination of both types of splice site consensus sequences, with a shift towards splicing using preferentially U2 sites in TALS patients' cells compared to controls.

This work is part of the PhD of Audric Cologne [1] defended in October 2019.

Cancer studies Circular RNAs (circRNAs) are a class of RNAs that is under increasing scrutiny, although their functional roles are debated. In [30], we analysed RNA-seq data of 348 primary breast cancers and developed a method to identify circRNAs that does not rely on unmapped reads or known splice junctions. We identified 95,843 circRNAs, of which 20,441 were found recurrently. Of the circRNAs that match exon boundaries of the same gene, 668 showed a poor or even negative ($R < 0.2$) correlation with the expression level of the linear gene. An *In silico* analysis showed that only a minority (8.5%) of circRNAs could be explained by known splicing events. Both these observations suggest that specific regulatory processes for circRNAs exist. We confirmed the presence of circRNAs of CNOT2, CREBBP, and RERE in an independent pool of primary breast cancers. We identified circRNA profiles associated with subgroups of breast cancers and with biological and clinical features, such as amount of tumour lymphocytic infiltrate and proliferation index. siRNA-mediated knockdown of circCNOT2 was shown to significantly reduce viability of the breast cancer cell lines MCF-7 and BT-474, further underlining the biological relevance of circRNAs. Furthermore, we found that circular, and not linear, CNOT2 levels are predictive for progression-free survival time to aromatase inhibitor (AI) therapy in advanced breast cancer patients, and found that circCNOT2 is detectable in cell-free RNA from plasma. We showed that circRNAs are abundantly present, show characteristics of being specifically regulated, are associated with clinical and biological properties, and thus are relevant in breast cancer.

Other cancer studies have concerned the automatic discovery of the 100-miRNA signature for cancer classification [21], an Integrative and comparative genomic analysis to identify clinically relevant pulmonary carcinoma groups and unveil the supra-carcinoids [5], [complete with 2 papers not yet entered in Hal], and finally the investigation of new therapeutic interventions that are needed to increase the immunogenicity of tumours and overcome the resistance to these immuno-therapies [29].

Infection studies *Mycoplasma hyopneumoniae* is an economically devastating pathogen in the pig farming industry, however little is known about its relation with the swine host. To improve our understanding on this interaction, we infected epithelial cells with *M. hyopneumoniae* to identify the effects of the infection on the expression of swine genes and miRNAs. In addition, we identified miRNAs differentially expressed (DE) in the extracellular milieu and in exosome-like vesicles released by infected cells. A total of 1,268 genes and 170 miRNAs were DE post-infection ($p < 0.05$). We identified the up-regulation of genes related to redox homeostasis and antioxidant defense, most of them putatively regulated by the transcription factor NRF2. Down-regulated genes were enriched in cytoskeleton and ciliary function, which could partially explain *M. hyopneumoniae* induced ciliostasis. Our predictions showed that DE miRNAs could be regulating the

aforementioned functions, since we detected down-regulation of miRNAs predicted to target antioxidant genes and up-regulation of miRNAs targeting ciliary and cytoskeleton genes. Based on these observations, *M. hyopneumoniae* seems to elicit an antioxidant response induced by NRF2 in infected cells; in addition, we propose that ciliostasis caused by this pathogen might be related to down-regulation of ciliary genes. The paper presenting these results has been submitted and is in revision.

Others Besides the above, a first step towards deep learning assisted genotype-phenotype association in whole genome-sized data has been explored in the context of predicting amyotrophic lateral sclerosis [34].

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 (2018-2020).
- 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: H el ene Touzet, University of Lille and CNRS.
- ERABLE participants: Vincent Lacroix (ERABLE coordinator), Audric Cologne, Eric Cumunel, Alex di Genova, Leandro I. S. de Lima, Arnaud Mary, Marie-France Sagot, Camille Sessegolo, Blerina Sinaimer.
- Type: ANR (2016-2020).
- Web page: <http://bioinfo.cristal.univ-lille.fr/aster/>.

8.2.1.2. 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/>

8.2.1.3. GrR

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

8.2.1.4. 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

8.2.1.5. 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

8.2.1.6. Networks

- Title: Networks
- Coordinator: Michel Mandjes, University of Amsterdam
- ERABLE participant(s): S. Pissis, L. Stougie
- Type: NWO Gravity Program (2014-2024)
- Web page: <https://www.thenetworkcenter.nl/>

8.2.1.7. 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

8.2.1.8. 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

8.2.1.9. *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), 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. *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.

