

2025 Activity Report


RESEARCH CENTRE: Inria Lyon Centre

IN PARTNERSHIP WITH: Université Claude Bernard (Lyon 1), Institut national des sciences appliquées de Lyon, Centrum Wiskunde & Informatica, Université de Rome la Sapienza

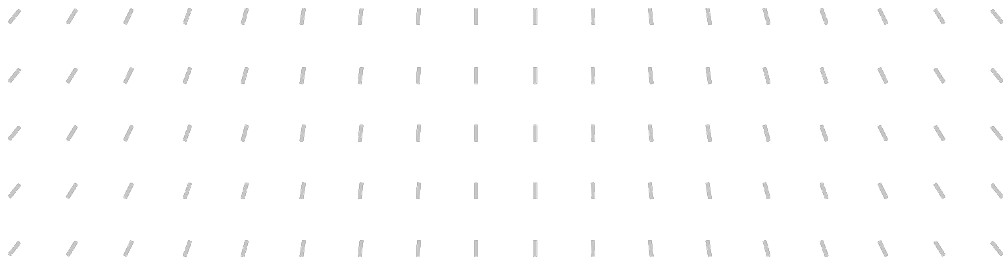

Project-Team

ERABLE

European Research team in Algorithms and Biology,
formal and Experimental



In collaboration with Laboratoire de Biométrie et Biologie Evolutive (LBBE)



Project-Team ERABLE

Creation of the Project-Team: 2015 July 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

- 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. – Digital health
 - B2.2. – Physiology and diseases
 - B2.2.3. – Cancer
 - B2.2.4. – Infectious diseases, Virology
 - B2.3. – Epidemiology

Contents

Project-Team ERABLE	1
1 Team members, visitors, external collaborators	5
2 Overall objectives	6
3 Research program	6
3.1 Two main goals	6
3.2 Different research axes	7
4 Application domains	9
4.1 Biology and Health	9
5 Social and environmental responsibility	9
5.1 Footprint of research activities	9
5.2 Expected impact of research results	9
6 Highlights of the year	10
7 Latest software developments, platforms, open data	10
7.1 Latest software developments	10
7.1.1 AmoCoala	10
7.1.2 ASPefm	11
7.1.3 BrumiR	11
7.1.4 Caldera	11
7.1.5 Capybara	12
7.1.6 Cassis	12
7.1.7 Coala	12
7.1.8 Cycads	13
7.1.9 DBGWAS	13
7.1.10 Eucalypt	13
7.1.11 Fast-SG	13
7.1.12 Gobbolino-Touché	14
7.1.13 HgLib	14
7.1.14 KissDE	14
7.1.15 KisSplice	15
7.1.16 KisSplice2RefGenome	15
7.1.17 KisSplice2RefTranscriptome	15
7.1.18 MetExplore	16
7.1.19 MetHg	16
7.1.20 Mirinho	16
7.1.21 Momo	16
7.1.22 Moomin	17
7.1.23 MultiPus	17
7.1.24 paSAMcs	17
7.1.25 Pitufolandia	18
7.1.26 Sasita	18
7.1.27 Smile	18
7.1.28 Totoro	18
7.1.29 Wengan	19

8	New results	19
8.1	General comments	19
8.2	General theoretical result	19
8.2.1	The Tape Reconfiguration Problem and its consequences for Dominating Set Reconfiguration	20
8.3	Axis 1: (Pan)Genomics and transcriptomics in general	20
8.3.1	Missing value replacement in strings and applications	20
8.3.2	McDag: Indexing maximal common subsequences for k strings	21
8.3.3	Dynamic programming alignments with skips	21
8.3.4	Models and algorithms for managing repeats in the <i>de novo</i> assembly of transcriptomes	21
8.4	Axis 2: Metabolism and (post)transcriptional regulation	22
8.4.1	Growth Balanced Analysis (GBA) and Elementary Growth Modes (EGMs)	22
8.4.2	Modelling energy metabolism dysregulations in neuromuscular diseases – A case study of calpainopathy	22
8.4.3	Logic programming-based Minimal Cut Sets to identify therapeutic targets in oncology	23
8.4.4	Systems biology in identifying metabolic reprogramming signatures resulting from the infection of human macrophages and fibroblasts by <i>Leishmania</i>	23
8.4.5	Ontologies and Genome-scale Metabolic Models (GEMs)	24
8.4.6	Production of polyhydroxyalkanoates by <i>Halomonas</i> sp. HG01 using various carbon sources: metabolic and genomic analysis	24
8.5	Axis 3: (Co)Evolution	24
8.5.1	Search for photosynthesis-related protein sites through chloroplast phylogenomics	25
8.6	Axis 4: Health in general	25
8.6.1	Broad versus limited gene panels to guide treatment in patients with advanced solid tumors: a randomized controlled trial	25
9	Bilateral contracts and grants with industry	26
9.1	Bilateral Grants with Industry	26
10	Partnerships and cooperations	26
10.1	International research visitors	26
10.1.1	Visits of international scientists	26
10.2	National initiatives	27
10.2.1	PEPR-ANR	27
10.2.2	ITMO aviesan	27
10.2.3	Others	27
11	Dissemination	27
11.1	Promoting scientific activities	28
11.1.1	Scientific events: organisation	28
11.1.2	Scientific expertise	29
11.1.3	Research administration	29
11.2	Teaching - Supervision - Juries - Educational and pedagogical outreach	29
11.2.1	Teaching	29
11.2.2	Supervision	30
11.2.3	Juries	30
11.3	Popularization	31
11.3.1	Specific official responsibilities in science outreach structures	31
11.3.2	Productions (articles, videos, podcasts, serious games, ...)	31
11.3.3	Participation in Live events	31
11.3.4	Others science outreach relevant activities	31
12	Scientific production	31
12.1	Publications of the year	31

1 Team members, visitors, external collaborators

Research Scientists

- Marie-France Sagot [Team leader, INRIA, Senior Researcher, until Oct 2025, HDR]
- Marie-France Sagot [Team leader, INRIA, Emeritus, from Oct 2025, HDR]
- Solon Pissis [CWI, Senior Researcher]
- Leen Stougie [CWI, Emeritus]
- Alain Viari [INRIA, Senior Researcher]

Faculty Members

- Sabine Peres [Team leader, UNIV LYON I, Professor, from Oct 2025, HDR]
- Roberto Grossi [UNIV PISE, Professor]
- Giuseppe Italiano [UNIV LUISS, Professor]
- Vincent Lacroix [UNIV LYON I, Associate Professor, HDR]
- Alberto Marchetti Spaccamela [SAPIENZA ROME, Professor, until Oct 2025]
- Alberto Marchetti-Spaccamela [SAPIENZA ROME, from Oct 2025, Emeritus]
- Arnaud Mary [UNIV LYON I, Associate Professor]
- Sabine Peres [UNIV LYON I, Professor, until Oct 2025, HDR]
- Nadia Pisanti [UNIV PISE, Associate Professor]
- Cristina Vieira [UNIV LYON I, Professor, HDR]

PhD Students

- Emma Crisci [INRIA]
- Sasha Darmon [UNIV LYON I]
- Pierre Gérenton [UNIV LYON I]
- Camille Siharath [UNIV LYON I]

Technical Staff

- François Gindraud [INRIA, Engineer]

Interns and Apprentices

- Arnaud Patey [INRIA, Intern, from May 2025 until Jul 2025]

Administrative Assistant

- Cecilia Navarro [INRIA]

External Collaborators

- Laurent Jacob [CNRS, HDR]
- Susana Vinga [Instituto Superior Técnico. Lisbon]

2 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 living systems, or systems that have been described as being at the edge of life such as viruses, or else living systems and chemical compounds (environment). It also includes the interaction between cells within a multicellular organism, or between transposable elements and their host genome.

The application objective of ERABLE is, through the use of mathematical models and algorithms, to better understand such close and often persistent interactions, with a longer term aim 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 objective 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 objective 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 sets of research goals that currently cover four main axes. We present here the research goals.

The first is related to the original areas of expertise of the team, namely combinatorial and statistical modelling and algorithms.

The second set of goals concern its main Life Science interest which is to better understand interactions

between living systems and their environment. This includes close and often persistent interactions between two living systems (symbiosis), interactions between living systems and viruses, and interactions between living systems and chemical compounds. It also includes interactions between cells within a multicellular organism, or interactions between transposable elements and their host genome.

Two major steps are constantly involved in the research done by the team: a first one of modelling (*i.e.* translating) a Life Science problem into a mathematical one, and a second of algorithm analysis and design. The algorithms developed are then applied to the questions of interest in Life Science using data from the literature or from collaborators. More recently, thanks to the recruitment of young researchers (PhD students and postdocs) in biology, the team has become able to start doing experiments and producing data or validating some of the results obtained on its own.