8.2.3.1. *AHeAD*

- Title: efficient Algorithms for HARnessing networked Data
- Coordinator: G. Italiano
- ERABLE participant(s): R. Grossi, G. Italiano
- Type: MUIR PRIN, Italian Ministry of Education, University and Research (2019-2022)
- Web page: <https://sites.google.com/view/aheadproject>

8.2.3.2. *CMACBioSeq*

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

8.2.3.3. *MyOwnResearch*

- Title: MyOwnResearch: Homogeneous subgroup identification in fatigue management across chronic immune diseases through single subject research design
- Coordinator: A. Schönhuth
- ERABLE participant(s): A. Schönhuth
- Type: Health Holland project (2018-2021)
- Web page: Not available

8.2.3.4. *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

8.3. European Initiatives

8.3.1. Collaborations in European Programs, Except FP7 & H2020

8.3.1.1. Pangaia

- Title: Pan-genome Graph Algorithms and Data Integration
- Coordinator: Paola Bonizzoni, University of Milan, Italy
- ERABLE participant(s): S. Pissis, A. Schönhuth, L. Stougie
- Type: H2020 MSCA-RISE (2020-2022)
- Web page: Not available

8.3.2. 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 then became a European Inria Team.

8.4. International Initiatives

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

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. Sinimeri, L. Stougie, A. Viari, I. Ziska
- Web page: <http://team.inria.fr/erable/en/projects/inria-associated-team-compasso/>

8.4.2. Participation in Other 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. This year (2019) is the final one. 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 participates in Network for Organismal Interactions Research (NOIR), a project funded by Conicyt in Chile within the call International Networking between Research Centers. The project started in 2019 and will last until the end of 2020. The coordinator on the Chilean side is Elena Vida from the Universidad Mayor, Santiago, Chile, and the Erable participants are Carol Moraga Quinteros, Mariana Ferrarini and Marie-France Sagot.

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 Portuguese partner. Perseids ended in December 2019.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

In 2019, ERABLE greeted the following International scientists:

- In France: Alexandra Carvalho and Susana Vinga, Assistant and Associate professors resp., Instituto Superior Técnico, Lisbon, Portugal; Helisson Faoro, researcher, Instituto Carlos Chagas, Fiocruz, Paraná, Brazil; Ariel Silber, professor, Universidade de São Paulo, Brazil; Arnaldo Zaha, professor at Universidade Federal do Rio Grande do Sul, Brazil.
- In Italy: Travis Gaggie, Associate professor, Dalhousie University; Nicola Prezza, postdoc, University of Pisa; Elena Arseneva, Assistant professor, St Petersburg State University, Blerina Sinaimeri, Junior Researcher, Inria (see below); Marie-France Sagot, Senior researcher, Inria (see below).
- In the Netherlands: Wiktor Zuba, PhD student, University of Warsaw; Lorraine Ayad, Lecturer, King's College London; Grigorios Loukides, Lecturer, King's College London; Martin Farach-Colton, Professor, Rutgers University; Grigorios Loukides, Lecturer, King's College London; Martin Dyer, Professor, University of Leeds.

8.5.1.1. Internships

In 2019, ERABLE in France greeted the following Internships:

- Phablo Moura, postdoc, University of Campinas, Brazil.
- Diego Pérez and Evelyn Sánchez, PhD students of Elena Vidal, Universidad Mayor, Santiago, Chile.

In the Netherlands, ERABLE greeted the following Internships: Luca Denti, University Bicocca of Milano, Italy, from October 2018 to January 2019, Mick van Dijk, TU Delft, from May 2018 to January 2019, Giulia Barnardini, University Bicocca of Milano, Italy, from September 2018 to November 2019.

8.5.2. Visits to International Teams

8.5.2.1. Sabbatical programme

From July 2019 to June 2020, Blerina Sinaimeri was on Sabbatical at Luiss University to work with Giuseppe Italiano, member of Erable.

8.5.2.2. Research Stays Abroad

In 2019, Marie-France Sagot visited Luiss University for 11 days as Visiting Professor from LUISS University to work with Blerina Sinaimeri who is on Sabbatical at Luiss University from July 2019 to June 2020, and with Giuseppe Italiano, member of Erable. While there, M.-F. Sagot also worked with Alberto Marchetti-Spaccamela from Sapienza University of Rome and from Erable.

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 of 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

- Leen Stougie was co-organiser of MAPSP 2019, Jun 2019, Hotel Zeeuwse Stromen, Renesse; and of the Networks Workshop on Random graphs, counting and sampling, Sep 2019, CWI, Amsterdam.

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 *APF*, *ATMOS*, and *CIAC*.
- Arnaud Mary was a member of the Program Committee of *MFCS*, and *WEPA*.
- Nadia Pisanti was a member of the Program Committee of *BIOINFORMATICS*, *CPM*, *ICCS*, *ISBRA*, *IWOCA*, and *WABI*.
- Marie-France Sagot was a member of the Program Committee of *BIBM*, *CIAC*, *CPM*, *PSC*, *RecombCG*, and *WABI*.

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

- Roberto Grossi is member of the Editorial Board of *Theory of Computing Systems (TOCS)* and of *RAIRO – Theoretical Informatics and Applications*.
- Giuseppe Italiano is member of the Editorial Board of *Algorithmica* and *Theoretical Computer Science*.
- Vincent Lacroix is recommender for *Peer Community in Genomics*, see <https://genomics.peercommunityin.org/>.
- 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: invited talk on “2-Connectivity on Directed Graphs”, 14th Computer Science Symposium in Russia (CSR 2019), Novosibirsk, Russia.
- Nadia Pisanti: Invited talk on “Mapping Reads on a Pan-Genome: Pattern Matching on Degenerate Texts”, 1st Workshop on Computational Pan-Genomics, Bielefeld, Germany; Invited talk on “On-line (approximate) Pattern Matching on Degenerate Texts and Applications”, 14th Workshop on Compression, Text and Algorithms (WCTA), Segovia, Spain.
- Solon Pissis: Invited talk on “Even Faster Elastic-Degenerate String Matching via Fast Matrix Multiplication”, Algorithms Group Seminar Series, 24 Oct 2019, University of Warsaw, Warsaw, Poland; Invited talk on “When linear space is impractical: computing absent words in output-sensitive space”, Bonsai Bioinformatics Seminar Series, 11 Jun 2019, Université de Lille, Lille, France; Invited talk on “Elastic-degenerate strings: a new representation for pattern matching in a collection of similar texts”, Computer Science Seminar Series, Feb 12 2019, University of Pisa, Pisa, Italy.
- Leen Stougie: Invited talk on “Fixed-Order Scheduling on parallel machines”, Workshop on Combinatorial Optimization, 26-27 September 2019, TU Berlin, Germany.
- Cristina Vieira: Invited talk on “Contribution of Transposable element to gene expression in *Drosophila* (and other)”, XI Symposium of Ecology, Genetic and *Drosophila* Evolution, November 2019, Pelotas, Brazil.

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.

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. He is member of the General Board of the Dutch Network on the Mathematics of Operations Research (Landelijk Netwerk Mathematische Besliskunde (LNMB)), and member of the Management Team of the Gravity project Networks.

Alain Viari 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 CoNRS8.

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 the M1 master in bioinformatics (<https://www.bioinfo-lyon.fr/>) and of the following courses (L3: Advanced Bioinformatics, M1: Methods for Data Analysis in Genomics, M1: Methods for Data Analysis in Transcriptomics, M1: Bioinformatics Project, M2: Ethics). He taught 96 hours in 2018-2019 and 192 hours in 2019-2020. Arnaud Mary is responsible for three courses of the Bioinformatics Curriculum at the University (L2: Introduction to Bioinformatics and Biostatistics, M1: Object Oriented Programming, M2: new course on Advanced Algorithms for Bioinformatics) and one at Insa (Discrete Mathematics). He taught 198 hours in 2019. Blerina Sinimeri taught 36 hours in 2019 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 and Audric Cologne were instructors in the NGS data analysis training for the CNRS Formation, a course coordinated by Annabelle Haudry, LBBE (<https://cnrsformation.cnrs.fr/stage-19026-Bioinformatique-pour-le-traitement-de-donnees-de-sequencage-%28NGS%29-Lyon.html>).

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 60 and 100 hours per year, again at the undergraduate and Master levels, in applied mathematics (*e.g.* Operational Research, Advanced Linear Programming), machine learning (Deep Learning) and computational biology (*e.g.* Biological Network Analysis, Algorithms for Genomics).

9.2.2. Supervision

The following PhDs were defended in ERABLE in 2019:

- Jasmijn Baaijens, CWI (supervisor: Alexander Schönhuth), Sep 2019
- Annelieke Baller, Vrije Universiteit Amsterdam (co-supervisor: Leen Stougie), Nov 2019
- Thomas Bosman, Vrije Universiteit Amsterdam (co-supervisor: Leen Stougie), Nov 2019
- Audric Cologne, University of Lyon 1 (funded by Inserm and Inria, co-supervisors: Patrick Edery – Federation of Health Research of Lyon-Est, Vincent Lacroix), Oct 2019
- 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), Apr 2019
- Nikos Parotsidis, University of Rome Tor Vergata, supervisor: Giuseppe Italiano, Mar 2019
- 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), Feb 2019

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)
- Nicolas Homberg, Inra, Inria & University of Lyon 1 (funded by Inra & Inria, co-supervisors: Christine Gaspin at Inra; 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)
- Camille Sessegolo, University of Lyon 1 (funded by ANR Aster; co-supervisors: Vincent Lacroix, Arnaud Mary)
- Michelle Sweering, CWI (co-supervisors: Solon Pissis and Leen Stougie)
- 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 Sinimeri)
- Irene Ziska, University Lyon 1 (funded by Inria Cordi-S, co-supervisors: Susana Vinga – Instituto Superior Técnico at Lisbon; Marie-France Sagot)

9.2.3. Juries

The following are the PhD or HDR juries to which members of ERABLE participated in 2019.