From a methodological point of view, the main characteristic of the team is to consider that, once a model is selected, the algorithms to explore such model should, whenever possible, be exact in the answer provided as well as exhaustive when more than one exists for a more accurate interpretation of the results. More recently, the team has also become interested in exploring the interface between exact algorithms on one hand, and probabilistic or statistical ones on the other such as used in machine learning approaches, notably “interpretable” versions thereof.

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. The first three axes are: (pan)genomics and transcriptomics in general, 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 may be artificially split into two different Axes.

Axis 1: (Pan)Genomics and transcriptomics in general

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. This means that at the modelling step, we need to consider the possibility, or the probability of errors or of missing information. 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: Health in general

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, 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 two different topics: (i) Infectiology, (ii) and Cancer. A third topic started a few years ago in collaboration with researchers from different universities and institutions in Brazil, and concerns tropical diseases, notably related to *Trypanosoma cruzi* (Chagas disease). This topic started to be developed more strongly from 2022 on, notably through the collaboration with Ariel Silber, full professor at the Department of Parasitology of the University of São Paulo, with whom we have projects in common, and since the middle of 2021 a PhD student in co-supervision with M.-F. Sagot from ERABLE. This student is Gabriela Torres Montanaro. Both Gabriela and Ariel have been visiting ERABLE at different occasions and will continue to do so, sometimes for long periods especially in the case of Gabriela.

Among the other two topics, 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 to be interested in infections affecting plants this time, and more particularly vine plants. Cancer on the other hand 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 responsibility of his team at CLB and pursued the main projects he had started.

Notice however that as concerns cancer, at the end of 2021 (October 1st), a new member joined the ERABLE team as full professor in the LBBE - University of Lyon, namely Sabine Peres. Sabine has also been working on cancer, in her case from a perspective of metabolism, in collaboration with Laurent Schwartz (Assistance Publique - Hôpitaux de Paris) and with Mario Jolicoeur, (Polytechnique Montréal, Canada).

Within Inria and beyond, the first application and the third one (Infectiology and Tropical diseases) may be seen as unique because of their specific focus (resp. microbiome and respiratory tract of swines / vine plants on one hand). In the first case, such uniqueness is also related to the fact that the work done involves a strong computational part but also experiments that in some cases (respiratory tract of swines) were *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, cancer, and since more recently, tropical diseases notably related to *Trypanosoma cruzi*.

5 Social and environmental responsibility

5.1 Footprint of research activities

There are three axes on which we would like to focus in the coming years.

Travelling is essential for the team, which is European and has many international collaborations. We would however like to continue to develop as much as possible travelling by train or even car. This is something we do already, for instance between Lyon and Amsterdam by train, and that we have done in the past, such as for instance between Lyon and Pisa by car, and between Rome and Lyon by train, or even in the latter case once between Rome and Amsterdam!

Computing is also essential for the team. We would like to continue our effort to produce resource-frugal software and develop better guidelines for the end users of our software so that they know better under which conditions our software is expected to be adapted, and which more resource-frugal alternatives exist, if any.

Having an impact on how data are produced is also an interest of the team. Much of the data produced is currently only superficially analysed. Generating smaller datasets and promoting data reuse could avoid not only data waste, but also economise on computer time and energy required to produce such data.

5.2 Expected impact of research results

As indicated earlier, the overall objective of the team is to arrive at a better understanding of close and often persistent interactions among living systems, between such living systems and viruses, between living systems and chemical compounds (environment), among cells within a multicellular organism, and between transposable elements and their host genome. There is another longer-term objective, much more difficult and riskier, a “dream” objective whose underlying motivation may be seen as social and is also environmental.

The main idea we thus wish to explore is inspired by the one universal concept underlying life. This is the concept of survival. Any living organism has indeed one single objective: to remain alive and reproduce. Not only that, any living organism is driven by the need to give its descendants the chance to perpetuate themselves. As such, no organism, and more in general, no species can be considered as “good” or “bad” in

itself. Such concepts arise only from the fact that resources, some of which may be shared among different species, are of limited availability. Conflict thus seems inevitable, and “war” among species the only way towards survival.

However, this is not true in all cases. Conflict is often observed, even actively pursued by, for instance, humans. Two striking examples that have been attracting attention lately, not necessarily in a way that is positive for us, are related to the use of antibiotics on one hand, and insecticides on the other, both of which, especially but not only the second can also have disastrous environmental consequences. Yet cooperation, or at least the need to stop distinguishing between “good” (mutualistic) and “bad” (parasitic) interactions appears to be, and indeed in many circumstances is of crucial importance for survival. The two questions which we want to address are: (i) what happens to the organisms involved in “bad” interactions with others (for instance, their human hosts) when the current treatments are used, and (ii) can we find a non-violent or cooperative way to treat such diseases?

Put in this way, the question is infinitely vast. It is not completely utopic. We had the opportunity in recent years to discuss such question with notably biologists with whom we were involved in two European projects (namely **BachBerry**, and **MicroWine**). In both cases, we had examples of bacteria that are "bad" when present in a certain environment, and "good" when the environment changes. In one of the cases at least, related to vine plants, such change in environment seems to be related to the presence of other bacteria. This idea is already explored in agriculture to avoid the use of insecticide. Such exploration is however still relatively limited in terms of scope, and especially, has not yet been fully investigated scientifically.

The aim will be to reach some proofs of concepts, which may then inspire others, including ourselves on a longer term, to pursue research along this line of thought. Such proofs will in themselves already require to better understand what is involved in, and what drives or influences any interaction.

6 Highlights of the year

The research of all team members, in particular of PhD students or Postdocs, is important for us and we prefer not to highlight any in particular.

7 Latest software developments, platforms, open data

We indicate in this section all the software that is either entirely new, or that is being constantly used or maintained and therefore usually continues to have new features or updates. ERABLE does not have any platform and the data we use comes either from the literature or from collaborators.

7.1 Latest software developments

7.1.1 AmoCoala

Name: Associations get Multiple for Our COALA

Keyword: Evolution

Functional Description: Despite an increasingly vaster literature on cophylogenetic reconstructions for studying host-parasite associations, understanding the common evolutionary history of such systems remains a problem that is far from being solved. Many of the most used algorithms do the host-parasite reconciliation analysis using an event-based model, where the events include in general (a subset of) cospeciation, duplication, loss, and host-switch. All known event-based methods then assign a cost to each type of event in order to find a reconstruction of minimum cost. The main problem with this approach is that the cost of the events strongly influence the reconciliation obtained. To deal with this problem, we developed an algorithm, called AMOCOALA, for estimating the frequency of the events based on an approximate Bayesian computation approach in presence of multiple associations.

URL: <https://team.inria.fr/erable/en/software/amocoala/>

Publication: [hal-03673256](https://hal.archives-ouvertes.fr/hal-03673256)

Contact: Blerina Sinimeri

Participants: Laura Urbini, Marie-France Sagot, Catherine Matias, Blerina Sinimeri

7.1.2 ASPefm

Keywords: Metabolic networks, ASP - Answer Set Programming

Functional Description: Elementary Flux Modes are minimal sets of enzymes that operate at steady state with all irreversible reactions proceeding in the appropriate direction. The enumeration of EFMs is a difficult task. It requires the resolution of combinatorial problems on metabolic networks, and the integration of appropriate biological constraints to help calculations. We propose to use the SAT-based power of ASP constraint logic programming resolution to reduce the hurdle of obtaining pathways of interest with EFMs on large-scale networks.

URL: <https://gitlab.inria.fr/erable/aspefm>

Contact: Sabine Peres

Participants: Maxime Mahout, Emma Crisci

7.1.3 BrumiR

Name: A toolkit for de novo discovery of microRNAs from sRNA-seq data.

Keywords: Bioinformatics, Structural Biology, Genomics

Functional Description: BRUMIR is an algorithm that is able to discover miRNAs directly and exclusively from sRNA-seq data. It was benchmarked with datasets encompassing animal and plant species using real and simulated sRNA-seq experiments. The results show that BRUMIR reaches the highest recall for miRNA discovery, while at the same time being much faster and more efficient than the state-of-the-art tools evaluated. The latter allows BRUMIR to analyse a large number of sRNA-seq experiments, from plant or animal species. Moreover, BRUMIR detects additional information regarding other expressed sequences (sRNAs, isomiRs, etc.), thus maximising the biological insight gained from sRNA-seq experiments. Finally, when a reference genome is available, BRUMIR provides a new mapping tool (BRUMIR2REFERENCE) that performs a posteriori an exhaustive search to identify the precursor sequences.