- Vincent Lacroix: External reviewer of the PhD of Patricia Sieber, supervised by Stefan Schuster at Friedrich-Schiller University of Jena, Germany; external reviewer of the PhD of Luca Denti, supervised by Paola Bonizzoni at University Bicocca of Milano, Italy.
- Arnaud Mary: External reviewer of the PhD of Karima Ennaoui, supervised by Lhouari Nourine at University of Clermont-Ferrand, France.
- Marie-France Sagot: External Reviewer of the PhD of Pierre Marijon, University of Lille, France, Dec 2019.
- Leen Stougie: Reading Committee of the PhD of Teun Janssen, TU Delft, Mar 2019; Chair Reading Committee of the PhD of Pieter Kleer, Vrije Universiteit Amsterdam, Sep 2019; Reading Committee of the PhD of Peter van der Gulik, Univ. of Amsterdam, Sep 2019; Chair Reading Committee of the PhD of Maaïke Hoogeboom, Vrije Universiteit Amsterdam, Dec 2019.
- Cristina Vieira: Member of the PhD Committee of Olivier Tabone, Faculté de Médecine Rockefeller, Jan 2019; Member of the PhD Committee of Sébastien Lemaire, ENS Lyon, Mar 2019; External Reviewer of the PhD of Natalia Martinez, Université Paris Sud, Oct 2019.

9.3. Popularization

9.3.1. Interventions

Carol Moraga Quinteros participated in the contest “DESCRYPThèse” of the doctoral school E2M2 of the University of Lyon 1, winning a prize for one of the best presentations in April 2019. The title of the talk was “BrumiR: un algorithme *de novo* pour prédire les petits ARNs sans génome de référence”.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] A. COLOGNE. *Exploration of minor splicing function during embryonic development with the Taybi-Linder Syndrome (TALS) model*, Université de Lyon, October 2019, <https://tel.archives-ouvertes.fr/tel-02363211>

- [2] L. ISHI SOARES DE LIMA. *De novo algorithms to identify patterns associated with biological events in de Bruijn graphs built from NGS data*, Université de Lyon ; Università degli studi di Roma "Tor Vergata", April 2019, <https://tel.archives-ouvertes.fr/tel-02280110>
- [3] T. PUSA. *Modélisation mathématique des impacts de l'environnement à l'aide de réseaux métaboliques et de la théorie des jeux*, Université de Lyon ; Università degli studi La Sapienza (Rome), February 2019, <https://tel.archives-ouvertes.fr/tel-02096971>

Articles in International Peer-Reviewed Journals

- [4] V. ACUÑA, R. GROSSI, G. F. ITALIANO, L. LIMA, R. RIZZI, G. SACOMOTO, M.-F. SAGOT, B. SINAIMERI. *On Bubble Generators in Directed Graphs*, in "Algorithmica", 2019, pp. 1-19 [DOI : 10.1007/s00453-019-00619-Z], <https://hal.inria.fr/hal-02284946>
- [5] N. ALCALA, N. LEBLAY, A. GABRIEL, L. MANGIANTE, D. HERVÁS, T. GIFFON, A.-S. SERTIER, A. FERRARI, J. DERKS, A. GHANTOUS, T. DELHOMME, A. CHABRIER, C. CUENIN, B. ABEDI-ARDEKANI, A. BOLAND, R. OLASO, V. MEYER, J. ALTMULLER, F. LE CALVEZ-KELM, G. DURAND, C. VOEGELE, S. BOYAULT, L. MOONEN, N. LEMAÎTRE, P. LORIMIER, A.-C. TOFFART, A. SOLTERMANN, J. CLEMENT, J. SAENGER, J. FIELD, M. BREVET, C. BLANC-FOURNIER, F. GALATEAU-SALLÉ, N. LE STANG, P. RUSSELL, G. WRIGHT, G. SOZZI, U. PASTORINO, S. LACOMME, J. VIGNAUD, V. HOFMAN, P. HOFMAN, O. T. BRUSTUGUN, M. LUND-IVERSEN, V. THOMAS DE MONTPREVILLE, L. A. MUSCARELLA, P. GRAZIANO, H. H. POPPER, J. STOJSIC, J.-F. DELEUZE, Z. HERCEG, A. VIARI, P. NUERNBERG, G. PELOSI, A.-M. C. DINGEMANS, M. MILIONE, L. ROZ, L. BRČIĆ, M. VOLANTE, M. PAPOTTI, C. CAUX, J. SANDOVAL, H. HERNANDEZ-VARGAS, E. BRAMBILLA, E. SPEEL, N. GIRARD, S. LANTUEJOL, J. MCKAY, M. FOLL, L. FERNANDEZ-CUESTA. *Integrative and comparative genomic analyses identify clinically relevant pulmonary carcinoid groups and unveil the supra-carcinoids*, in "Nature Communications", December 2019, vol. 10, n^o 1, pp. 1-21 [DOI : 10.1038/s41467-019-11276-9], <https://hal.inria.fr/hal-02339242>
- [6] 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", 2019, vol. 27, n^o 3, pp. 268-281 [DOI : 10.1016/j.TIM.2018.11.004], <https://hal.inria.fr/hal-01964983>
- [7] J. A. BAAIJENS, A. SCHÖNHUTH. *Overlap graph-based generation of haplotigs for diploids and polyploids*, in "Bioinformatics", November 2019, vol. 35, n^o 21, pp. 4281-4289 [DOI : 10.1093/BIOINFORMATICS/BTZ255], <https://hal.inria.fr/hal-02344853>
- [8] A. C. BALLER, M. VAN EE, M. HOOGEBOOM, L. STOUGIE. *Complexity of inventory routing problems when routing is easy*, in "Networks", October 2019, pp. 1-11 [DOI : 10.1002/NET.21908], <https://hal.inria.fr/hal-02422721>
- [9] S. K. BARUAH, V. BONIFACI, R. BRUNI, A. MARCHETTI-SPACCAMELA. *ILP models for the allocation of recurrent workloads upon heterogeneous multiprocessors*, in "Journal of Scheduling", April 2019, vol. 22, n^o 2, pp. 195-209 [DOI : 10.1007/s10951-018-0593-x], <https://hal.inria.fr/hal-02339161>
- [10] G. BERNARDINI, P. BONIZZONI, L. DENTI, M. PREVITALI, A. SCHÖNHUTH. *MALVA: Genotyping by Mapping-free Allele Detection of Known Variants*, in "iScience", 2019, vol. 18, pp. 20-27 [DOI : 10.1016/j.ISCI], <https://hal.inria.fr/hal-02344254>