URL: <https://github.com/camoragaq/BrumiR>

Publication: [hal-03831360](https://hal.archives-ouvertes.fr/hal-03831360)

Contact: Carol Moraga Quinteros

Participants: Carol Moraga Quinteros, Marie-France Sagot

7.1.4 Caldera

Keywords: Genomics, Graph algorithmics

Functional Description: CALDERA extends DBGWAS by performing one test for each closed connected subgraph of the compacted De Bruijn graph built over a set of bacterial genomes. This allows to test the association between a phenotype and the presence of a causal gene which has several variants. CALDERA exploits Tarone's concept of testability to avoid testing sequences which cannot possibly be associated with the phenotype.

URL: https://github.com/HectorRDB/Caldera_Recomb

Contact: Laurent Jacob

7.1.5 Capybara

Name: equivalence CLASS enumeration of coPhylogenY event-BAsed ReconciliAtions

Keywords: Bioinformatics, Evolution

Functional Description: Phylogenetic tree reconciliation is the method of choice in analysing host-symbiont systems. Despite the many reconciliation tools that have been proposed in the literature, two main issues remain unresolved: listing suboptimal solutions (*i.e.*, whose score is “close” to the optimal ones), and listing only solutions that are biologically different “enough”. The first issue arises because the optimal solutions are not always the ones biologically most significant, providing many suboptimal solutions as alternatives for the optimal ones is thus very useful. The second one is related to the difficulty to analyse an often huge number of optimal solutions. Capybara addresses both of these problems in an efficient way. Furthermore, it includes a tool for visualising the solutions that significantly helps the user in the process of analysing the results.

URL: <https://github.com/Helio-Wang/Capybara-app>

Publication: [hal-02917341](https://hal.archives-ouvertes.fr/hal-02917341)

Contact: Yishu Wang

Participants: Yishu Wang, Arnaud Mary, Marie-France Sagot, Blerina Sinimeri

7.1.6 Cassis

Keywords: Bioinformatics, Genomics

Functional Description: Implements methods for the precise detection of genomic rearrangement break-points.

Contact: Marie-France Sagot

Participants: Christian Baudet, Christian Gautier, Claire Lemaitre, Eric Tannier, Marie-France Sagot

7.1.7 Coala

Name: CO-evolution Assessment by a Likelihood-free Approach

Keywords: Evolution, Phylogenomics

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.

URL: <http://team.inria.fr/erable/en/software/coala/>

Publication: [hal-01092972](https://hal.archives-ouvertes.fr/hal-01092972)

Contact: Blerina Sinimeri

Participants: Beatrice Donati, Blerina Sinimeri, Catherine Matias, Christian Baudet, Christian Gautier, Marie-France Sagot, Pierluigi Crescenzi

7.1.8 Cycads

Keyword: Metabolism

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

Contact: Hubert Charles

Participants: Augusto Vellozo, Hubert Charles, Marie-France Sagot, Stefano Colella

7.1.9 DBGWAS

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.

URL: <https://gitlab.com/leoisl/dbgwas>

Contact: Laurent Jacob

7.1.10 Eucalypt

Keywords: Evolution, Phylogenomics

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.

URL: <http://team.inria.fr/erable/en/software/eucalypt/>

Publication: [hal-01092977](https://hal.archives-ouvertes.fr/hal-01092977)

Contact: Blerina Sinimeri

Participants: Beatrice Donati, Blerina Sinimeri, Christian Baudet, Marie-France Sagot, Pierluigi Crescenzi

7.1.11 Fast-SG

Keyword: Genome assembly

Functional Description: FAST-SG enables the optimal hybrid assembly of large genomes by combining short and long read technologies.

URL: <https://github.com/adigenova/fast-sg>

Publication: [hal-01842462](https://hal.archives-ouvertes.fr/hal-01842462)

Contact: Alex Di Genova

Participants: Alex Di Genova, Marie-France Sagot, Alejandro Maass, Gonzalo Ruz Heredia

7.1.12 Gobbolino-Touché

Keywords: Graph algorithmics, Metabolism

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.

URL: <https://team.inria.fr/erable/en/software/gobbolino/>

Contact: Marie-France Sagot

Participants: Etienne Birmele, Fabien Jourdan, Ludovic Cottret, Marie-France Sagot, Paulo Vieira Milreu, Pierluigi Crescenzi, Vicente Acuña, Vincent Lacroix

7.1.13 HgLib

Name: 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 Contributions: Initial version

URL: <https://gitlab.inria.fr/kirikomics/hglib>

Contact: Arnaud Mary

Participants: Martin Wannagat, David Parsons, Arnaud Mary, Irene Ziska

7.1.14 KissDE

Keywords: Graph algorithmics, Transcriptomics, Genomics

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

URL: <http://kisssplice.prabi.fr/tools/kissDE/>

Contact: Vincent Lacroix

Participants: Camille Marchet, Aurélie Siberchicot, Audric Cologne, Clara Benoît-Pilven, Janice Kielbassa, Lilia Brinza, Vincent Lacroix

7.1.15 KisSplice

Keywords: RNA-seq, De Bruijn graphs

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 Contributions: 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+.

KISPLICE 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 KISPLICE. 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.

URL: <http://kissplice.prabi.fr/>

Publication: [hal-00784407v1](#)

Contact: Vincent Lacroix

Participants: Alice Julien-Laferriere, Pierre Peterlongo, Rayan Chikhi, Vincent Miele, François Gindraud, Leandro Ishi Soares De Lima, Camille Marchet, Gustavo Akio Tominaga Sacomoto, Marie-France Sagot, Vincent Lacroix

7.1.16 KisSplice2RefGenome

Keywords: Bioinformatics, NGS, Transcriptomics

Functional Description: KISPLICE identifies variations in RNA-seq data, without a reference genome. In many applications however, a reference genome is available. KISPLICE2REFGENOME enables to facilitate the interpretation of the results of KISPLICE after mapping them to a reference genome.

URL: <http://kissplice.prabi.fr/tools/kiss2refgenome/>

Publication: [hal-02305628](#)

Contact: Vincent Lacroix

Participants: Audric Cologne, Camille Marchet, Camille Sessegolo, Alice Julien-Laferriere, Vincent Lacroix

7.1.17 KisSplice2RefTranscriptome

Keywords: Bioinformatics, NGS, Transcriptomics

Functional Description: KISPLICE2REFTRANSCRIPTOME enables to combine the output of KISPLICE 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.

URL: <http://kissplice.prabi.fr/tools/kiss2rt/>

Publication: [hal-02305628](#)

Contact: Vincent Lacroix

Participants: Helene Lopez Maestre, Mathilde Boutigny, Vincent Lacroix

7.1.18 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.

URL: <https://metexplore.toulouse.inra.fr/index.html/>

Contact: Fabien Jourdan

Participants: Fabien Jourdan, Hubert Charles, Ludovic Cottret, Marie-France Sagot

7.1.19 MetHg

Keywords: Hypergraphs, Metabolic networks, Rust

Functional Description: Rust directed hypergraph library, with a focus on modelling metabolic networks. Data is stored in dense arrays with layouts similar to Apache Columnar for efficiency. Supports both uses as a model database for generating linear programming problems, or combinatorial graph searches. This can be compiled to Wasm, and is being used for the rewrite as a client-side only app of a web visualisation tool previously developed in the team and called Dinghy.

URL: <https://gitlab.inria.fr/erable/methg/>

Contact: François Gindraud

7.1.20 Mirinho

Keywords: Bioinformatics, Computational biology, Genomics, Structural Biology

Functional Description: Predicts, at a genome-wide scale, microRNA candidates.

URL: <http://team.inria.fr/erable/en/software/mirinho/>

Publication: [hal-01166487](#)

Contact: Marie-France Sagot

Participants: Christian Gautier, Christine Gaspin, Cyril Fournier, Marie-France Sagot, Susan Higashi

7.1.21 Momo

Name: 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.

URL: <http://team.inria.fr/erable/en/software/momo/>

Publication: [hal-02490353](#)

Contact: Marie-France Sagot

Participants: Ricardo Luiz De Andrade Abrantes, Nuno Mira, Susana Vinga, Marie-France Sagot

7.1.22 Moomin

Name: 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>

URL: <https://github.com/htpusa/moomin>

Publication: [hal-02284835v1](#)

Contact: Marie-France Sagot

Participants: Henri Taneli Pusa, Mariana Ferrarini, Ricardo Luiz De Andrade Abrantes, Arnaud Mary, Alberto Marchetti-Spaccamela, Leendert Stougie, Marie-France Sagot

7.1.23 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.

URL: <https://team.inria.fr/erable/en/software/multipus/>

Publication: [hal-01394119](#)

Contact: Marie-France Sagot

Participants: Alberto Marchetti-Spaccamela, Alice Julien-Laferriere, Arnaud Mary, Delphine Parrot, Laurent Bulteau, Leendert Stougie, Marie-France Sagot, Susana Vinga

7.1.24 paSAMcs

Keyword: Metabolism

Functional Description: Computation of Minimal Cut Sets using Answer Set Programming (ASP), and more precisely [aspefm](#).