- [11] G. BERNARDINI, N. PISANTI, S. PISSIS, G. ROSONE. *Approximate pattern matching on elastic-degenerate text*, in "Theoretical Computer Science", August 2019, pp. 1-30 [DOI : 10.1016/J.TCS.2019.08.012], <https://hal.inria.fr/hal-02298622>
- [12] V. BONIFACI, A. WIESE, S. K. BARUAH, A. MARCHETTI-SPACCAMELA, S. STILLER, L. STOUGIE. *A Generalized Parallel Task Model for Recurrent Real-time Processes*, in "ACM Transactions on Parallel Computing", June 2019, vol. 6, n^o 1, pp. 63-72 [DOI : 10.1109/RTSS.2012.59], <https://hal.inria.fr/hal-02347467>
- [13] M. BONNIN, N. FARES, B. TESTONI, Y. ESTORNES, K. WEBER, B. VANBERVLIET, L. LEFRANCOIS, A. GARCIA, A. KFOURY, F. PEZ, I. COSTE, P. SAINTIGNY, A. VIARI, K. LANG, B. GUEY, V. HERVIEU, B. BANCEL, B. BARTOCH, D. DURANTEL, T. RENNO, P. MERLE, S. LEBECQUE. *Toll-like receptor 3 downregulation is an escape mechanism from apoptosis during hepatocarcinogenesis*, in "Journal of Hepatology", October 2019, vol. 71, n^o 4, pp. 763-772 [DOI : 10.1016/J.JHEP.2019.05.031], <https://hal.inria.fr/hal-02367101>
- [14] T. CALAMONERI, A. MONTI, B. SINAIMERI. *Co-divergence and tree topology*, in "Journal of Mathematical Biology", August 2019, vol. 79, n^o 3, pp. 1149-1167 [DOI : 10.1007/s00285-019-01385-w], <https://hal.inria.fr/hal-02298643>
- [15] A. COLOGNE, C. BENOIT-PILVEN, A. BESSON, A. PUTOUX, A. CAMPAN-FOURNIER, M. B. BOBER, C. E. DE DIE-SMULDERS, A. PAULUSSEN, L. PINSON, A. TOUTAIN, C. M. ROIFMAN, A.-L. LEUTENEGGER, S. MAZOYER, P. EDERY, V. LACROIX. *New insights into minor splicing-a transcriptomic analysis of cells derived from TALS patients*, in "RNA", 2019, pp. 1-21 [DOI : 10.1261/RNA.071423.119], <https://hal.inria.fr/hal-02305628>
- [16] M. FABLET, A. JACQUET, R. REBOLLO, A. A. HAUDRY, C. REY, J. SALCES-ORTIZ, P. BAJAD, N. BURLET, M. JANTSCH, M. P. GARCIA GUERREIRO, C. VIEIRA. *Dynamic Interactions Between the Genome and an Endogenous Retrovirus: Tirant in Drosophila simulans Wild-Type Strains*, in "G3", March 2019, vol. 9, n^o 3, pp. 855-865 [DOI : 10.1534/G3.118.200789], <https://hal.archives-ouvertes.fr/hal-02004384>
- [17] L. GEORGIADIS, G. F. ITALIANO, A. KARANASIOU. *Approximating the smallest 2-vertex connected spanning subgraph of a directed graph*, in "Theoretical Computer Science", September 2019, pp. 1-16 [DOI : 10.1016/J.TCS.2019.09.040], <https://hal.inria.fr/hal-02335015>
- [18] L. GUILLOT, L. DELAGE, A. VIARI, Y. VANDENBROUCK, E. COM, A. A. RITTER, R. LAVIGNE, D. MARIE, P. PETERLONGO, P. POTIN, C. PINEAU. *Peptimapper: proteogenomics workflow for the expert annotation of eukaryotic genomes*, in "BMC Genomics", January 2019, vol. 20, n^o 1, 56 p. [DOI : 10.1186/s12864-019-5431-9], <https://hal.inria.fr/hal-01987197>
- [19] E. LERAT, J. CASACUBERTA, C. CHAPARRO, C. VIEIRA. *On the Importance to Acknowledge Transposable Elements in Epigenomic Analyses*, in "Genes", March 2019, vol. 10, n^o 4, 258 p. [DOI : 10.3390/GENES10040258], <https://hal.archives-ouvertes.fr/hal-02093613>
- [20] L. LIMA, C. MARCHET, S. CABOCHE, C. DA SILVA, B. ISTACE, J.-M. AURY, H. TOUZET, R. CHIKHI. *Comparative assessment of long-read error correction software applied to Nanopore RNA-sequencing data*, in "Briefings in Bioinformatics", June 2019, pp. 1-18 [DOI : 10.1093/BIB/BBZ058], <https://hal.archives-ouvertes.fr/hal-02394395>