URL: <https://github.com/maxm4/paSAMcs>

Contact: Sabine Peres

Participants: Sabine Peres, Maxime Mahout

7.1.25 Pitufolandia

Keywords: Bioinformatics, Graph algorithmics, Systems Biology

Functional Description: The algorithms in PITUFOLANDIA (PITUFO / PITUFINA / PAPAITUFO) 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.

URL: <https://team.inria.fr/erable/en/software/pitufo/>

Contact: Marie-France Sagot

Participants: Vicente Acuña, Paulo Vieira Milreu, Alberto Marchetti-Spaccamela, Leendert Stougie, Martin Wannagat, Marie-France Sagot

7.1.26 Sasita

Keywords: Bioinformatics, Graph algorithmics, Systems Biology

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

URL: <https://team.inria.fr/erable/en/software/sasita/>

Publication: hal-01368653

Contact: Marie-France Sagot

Participants: Vicente Acuña, Ricardo Luiz De Andrade Abrantes, Paulo Vieira Milreu, Alberto Marchetti-Spaccamela, Leendert Stougie, Martin Wannagat, Marie-France Sagot

7.1.27 Smile

Keywords: Bioinformatics, Genomic sequence

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

URL: <https://gitlab.inria.fr/nhomberg/smile>

Publication: tel-04366914

Contact: Marie-France Sagot

Participants: Marie-France Sagot, Nicolas Homberg

7.1.28 Totoro

Name: Transient respOnse to meTabOlic pertuRbation inferred at the whole netwOrk level

Keywords: Bioinformatics, Graph algorithmics, Systems Biology

Functional Description: TOTORO is a constraint-based approach that integrates internal metabolite concentrations that were measured before and after a perturbation into genome-scale metabolic reconstructions. It predicts reactions that were active during the transient state that occurred after the perturbation. The method is solely based on metabolomic data.

URL: <https://gitlab.inria.fr/erable/totoro>

Publication: hal-03584295

Contact: Irene Ziska

Participants: Irene Ziska, Arnaud Mary, Marie-France Sagot

7.1.29 Wengan

Name: 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.

URL: <https://github.com/adigenova/wengan>

Publication: hal-03065904

Contact: Marie-France Sagot

Participants: Alex Di Genova, Marie-France Sagot

8 New results

8.1 General comments

We present in this section the main results obtained in 2025.

As in previous years, 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.

We would like also to call attention to two main facts.

The first one was already pointed out in our reports for the previous years. It concerns the fact that we choose in general not to detail the results on the more theoretical aspects of computer science when these are initially addressed in contexts not directly related to computational biology even though they could be relevant for different problems in the life sciences areas of research, or could become more specifically so in a near future. Examples of these for 2025 are [2, 3, 5, 6, 7, 15]. We also chose not to detail some of the results related to text algorithms even though these may, or have already more direct applications in biology [14, 8].

This year, as was the case in 2024, there is an exception in the sense that we obtained results – theoretical – that have already been shown to be potentially important in different aspects of computational biology and that are of the team’s interest. Because of this, we chose to provide more details on the paper in the first section below.

The second fact we want to call attention to is that in 2025, as was already the case for 2024 but things are now accelerating, represents a transition period for the ERABLE team. Indeed, due to the fact that various of the more senior members retired already (namely, Alberto Marchetti-Spaccamela since November, Leen Stougie since January, and the team’s leader Marie-France Sagot since October - the new leader is since October 2025 Sabine Peres) or will retire soon (Alain Viari), there are and will continue to have changes in the overall composition of the team and in the scientific topics it continues to address in the future.

8.2 General theoretical result

One main general theoretical result was obtained in 2025. This addressed reconfiguration problems, an area related to enumeration problems which is one of the topics at the heart of ERABLE’s scientific interests. The objective is to study how the states of a system can evolve by small modifications. More precisely, we ask under what conditions there exists a sequence of local transformations that allows one to move from one solution of an optimisation problem to another solution, while maintaining at each step the property of being a valid solution to the problem.

8.2.1 The Tape Reconfiguration Problem and its consequences for Dominating Set Reconfiguration

Participants: Arnaud Mary.

A dominating set of a graph $G = (V, E)$ is a set of vertices $D \subseteq V$ whose closed neighbourhood is V , i.e., $N[D] = V$. We view a dominating set as a collection of tokens placed on the vertices of D . In the token sliding variant of the Dominating Set Reconfiguration problem (TS-DSR), we seek to transform a source dominating set into a target dominating set in G by sliding tokens along edges, and while maintaining a dominating set all along the transformation. TS-DSR is known to be PSPACE-complete even restricted to graphs of pathwidth w , for some non-explicit constant w and to be XL-complete parameterized by the size k of the solution. The first contribution of the paper [16] consisted in using a novel approach to provide the first explicit constant for which the TS-DSR problem is PSPACE-complete, a question that was left open in the literature. From a parameterized complexity perspective, the token jumping variant of DSR, i.e., where tokens can jump to arbitrary vertices, is known to be FPT when parameterized by the size of the dominating sets on nowhere dense classes of graphs. However, in contrast, no non-trivial result was known about TS-DSR. We proved that DSR is actually much harder in the sliding model since it is XL-complete when restricted to bounded pathwidth graphs and even when parameterized by k plus the feedback vertex set number of the graph. This gives, for the first time, a difference of behaviour between the complexity under token sliding and token jumping for some problems on graphs of bounded treewidth. All our results were obtained using a brand new method, based on the hardness of the so-called Tape Reconfiguration problem, a problem we believe to be of independent interest.

8.3 Axis 1: (Pan)Genomics and transcriptomics in general

We start by presenting the results obtained within this axis that are related more to sequence analysis and alignment, although they are important also for pangenome analysis, in particular the last-but-one result presented below.

We then end this section by presenting a result obtained in 2025 in the area of transcriptomics.

8.3.1 Missing value replacement in strings and applications

Participants: Alberto Marchetti-Spaccamela, Solon Pissis, Leen Stougie.

Missing values arise routinely in real-world sequential (string) datasets due to: (1) imprecise data measurements; (2) flexible sequence modelling, such as binding profiles of molecular sequences; or (3) the existence of confidential information in a dataset which has been deleted deliberately for privacy protection. In order to analyse such datasets, it is often important to replace each missing value, with one or more valid letters, in an efficient and effective way. In the paper [1], we formalised this task as a combinatorial optimisation problem: the set of constraints includes the context of the missing value (i.e., its vicinity) as well as a finite set of user-defined forbidden patterns, modelling, for instance, implausible or confidential patterns; and the objective function seeks to minimise the number of new letters we introduce. Algorithmically, our problem translates to finding shortest paths in special graphs that contain forbidden edges representing the forbidden patterns. Our work made the following contributions: (1) we designed a linear-time algorithm to solve this problem for strings over constant-sized alphabets; (2) we showed how our algorithm can be effortlessly applied to fully sanitize a private string in the presence of a set of fixed-length forbidden patterns; (3) we proposed a methodology for sanitizing and clustering a collection of private strings that utilizes our algorithm and an effective and efficiently computable distance measure; and (4) we presented extensive experimental results showing that our methodology can efficiently sanitize a collection of private strings while preserving clustering quality, outperforming the state of the art and baselines. To arrive at our theoretical results, we employed techniques from formal languages and combinatorial pattern matching.

8.3.2 McDag: Indexing maximal common subsequences for k strings

Participants: Roberto Grossi.

Maximal Common Subsequences (MCSs), i.e., inclusion-maximal sequences of non-contiguous symbols common to two or more strings, have only recently received attention in the area of sequence comparison, despite being a basic notion and a natural generalisation of more common tools like Longest Common Substrings/Subsequences. In the paper [4], we simplified and engineered recent advancements as concerns MCSs into a practical tool that can index MCSs of real genomic data, and showed that its definition can be generalised to multiple strings. We demonstrated that our tool can index pairs of sequences exceeding 10,000 base pairs within minutes, utilising only 4-7% more than the minimum required nodes. For three or more sequences, we observed experimentally that the minimum index may exhibit a significant increase in the number of nodes.

8.3.3 Dynamic programming alignments with skips

Participants: Nadia Pisanti.

The outcome of a Multiple Sequence Alignment (MSA) can be compactly represented by means of Elastic-Degenerate Strings (ED-strings) by collapsing conserved fragments into standard linear strings, while representing gaps and variants as sets of alternative strings. These alternative variants can differ in size and can possibly include the empty string. In 2022, Lee *et al.* introduced Partial Order Alignment (POA) to enable the alignment of a string against a graph-like structure derived from an MSA. However, the POA edit transcript (the sequence of edit operations that describe the alignment) does not reflect the possible elasticity of the MSA (such as different gaps sizes in the aligned string), leaving room for a possible misalignment and its propagation in progressive MSA strategies. In the paper [11], we proposed a dynamic programming based method that optimally aligns a string to an ED-string, the latter compactly representing an MSA, overcoming the ambiguity in the POA edit transcript while maintaining its time and space complexity. Moreover, since pangenomes can also be represented using ED-strings, our algorithm paves the way to a new class of sequence to graph alignment methods capable of taking into account possible gaps in the pangenome representation, thus offering a richer and more flexible model for pangenomic analysis.