- [21] A. LOPEZ-RINCON, M. MARTINEZ-ARCHUNDIA, G. U. MARTINEZ-RUIZ, A. SCHÖNHUTH, A. TONDA. *Automatic discovery of 100-miRNA signature for cancer classification using ensemble feature selection*, in "BMC Bioinformatics", December 2019, vol. 20, n^o 1, pp. 1-17 [DOI : 10.1186/s12859-019-3050-8], <https://hal.inria.fr/hal-02344257>
- [22] P. MARIN, J. GENITONI, D. D. BARLOY, S. MAURY, P. GIBERT, C. K. GHALAMBOR, C. VIEIRA. *Biological Invasion: The Influence of the Hidden Side of the (Epi)Genome*, in "Functional Ecology", 2019, pp. 1-55, forthcoming [DOI : 10.1111/1365-2435.13317], <https://hal-agrocampus-ouest.archives-ouvertes.fr/hal-02063295>
- [23] A. MARINO, L. VERSARI, A. CONTE, R. GROSSI, G. PUNZI, T. UNO. *Polynomial-Delay Enumeration of Maximal Common Subsequences*, in "SIAM Journal on Discrete Mathematics", October 2019, vol. 33, n^o 2, pp. 189-202 [DOI : 10.1007/978-3-030-32686-9_14], <https://hal.inria.fr/hal-02338458>
- [24] A. MARY, Y. STROZECKI. *Efficient enumeration of solutions produced by closure operations*, in "Discrete Mathematics and Theoretical Computer Science", June 2019, pp. 1-30, <https://arxiv.org/abs/1712.03714> - 30 pages, 1 figure. Long version of the article arXiv:1509.05623 of the same name which appeared in STACS 2016. Final version for DMTCS journal [DOI : 10.23638/DMTCS-21-3-22], <https://hal.inria.fr/hal-02373737>
- [25] N. PREZZA, N. PISANTI, M. SCIORTINO, G. ROSONE. *SNPs detection by eBWT positional clustering*, in "Algorithms for Molecular Biology", December 2019, vol. 14, n^o 1, pp. 1-13 [DOI : 10.1186/s13015-019-0137-8], <https://hal.inria.fr/hal-02335605>
- [26] T. PUSA, M. G. FERRARINI, R. ANDRADE, A. MARY, A. MARCHETTI-SPACCAMELA, L. STOUGIE, M.-F. SAGOT. *MOOMIN – Mathematical exploration of 'Omics data on a Metabolic Network*, in "Bioinformatics", August 2019, pp. 1-10 [DOI : 10.1093/BIOINFORMATICS/BTZ584], <https://hal.inria.fr/hal-02284835>
- [27] T. PUSA, M. WANNAGAT, M.-F. SAGOT. *Metabolic Games*, in "Frontiers in Applied Mathematics and Statistics", April 2019, vol. 5, pp. 1-13 [DOI : 10.3389/FAMS.2019.00018], <https://hal.inria.fr/hal-02336595>
- [28] C. SESSEGOLO, C. CRUAUD, C. DA SILVA, A. COLOGNE, M. DUBARRY, T. DERRIEN, V. LACROIX, J.-M. AURY. *Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules*, in "Scientific Reports", December 2019, vol. 9, n^o 1, pp. 1-12 [DOI : 10.1038/s41598-019-51470-9], <https://hal.inria.fr/hal-02335574>
- [29] T. SHEKARIAN, E. SIVADO, A.-C. JALLAS, S. DEPIL, J. KIELBASSA, I. JANOUÉIX-LEROSEY, G. HUTTER, N. GOUTAGNY, C. BERGERON, A. VIARI, S. VALSESIA-WITTMANN, C. CAUX, A. MARABELLE. *Repurposing rotavirus vaccines for intratumoral immunotherapy can overcome resistance to immune checkpoint blockade*, in "Science Translational Medicine", October 2019, vol. 11, n^o 515, eaat5025 p. [DOI : 10.1126/SCITRANSLMED.AAT5025], <https://hal.inria.fr/hal-02347852>
- [30] M. SMID, S. M. WILTING, K. UHR, F. G. RODRÍGUEZ-GONZÁLEZ, V. DE WEERD, W. J. PRAGER-VAN DER SMISSEN, M. VAN DER VLUGT-DAANE, A. VAN GALEN, S. NIK-ZAINAL, A. P. BUTLER, S. MARTIN, H. R. DAVIES, J. STAAF, M. J. VAN DE VIJVER, A. L. RICHARDSON, G. MACGROGAN, R. SALGADO, G. G. VAN DEN EYNDEN, C. A. PURDIE, A. M. THOMPSON, C. CALDAS, P. N. SPAN, F. C. SWEEP, P. T. SIMPSON, S. R. LAKHANI, S. VAN LAERE, C. DESMEDT, A. PARADISO, J. EYFJORD, A. BROEKS, A. VINCENT-SALOMON, A. P. FUTREAL, S. KNAPPSKOG, T. KING, A. VIARI, A.-L. BØRRESEN-DALE, H. G. STUNNENBERG, M. STRATTON, J. A. FOEKENS, A. M. SIEUWERTS, J. W.