8.3.4 Models and algorithms for managing repeats in the *de novo* assembly of transcriptomes

Participants: Sasha Darmon, Vincent Lacroix, Arnaud Mary.

With the advent of short-read RNA-seq technologies, transcriptome assembly has become both more accessible and also more complicated. This problem, known as *de novo* transcriptome assembly, remains the only option for transcriptomic exploration in most non-model organisms, where no reference genome is available or where existing references are too divergent. Inexact repeats in the transcriptome generate complex regions in the assembly graph that are difficult to resolve. Among the most problematic repeats are transposable elements (TEs)—mobile sequences capable of copying and inserting themselves throughout the genome. Their high copy number and sequence similarity introduce ambiguities in read mapping and transcript structure inference. These issues are especially severe in *de novo* assemblies where no reference exists to anchor and disambiguate repetitive reads, leading to tangled graph structures and misassemblies. We specifically utilise De Bruijn graphs, an efficient data structure where each transcript corresponds to a path within the graph. Our research focuses on characterising complex regions that contain families of repeats and replacing them with consensus nodes. The objective of the novel method we developed [18] is to operate *de novo*, without relying on genomic references or repeat consensus sequences. This *de novo*

approach aims to avoid the ambiguous mapping of TEs, utilising widely available short-read sequences and making it applicable to non-model species.

8.4 Axis 2: Metabolism and (post)transcriptional regulation

As in 2024, the work of ERABLE in 2025 concentrated more on metabolism. The team is however still interested in (post)transcriptional regulation, notably small RNAs, and should notably pick up again a collaboration with an ex-PhD student of ERABLE, namely Carol Moraga Quinteros who has now a permanent position as Associate Professor at the University of O'Higgins in Chile.

As concerns the work done on metabolism, this involved notably the continuation of the work of two PhD students of the team, namely Emma Crisci and Camille Siharath. These are briefly described below with manuscripts in preparation. Both will also be defending their PhD in 2026. As may be seen, some of these works involve also health-related questions. We nevertheless decided to present them in this section, and to just mention it in Axis 4 below.

8.4.1 Growth Balanced Analysis (GBA) and Elementary Growth Modes (EGMs)

Participants: Emma Crisci, Sabine Peres.

Elementary Flux Modes (EFM) allow the description of the minimal sets of reactions in a metabolic network under steady-state conditions, representing unique and feasible pathways. They fully characterise the solution space but a combinatorial explosion prevents their calculation when the network is large. Furthermore, it is not necessary to calculate all EFMs as many are not biologically relevant. In a paper published in 2024, we had introduced the software ASP_{EFM} which combines the use of Answer Set Programming and Linear Programming, and further proposes to integrate different types of constraints in the computation of EFMs such as equilibrium constants, Boolean regulatory rules, growth yields and growth medium. In 2025, we chose instead to use the CLINGO_{LPx} solver, which allows us to save a considerable amount of time in the enumeration of EFMs compared to our previous methods. We coded a new thermodynamic extension to add to CLINGO_{LPx}. We also started to collaborate with two other researchers, namely Wolfram Liebermeister from INRAe in Paris, and Noor Elad from the Weizmann Institute of Science, Israel, to develop a new extension that allows to add a new biological constraint on the notion of maximum enzymatic cost. We also took a close interest in Growth Balanced Analysis (GBA), and in particular Elementary Growth Modes (EGMs) which are equivalence classes of Elementary Growth States. This new modelling method allows the notion of growth to be incorporated directly into the networks without using a fictitious biomass equation. The major advantage of this type of method is that it gives access to metabolite concentrations, which is not the case for the methods we used before. We are currently implementing a method for doing Growth Balance Analysis, more specifically for enumerating the EGMs of a model. This presents many algorithmic challenges.

8.4.2 Modelling energy metabolism dysregulations in neuromuscular diseases – A case study of calpainopathy

Participants: Sabine Peres, Camille Siharath.

As a reminder of last year Inria's Annual Report, the objective of Camille Siharath's PhD is to develop a metabolic model of skeletal muscle tissue to better understand the reorganisations associated with certain neuromuscular pathologies, and to identify potential therapeutic targets. In the second year, Camille's focus was on exploring new methods for introducing kinetic and transcriptomic constraints in the model. As concerns the integration of kinetic constraints, different approaches were considered. Among those enabling to overcome the limitations of FBA, some methods, such as KINETICEFM (developed in the team), are based on elementary flux modes (EFMs) that correspond to minimal subsets of reactions capable of functioning

autonomously in a steady state. Their use opens up the possibility of integrating kinetic constraints in a more refined manner, but at the cost of significant computational complexity. Another approach is to take into account the physical limitation imposed by molecular crowding to try to better reflect the actual enzyme capacities. As concerns now transcriptomic constraints, the idea was to use data to more directly link the gene expression in pathological contexts to the simulated metabolic capacities. Different approaches to integrating such data, namely iMAT, GIMME, MADE, TIGER-MADE, and RIPTiDE, were compared. While conventional methods rely on binary activation of fluxes, we were able to establish that RIPTiDE stands out for its quantitative approach and parsimonious sampling, which facilitates the identification of key reactions. This method is currently being implemented and should enable more detailed links to be established between pathological transcriptomic profiles and simulated metabolic dysregulations. Both developments aim to make the model more predictive and flexible, paving the way for its application to other neuromuscular diseases and personalised medicine approaches.

8.4.3 Logic programming-based Minimal Cut Sets to identify therapeutic targets in oncology

Participants: Sabine Peres.

Within the *Mitotic* project, we developed, with Jérémie Muller-Prokob (a former Master's student, currently a PhD student at the University of Düsseldorf), an innovative methodological approach for identifying therapeutic targets in oncology, based on genome-wide metabolic modelling and logic programming. This work is currently being prepared for publication. It relies on the use of metabolic models of cancer and healthy cells to identify minimal sets of metabolic perturbations capable of inhibiting tumor viability while preserving the essential functions of non-pathological cells. The method combines metabolic flux analysis, the enumeration of minimal cut sets, and the integration of biological constraints derived from transcriptomic data and existing therapeutic knowledge. Extensive model curation and validation work was carried out to ensure the robustness and transferability of the approach. The results obtained demonstrate the ability of the proposed framework to find known targets, suggest new potential targets and improve the specificity of predictions, while maintaining a computational performance compatible with large-scale exploration.

8.4.4 Systems biology in identifying metabolic reprogramming signatures resulting from the infection of human macrophages and fibroblasts by *Leishmania*

Participants: Marie-France Sagot.

As mentioned in the section “Partnerships and cooperations”, we have since 2024 been working with the PhD student of a researcher from Fiocruz, in Salvador, Bahia, who had some 14 years ago visited the team himself as a PhD student, more precisely as a “sandwich” PhD (“sandwich” PhDs refer to Brazilian PhD students who obtain a funding to spend one year working with a researcher outside Brazil). The “sandwich” student in the case was Pablo Ivan Pereira Ramos, who was doing his PhD with a researcher, Marisa Nicolas, at the LNCC (“Laboratório Nacional de Ciência Computacional”), Brazil, in the group of Ana Tereza Ribeiro de Vasconcelos with whom ERABLE has had a long-term collaboration, including via a CNRS LIA (“Laboratoire International Associé”) and also a Capes-Cofecub project. After his PhD, Pablo I. P. Ramos got a position at Fiocruz where he is now a senior researcher. This time, he wanted to send one of his current PhD students, Lucas Gentil Azevedo, to spend 10 months working in ERABLE. Lucas G. Azevedo obtained for this in 2024 a TerrEE scholarship from Campus France and arrived in Lyon in September 2024. The main topic we have been working on aims to increase our knowledge about the molecular mechanisms of a *Leishmania* infection during the amastigote life stages present in the human host, particularly to identify the metabolic reprogramming signatures of the host resulting from infection in immune system cells. From these results, it is hoped to pave the way for the identification of new biomarkers and effective drug targets in the fight against leishmaniasis, as well as to contribute to a deeper understanding of pathogen-host interactions in leishmaniasis, as well as, on a longer term, of other medically important pathogens. Two papers are in

preparation related to this work. This work is been conducted with also Mariana G. Ferrarini, an ex-PhD student and postdoc in ERABLE who is now group leader at the Max Planck Institute for Chemical Ecology in Jena, Germany, as well as with Ariel Silber, a Professor at the Department of Parasitology of the Institute of Biomedical Sciences at the University of São Paulo, Brazil, who is an expert both of diseases related with parasites, in his case Trypanomomas, and of metabolism. On the other hand, Lucas and Pablo, together with Mariana and Ariel, are also participating in a work that we are conducting with Renata Wassermann, who is Professor at the University of São Paulo like Ariel, but in her case in the Department of Computer Science of the Institute of Mathematics and Statistics. This is briefly described in the next section below.