- MARTENS. *The circular RNome of primary breast cancer*, in "Genome Research", March 2019, vol. 29, n^o 3, pp. 356-366 [DOI : 10.1101/GR.238121.118], <https://hal.inria.fr/hal-02338656>
- [31] L. TONON, G. FROMONT, S. BOYVAULT, 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", January 2019, vol. 75, n^o 1, pp. 11-15 [DOI : 10.1016/J.EURURO.2018.08.026], <https://hal.inria.fr/hal-01921597>
- [32] O. TRÉDAN, Q. WANG, D. PISSALOUX, P. CASSIER, A. DE LA FOUCHARDIÈRE, J. FAYETTE, F. DESSEIGNE, I. RAY-COQUARD, C. DE LA FOUCHARDIÈRE, D. FRAPPAZ, P. E. HEUDEL, A. BONNEVILLE-LEVAR, A. FLECHON, M. SARABI, P. GUIBERT, T. BACHELOT, M. PÉROL, B. YOU, N. BONNIN, O. COLLARD, C. LEYRONNAS, V. ATTIGNON, C. BAUDET, E. SOHIER, J.-P. VILLEMEN, A. VIARI, S. BOYVAULT, S. LANTUEJOUL, S. PAINDAVOINE, I. TREILLEUX, C. RODRIGUEZ, V. AGRAPART, V. CORSET, G. GARIN, S. CHABAUD, D. PEROL, J.-Y. BLAY. *Molecular screening program to select molecular-based recommended therapies for metastatic cancer patients: analysis from the ProFiLER trial*, in "Annals of Oncology", May 2019, vol. 30, n^o 5, pp. 757-765 [DOI : 10.1093/ANNONC/MDZ080], <https://hal.inria.fr/hal-02367110>
- [33] 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", 2019, pp. 1-11 [DOI : 10.1109/TCBB.2018.2838667], <https://hal.inria.fr/hal-01842451>
- [34] B. YIN, M. BALVERT, R. VAN DER SPEK, B. E. DUTILH, S. BOHTE, J. VELDINK, A. SCHÖNHUTH. *Using the structure of genome data in the design of deep neural networks for predicting amyotrophic lateral sclerosis from genotype*, in "Bioinformatics", July 2019, vol. 35, n^o 14, pp. i538-i547 [DOI : 10.1093/BIOINFORMATICS/BTZ369], <https://hal.inria.fr/hal-02344253>
- International Conferences with Proceedings**
- [35] A. ABBOUD, L. GEORGIADIS, G. F. ITALIANO, R. KRAUTHGAMER, N. PAROTSIDIS, O. TRABELSI, P. UZNAŃSKI, D. WOLLEB-GRAF. *Faster Algorithms for All-Pairs Bounded Min-Cuts*, in "ICALP 2019 - 46th International Colloquium on Automata, Languages and Programming", Patras, Greece, July 2019, pp. 1-15 [DOI : 10.4230/LIPIcs.ICALP.2019.7], <https://hal.inria.fr/hal-02335025>
- [36] G. BERNARDINI, P. GAWRYCHOWSKI, N. PISANTI, S. PISSIS, G. ROSONE. *Even Faster Elastic-Degenerate String Matching via Fast Matrix Multiplication*, in "ICALP 2019 - 46th International Colloquium on Automata, Languages and Programming", Patras, Greece, July 2019, pp. 1-15 [DOI : 10.4230/LIPIcs.ICALP.2019.21], <https://hal.inria.fr/hal-02298621>
- [37] M. BONAMY, N. BOUSQUET, M. HEINRICH, T. ITO, Y. KOBAYASHI, A. MARY, M. MÜHLENTHALER, K. WASA. *The Perfect Matching Reconfiguration Problem*, in "MFCS 2019 - 44th International Symposium on Mathematical Foundations of Computer Science", Aachen, Germany, August 2019, pp. 1-14 [DOI : 10.4230/LIPIcs.MFCS.2019.80], <https://hal.inria.fr/hal-02335588>
- [38] A. CONTE, R. GROSSI, M. M. KANTÉ, A. MARINO, T. UNO, K. WASA. *Listing Induced Steiner Subgraphs as a Compact Way to Discover Steiner Trees in Graphs*, in "MFCS 2019 - 44th International Symposium on Mathematical Foundations of Computer Science", Aachen, Germany, August 2019, pp. 1-14 [DOI : 10.4230/LIPIcs.MFCS.2019.73], <https://hal.inria.fr/hal-02335601>