8.4.5 Ontologies and Genome-scale Metabolic Models (GEMs)

Participants: Marie-France Sagot.

It is also in 2024 that we started a collaboration with Renata Wassermann, who had done her Bachelor's degree in Computer Science at the University of São Paulo (USP) at the same time as M.-F. Sagot, the two maintaining contact since then. Renata had then done her PhD at CWI in the Netherlands, in the areas of logic, knowledge representation and model revision, before returning to Brazil where she got a position at USP as Associate Professor in 2005. The collaboration we established in 2024 involved also a PhD student of Renata, Nahim Alves de Souza, who has visited ERABLE twice in 2025. The main objectives of this collaboration is to apply ontology concepts and logic to (1) provide a more expressive representation for metabolic networks by adding semantic information, (2) accelerate the process of Genome-scale Metabolic Model (GEM) reconstructions, (3) help to find problems/inconsistencies in the networks reconstructed by using reasoning and logical inferences, and (4) compare two reconstructed networks in order to find commonalities and differences. As in the previous case, this work involves also Mariana G. Ferrarini and Ariel Silber, as well as the PhD Ariel and M.-F. Sagot have in common, namely Gabriela T. Montanaro, as well as Lucas G. Azevedo and Pablo I. P. Ramos. Two papers are in preparation related to this work.

8.4.6 Production of polyhydroxyalkanoates by *Halomonas* sp. HG01 using various carbon sources: metabolic and genomic analysis

Participants: Marie-France Sagot.

Halomonas sp. HG01, a moderate halophilic bacterium isolated from a northern Peru salt mine, is promising as a polyhydroxyalkanoate (PHA) producer. In a collaboration involving biologists from two universities in the state of São Paulo as well as ex-members of ERABLE, namely Mariana Ferrarini (now at Max Planck for Chemical Biology in Germany and Alex di Genova at O'Higgins University in Chile, experimental data and genome analysis were used to evaluate its capabilities and presented in a paper [9] accepted at the end of 2025. Shaken flask experiments revealed that HG01 can accumulate 70-86 wt.% poly(3-hydroxybutyrate) [P(3HB)] from various carbon sources, including glucose, sucrose, and fructose. In a fed-batch bioreactor, it achieved a cell dry weight (CDW) of 12.2 g/L with 63% P(3HB) content after 72 hours. The PHA synthase enzyme exhibited substrate specificity for C4 to C5 compounds. HG01 strain also produced poly(3hydroxybutyrate-co-3-hydroxyvalerate) [P(3HB-co-3HV)] from propionic and valeric acids, with a maximum 3HV content of 25.53 mol% monomers into the polymer (69.40 %wt) when valeric acid was used. The complete bacterial 3.66 Mbp genome sequence revealed metabolic pathways for carbohydrate and fatty acid catabolism, PHA biosynthesis, and stress tolerance factor. This genetic information enhanced our understanding of PHA synthesis and supports the development of metabolic engineering strategies, positioning *Halomonas* sp. HG01 as a promising candidate for biotechnological applications.

8.5 Axis 3: (Co)Evolution

The work of ERABLE in 2025 on (co)evolution/(co)phylogenetics was strongly reduced. The topic still interests ERABLE and will be picked up again, involving notably Blerina Sinimeri and also Arnaud Mary

and Susana Vinga, one of ERABLE's external members, as well as Marie-France Sagot. There is however one preliminary result related to evolution that was presented in a poster at Alphy/AIEM in 2025 and involved Pierre Gérenton and Vincent Lacroix. This is briefly mentioned below.

8.5.1 Search for photosynthesis-related protein sites through chloroplast phylogenomics

Participants: Pierre Gérenton, Vincent Lacroix.

Since the late 90's, phylogeneticists have been interested in uncovering sites associated to phenotypes, across species. Several methods were developed to study them, including dn/ds methods or profile methods. Profile methods inform about the direction of the selection, and one of them, namely Pelican (Duchemin, 2023), allows us to detect proteic sites related to a phenotype of interest at the genome scale. In the context of his PhD, Pierre Gérenton is trying to uncover genotype-phenotype associations in plants (Viridiplantae), a difficult kingdom to analyse because of multiple genome-wide duplications and multicopy gene families. As a first step, it was decided to look for chloroplastic proteic sites related to the photosynthetic pathway (C3/C4/CAM). The preliminary analysis done and the conclusion about the findings reached was presented in the poster [19].

8.6 Axis 4: Health in general

As indicated in Axis 2 above, some of the work on metabolism developed in 2025 concerned health-related questions. This includes notably the PhD work of Camille Siharath with Sabine Peres and Olivier Biondi, the work of Sabine with Jérémie Prokob, as well as the work of Marie-France Sagot with some of her Brazilian collaborators.

Besides this, there are other ones concerning cancer, and more precisely the work of Alain Viari. We highlight the results of the main one below and just cite here the two other ones [10, 12].

8.6.1 Broad versus limited gene panels to guide treatment in patients with advanced solid tumors: a randomized controlled trial

Participants: Alain Viari.

Large genomic programs have contributed to improving drug development in cancer. To assess the potential benefit of using larger gene panels to guide molecular-based treatments, we conducted a multicenter randomized trial in patients with advanced and/or metastatic solid cancer. Molecular alterations were determined using either a panel of 324 cancer-related genes (Foundation OneCDX (F1CDX)) or a limited panel of 87 single-nucleotide/indel genes and genome-wide copy number variations (CTL) and reviewed by a molecular tumor board to identify molecular-based recommended therapies (MBRTs). Using paired data from both panels for each patient, the primary endpoint was the proportion of patients with an MBRT identified. Main secondary endpoints included the number of patients with at least one actionable alteration leading to MBRT identification, the number of patients with and without MBRTs initiated, progression-free survival, best overall response, duration of response and safety. Among the 741 patients screened, 45.7% had quality-checked tumor samples. MBRTs were identified with F1CDX in 175 (51.6%) patients and with CTL in 125 (36.9%) patients, translating to a significant increase of 14.8 percentage points ($P < 0.001$) with the more comprehensive gene panel versus the more limited panel, meeting the primary endpoint. However, no differences in clinical outcomes were observed in these patients with advanced and/or metastatic cancer in need of treatment beyond standard genomic alterations. These findings which were presented in the paper [13] illustrate the potential for larger gene panels to increase the number of molecularly matched therapies. Larger studies will be needed in the future to assess the clinical benefit of expanded MBRTs.

9 Bilateral contracts and grants with industry

9.1 Bilateral Grants with Industry

Participants: Vincent Lacroix, Arnaud Mary, Sabine Peres.

The TYP'OMICS project is a scientific innovation initiative aimed at preserving and strengthening the unique characteristics of traditional cheeses with quality labels, using Reblochon AOP as a model. The project, coordinated by **Actalia** and that will last from 2025 to 2029 with a total grant of 499851 euros for the two partners, Actalia and ERABLE, is funded by the CASDAR ("Compte d'affectation spéciale développement agricole et rural") call that is part of the France 2030 regionalized initiative, under the "Collaborative Projects / Regionalized I-Démo" action in the Auvergne-Rhône-Alpes region. The mission of ERABLE involves deploying breakthrough technologies, such as genome-scale metabolic modelling and flux analysis methods, to characterise and select the most efficient microorganisms for the dairy industry. Beyond individual analysis, our work aims to construct meta-networks to simulate interactions within complex microbial communities and predict the production of aromatic compounds of interest.

10 Partnerships and cooperations

10.1 International research visitors

10.1.1 Visits of international scientists

Participants: Arnaud Mary, Sabine Peres, Marie-France Sagot.

Visit of Nahim Alves de Souza, PhD student of Prof. Renata Wassermann, as well as of Gabriela T. Montanaro, PhD student of Ariel Silber co-supervised by M.-F. Sagot, both from the University of São Paulo (USP), Brazil, from March 22 to April 19 to Erable in Lyon. This had various objectives. The main ones were related with the problem of the representation of genome-scale metabolic networks (also called GEMs) which is the main topic of Nahim's PhD and intervenes greatly also in the work of Gabriela. These objectives are: (1) provide a more expressive representation for metabolic networks by adding semantic information, (2) accelerate the process of GEMs reconstruction, (3) help to find problems/inconsistencies in the networks by using reasoning and logical inferences, (4) compare two network representations in order to find commonalities and differences. This visit was also the occasion to discuss with Gabriela on her PhD. Moreover, in the last week of both Nahim and Gabriela's visit, from April 12 to 19, we were able to also have the visit of Renata Wassermann.

Later in the year, Nahim re-visited the team from October 4 to 31, together with Lucas Gentil Azevedo from October 5 to 30. They were joined once again by Renata Wassermann from October 11 to 18, and then from October 19 to 26 by the PhD supervisor of Lucas.

From April 22 to May 17, we also had the visit of Bertrand Marchand, postdoc at University of Québec, Montréal, Canada, for a general discussion on possible collaborations.

Other international visits to the team

Lucas Gentil Azevedo

Status PhD

Institution of origin: Fiocruz-Bahia

Country: Brazil

Dates: September 1st, 2024 until June 31, 2025

Context of the visit: Lucas Gentil Azevedo's supervisor at Fiocruz-Brazil, Pablo Ivan Pereira Ramos, had been a "sandwich" PhD in the team in 2010-2011 and the visit of Lucas to France, funded by Campus-France, was an occasion to pick up the collaboration with Pablo on topics related to genomics and metabolism, and to the Leishmania parasite.

Mobility program/type of mobility: "Sandwich" PhD funded by Campus-France.

10.2 National initiatives

10.2.1 PEPR-ANR

Participants: Sabine Peres.

Title: Multi-size Hybrid Cell Models.

Coordinator: Alberto Tonda.

Type: Program PEPR Biomasse, Biotechnologie durables pour les produits chimiques et les carburants.

Duration: 2025-2029.

10.2.2 ITMO aviesan

Participants: Sabine Peres.

Title: Ressources Balances Analyses pour découvrir la vulnérabilité métabolique dans le cancer et identifier de nouvelles thérapies (MITOTIC).

Coordinator: Sabine Peres.

Type: Program "Mathématiques et Informatique" 2021 of ITMO Cancer aviesan INSERM.

Duration: 2021-2024, extended to 2025.

10.2.3 Others

Optimal

Participants: Leen Stougie.

Title: Optimization for and with Machine Learning.

Coordinator: Dick den Hertog.

Type: NWO ENW-Groot Program.

11 Dissemination

Participants: Emma Crisci, Sasha Darmon, Roberto Grossi, Giuseppe Italiano, Vincent Lacroix, Alberto Marchetti-Spaccamela, Arnaud Mary, Sabine Peres, Nadia Pisanti, Solon Pissis, Marie-France Sagot, Camille Siharath, Leen Stougie, Alain Viari.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

General chair, scientific chair

- Giuseppe Italiano is President of the Steering Committee of the International *Colloquium on Automata, Languages and Programming (ICALP)*.
- Roberto Grossi is member of the Steering Committee of *Symposium on Combinatorial Pattern Matching (CPM)*.
- Arnaud Mary is member of the Steering Committee of *Workshop on Enumeration Problems and Applications (WEPA)*.
- Sabine Peres is member of the Steering Committee of *Metabolic Pathway Analysis (MPA)*.
- Nadia Pisanti is member of the Steering Committee of *Workshop on Algorithms in BioInformatics (WABI)*.
- 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)*.

Member of the organizing committees

- Arnaud Mary was co-organiser of the First Edition of the Colloquium **Combinatorics and Life Sciences** that took place in Lyon from Sept 29 to Oct 2, 2025.
- Sabine Peres was co-organiser of the **23rd International Conference on Computational Methods in Systems Biology (CMSB)** which took place in Lyon from Sept 10 to 12, 2025 and whose Proceedings was published here [17]. She was also co-organiser of the **annual meeting of the GT BIOSS** which took place in Paris from 26 to 27 May 2025. She co-organised **virtual seminars every month**.
- Marie-France Sagot was co-organiser of the First Edition of the Colloquium **Combinatorics and Life Sciences** that took place in Lyon from Sept 29 to Oct 2, 2025; and co-organiser of the Fifth Edition of the **Workshop Metabolism and mathematical models: Two for a tango**, held virtually, Nov 20-21, 2025.

Chair of conference program committees

- Sabine Peres was co-Chair of the **23rd International Conference on Computational Methods in Systems Biology (CMSB)** which took place in Lyon from Sept 10 to 12, 2025 and whose Proceedings was published here [17].

Member of conference program committees

- Vincent Lacroix was a member of the Program Committee of *JOBIM* and *SeqBim*.
- Solon Pissis was a member of the Program Committee of *PSC* and *WABI*.

Member of the editorial boards

- Roberto Grossi is member of the Editorial Board of *Theory of Computing Systems (TOCS)*.
- Giuseppe Italiano is member of the Editorial Board of *ACM Transactions on Algorithms*, of *Algorithmica* and *Theoretical Computer Science*.
- Vincent Lacroix is recommender for *Peer Community in Genomics*.
- Nadia Pisanti is since 2017 member of the Editorial Board 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*, *Computer Science Review*, and *Lecture Notes in Bioinformatics*.
- Blerina Sinimeri is member of the Editorial Board of *Information Processing Letters* and of *Theoretical Computer Science*.
- 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*.

Reviewer - reviewing activities Members of ERABLE have reviewed papers for a number of journals including: *Theoretical Computer Science*, *Algorithmica*, *SIAM Journal on Computing*, *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*, *PLoS Computational Biology*.

11.1.2 Scientific expertise

- Giuseppe Italiano is since 2024 President of the European Association for Theoretical Computer Science (EATCS). He is since 2025 Deputy Rector for Artificial Intelligence and Digital Skills at LUISS University, Rome, besides having a number of other responsibilities at LUISS. He is also member of the Advisory Board of MADALGO - Center for MASSive Data ALGORITHmics, Aarhus, Denmark.
- Sabine Peres is since 2022 Head of the Master's degree in bioinformatics - University Lyon 1, member of the Advisory committee section 67-68 University Lyon 1, and internal member of the E2M2 doctoral school of the University of Lyon 1. She is also member of the coordination committee of DigitBioMed (Digital Sciences for Biology and Health) of the SFRI (Structuration de la Formation par la Recherche dans les Initiatives d'excellence).
- 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 was from 2014 to 2025 member of the Scientific Advisory Board of CWI, and from 2022 to 2025 member of the Scientific Advisory Board of the Dept. of Computational Biology at the Univ. of Lausanne, Switzerland. From 2022 to 2025 also, she was member of the Scientific Advisory Board of the MATOMIC project funded by the Novo Nordisk Foundation, Denmark, and coordinated by Prof. Daniel Merkle, Univ. of South Denmark.
- 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”.

11.1.3 Research administration

Marie-France Sagot was from 2021 until October 2025, member of the “Conseil Scientifique (COS)” and of the “COMité des Moyens Incitatifs (COMI)” for Inria Lyon.

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

11.2.1 Teaching

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. She teaches genetics 192 hours per year at the University and at the ENS-Lyon.
- Vincent Lacroix is responsible for the M1 master in bioinformatics 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 192 hours in 2025.
- Arnaud Mary is co-responsible for the M1 master in bioinformatics and of the following two courses : (M1: Advanced python programming for bioinformatics, M2: Advanced Algorithms for Bioinformatics). He taught 198 hours in 2025.
- Sabine Peres is co-responsible for the M2 master in bioinformatics. She is also responsible for four courses at the University, one at the Licence level and three at the Master level (L2: Mathematics life science, Python programming, M2 Bioinformatics: Modelling of metabolic networks; M2 Integrative Biology and Physiology: Modelling in Physiology, M2 Biodiversity, ecology and evolution: Python programming - simulation of population genetics).

Besides the above, all the French PhD students, namely Emma Crisci, Sasha Darmon, Pierre Gérenton, and Camille Siharath teach at the University on average 64 hours per academic year.

The ERABLE team regularly welcomes M1 and M2 interns from the bioinformatics Master. All French members of the ERABLE team are affiliated to the doctoral school E2M2, Ecology-Evolution-Microbiology-Modelling.

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).

11.2.2 Supervision

The following are the PhDs in progress in 2025:

- Emma Crisci, University Lyon & Inria (supervisor: Sabine Peres, together with Arnaud Mary)
- Sasha Darmon, University Lyon (supervisor: Vincent Lacroix, together with Arnaud Mary)
- Camille Siharath, University Lyon (supervisors: Sabine Peres and Olivier Biondi, University Evry Paris-Saclay)
- Michelle Sweering, CWI (co-supervisors: Solon Pissis and Leen Stougie)

11.2.3 Juries

The following are the PhD and HDR (Habilitation) juries to which members of ERABLE participated in 2025:

- Sabine Peres: Member of the Habilitation jury of Etienne Rajon, University Lyon 1, June 2025; Reviewer of the Habilitation of Clémence Frioux, University of Bordeaux, September 2025; President of the PhD jury of Arthur Lequertier, University Paris-Saclay, December 2025.
- Vincent Lacroix: Member of the Habilitation jury of Matthieu Boulesteix, University of Lyon 1, February 2025; Reviewer of the PhD of Ali Hamraoui, University of Paris Sciences and Letters, November 2025; Member of the PhD jury of Sylvère Bastien, University of Lyon 1, December 2025.
- Marie-France Sagot: Reviewer of the PhD of Jonas Coelho Kasman, University of Leipzig, Germany, April 2025; Reviewer of the PhD of Anupam Gautam, University of Tübingen, Germany, May 2025.
- Arnaud Mary: Member of the jury of Mostafa Gholami, University of Caen, November 2025.