- [39] A. CONTE, R. GROSSI, A. MARINO, L. TATTINI, L. VERSARI. *A Fast Discovery Algorithm for Large Common Connected Induced Subgraphs*, in "WEPA 2019 - Workshop on Enumeration Problems & Applications", Awaji Island, Japan, October 2019, pp. 1-26, <https://hal.inria.fr/hal-02338435>
- [40] A. CONTE, R. GROSSI, G. PUNZI, T. UNO. *Polynomial-Delay Enumeration of Maximal Common Subsequences*, in "SPIRE 2019 - 26th International Symposium on String Processing and Information Retrieval", Segovia, Spain, October 2019, pp. 189-202 [DOI : 10.1007/978-3-030-32686-9_14], <https://hal.inria.fr/hal-02338437>
- [41] M. EQUI, R. GROSSI, V. MAKINEN. *On the Complexity of Exact Pattern Matching in Graphs: Binary Strings and Bounded Degree*, in "ICALP 2019 - 46th International Colloquium on Automata, Languages and Programming", Patras, Greece, July 2019, pp. 1-15, <https://hal.inria.fr/hal-02338498>
- [42] F. GERACI, R. MASSIDDA, N. PISANTI. *Technology and Species independent Simulation of Sequencing data and Genomic Variants*, in "BIBE 2019 - 19th annual IEEE International Conference on BioInformatics and BioEngineering", Athens, Greece, IEEE, October 2019, pp. 1-8, <https://hal.inria.fr/hal-02336600>
- [43] N. HJULER, G. F. ITALIANO, N. PAROTSIDIS, D. SAULPIC. *Dominating Sets and Connected Dominating Sets in Dynamic Graphs*, in "STACS 2019 - 36th International Symposium on Theoretical Aspects of Computer Science", Berlin, Germany, March 2019, pp. 1-17 [DOI : 10.4230/LIPICS.STACS.2019.35], <https://hal.inria.fr/hal-02335028>
- [44] G. F. ITALIANO, S. LATTANZI, V. MIRROKNI, N. PAROTSIDIS. *Dynamic Algorithms for the Massively Parallel Computation Model*, in "SPAA 2019 - 31st ACM Symposium on Parallelism in Algorithms and Architectures", Phoenix, France, ACM Press, June 2019, pp. 49-58, <https://arxiv.org/abs/1905.09175> [DOI : 10.1145/3323165.3323202], <https://hal.inria.fr/hal-02335011>

Scientific Books (or Scientific Book chapters)

- [45] N. PISANTI. *Algorithms Foundations*, in "Encyclopedia of Bioinformatics and Computational Biology", Elsevier, 2019, vol. 1, pp. 1-4 [DOI : 10.1016/B978-0-12-809633-8.20315-4], <https://hal.inria.fr/hal-01964689>