11.3 Popularization

11.3.1 Specific official responsibilities in science outreach structures

Sasha Darmon is president of the science outreach association [Démésures](#).

11.3.2 Productions (articles, videos, podcasts, serious games, ...)

Sasha Darmon participated in the production, development and publication of a popular scientific comic book which may be recovered [here](#).

11.3.3 Participation in Live events

Emma Crisci participated to several popular science events in link with [Pint of Science](#).

Sasha Darmon participated to the 4th Meeting of the Inria Centre in Lyon where he shared his commitment to mediation, notably around the popular scientific comic book (BD) "Mission Z, in pursuit of intelligence", intended for middle school students and mentioned in the previous section.

Sasha also organised and animated five science activities during the "Fête de la Science" event in the LBBE in October 2025, and he participated in a meeting between teenagers and PhD students at the Annonay Theatre, reaching an audience of 800 teenagers.

Emma, Sasha and Camille Siharath participated to the GeekTouch event which took place in May 2025 where they co-hosted a science popularisation stand with the [Démésures](#) association.

11.3.4 Others science outreach relevant activities

Vincent Lacroix continues to participate in the development of a webservice called Alimempreinte (see [here](#)), which enables to calculate and compare the carbon footprint of different meals based on the list of ingredients. This tool is of interest for anyone who wishes to understand and reduce the carbon footprint of his/her diet. Alimempreinte has an average of 120 unique visitors per month.

Vincent is also part of the group that teaches the Climate & Transitions course for first-year undergraduates. This year, a humanities component has been added. The course is freely accessible [here](#).

12 Scientific production

12.1 Publications of the year

International journals

- [1] G. Bernardini, C. Liu, G. Loukides, A. Marchetti-Spaccamela, S. P. Pissis, L. Stougie and M. Sweering. 'Missing value replacement in strings and applications'. In: *Data Mining and Knowledge Discovery* 39.2 (22nd Jan. 2025), p. 12. DOI: [10.1007/s10618-024-01074-3](https://doi.org/10.1007/s10618-024-01074-3). URL: <https://inria.hal.science/hal-05435081> (cit. on p. 20).
- [2] V. Bonifaci and A. Marchetti-Spaccamela. 'Feasibility analysis of recurrent DAG tasks is PSPACE-hard'. In: *Theoretical Computer Science* 1030 (Mar. 2025), p. 115062. DOI: [10.1016/j.tcs.2024.115062](https://doi.org/10.1016/j.tcs.2024.115062). URL: <https://inria.hal.science/hal-05435040> (cit. on p. 19).
- [3] T. Bosman, M. van Ee, E. Ergen, C. Imreh, A. Marchetti-Spaccamela, M. Skutella and L. Stougie. 'Total Completion Time Scheduling Under Scenarios'. In: *Theory of Computing Systems* 69 (22nd Sept. 2025). DOI: [10.1007/s00224-025-10232-z](https://doi.org/10.1007/s00224-025-10232-z). URL: <https://inria.hal.science/hal-05435070> (cit. on p. 19).
- [4] G. Buzzega, A. Conte, R. Grossi and G. Punzi. 'McDag: indexing maximal common subsequences for k strings'. In: *Algorithms for Molecular Biology* 20.1 (19th Apr. 2025), p. 6. DOI: [10.1186/s13015-025-00271-z](https://doi.org/10.1186/s13015-025-00271-z). URL: <https://inria.hal.science/hal-05435089> (cit. on p. 21).
- [5] A. Conte, R. Grossi, Y. Kobayashi, K. Kurita, D. Rucci, T. Uno and K. Wasa. 'Enumerating Graphlets with Amortized Time Complexity Independent of Graph Size'. In: *Algorithmica* 87.9 (13th May 2025), pp. 1247–1273. DOI: [10.1007/s00453-025-01312-0](https://doi.org/10.1007/s00453-025-01312-0). URL: <https://inria.hal.science/hal-05435077> (cit. on p. 19).

- [6] A. Conte, R. Grossi, G. Loukides, N. Pisanti, S. P. Pissis and G. Punzi. ‘Fast Assessment of Eulerian Trails in Graphs with Applications’. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 20.1, Article 14 (8th Dec. 2025), pp. 1–33. doi: [10.1145/3771997](https://doi.org/10.1145/3771997). URL: <https://inria.hal.science/hal-05435063> (cit. on p. 19).
- [7] A. Friebe, A. Marchetti-Spaccamela, T. Cucinotta, A. V. Papadopoulos, T. Nolte and S. Baruah. ‘Resource Management for Stochastic Parallel Synchronous Tasks: Bandits to the Rescue’. In: *Real-Time Systems* 61.3-4 (14th July 2025), pp. 359–402. doi: [10.1007/s11241-025-09454-8](https://doi.org/10.1007/s11241-025-09454-8). URL: <https://inria.hal.science/hal-05435088> (cit. on p. 19).
- [8] E. Gabory, M. N. Mwaniki, N. Pisanti, S. P. Pissis, J. Radoszewski, M. Sweering and W. Zuba. ‘Elastic-degenerate string comparison’. In: *Information and Computation* 304 (May 2025), p. 105296. doi: [10.1016/j.ic.2025.105296](https://doi.org/10.1016/j.ic.2025.105296). URL: <https://inria.hal.science/hal-05434483> (cit. on p. 19).
- [9] C. W. Guzmán-Moreno, J. Cardinali-Rezende, A. Di Genova, M. Ferrarini, M.-F. Sagot, K. L. Rodrigues, L. F. Silva, M. K. Taciro and J. G. C. Gomez. ‘Production of polyhydroxyalkanoates by *Halomonas* sp. HG01, a halophilic bacterium from northern Peru, using various carbon sources: metabolic and genomic analysis’. In: *International Journal of Biological Macromolecules* 337 (Jan. 2026). doi: [10.1016/j.ijbiomac.2025.149399](https://doi.org/10.1016/j.ijbiomac.2025.149399). URL: <https://inria.hal.science/hal-05449695> (cit. on p. 24).
- [10] C. Morin, H. Paraqindes, F. N. van Long, C. Isaac, E. Thomas, D. Pedri, C. A. Pulido-Vicuna, A.-P. Morel, V. Marchand, Y. Motorin, M. Carrere, J. Auclair, V. Attignon, R. M. Pommier, E. Ruiz, F. Bourdelais, F. Catez, S. Durand, A. Ferrari, A. Viari, J.-C. Marine, A. Puisieux, J.-J. Diaz, C. Moyret-Lalle and V. Marcel. ‘Specific modulation of 28S_Um2402 rRNA 2’-ribose methylation as a novel epitranscriptomic marker of ZEB1-induced epithelial–mesenchymal transition in different mammary cell contexts’. In: *NAR Cancer* 7.1 (2025), zcaf001. doi: [10.1093/narcan/zcaf001](https://doi.org/10.1093/narcan/zcaf001). URL: <https://cnrs.hal.science/hal-05021643> (cit. on p. 25).
- [11] N. Moses Mwaniki and N. Pisanti. ‘Dynamic Programming Alignments With Skips’. In: *IEEE Access* 13 (2025), pp. 154267–154282. doi: [10.1109/ACCESS.2025.3597547](https://doi.org/10.1109/ACCESS.2025.3597547). URL: <https://inria.hal.science/hal-05435065> (cit. on p. 21).
- [12] C. Patte, R. Pommier, A. Ferrari, F. Fei-Lei Chung, M. Ouzounova, P. Moullé, M. Richaud, R. Khoueiry, M. Hervieu, S. Breusa, M. Allio, N. Rama, L. Gérard, V. Hervieu, G. Poncet, T. Fenouil, V. Cahais, A.-S. Sertier, A. Boland, D. Bacq-Daian, B. Ducarouge, J. Marie, J.-F. Deleuze, A. Viari, J.-Y. Scoazec, C. Roche, P. Mehlen, T. Walter and B. Gibert. ‘Comprehensive molecular portrait reveals genetic diversity and distinct molecular subtypes of small intestinal neuroendocrine tumors’. In: *Nature Communications* 16.1 (4th Mar. 2025), p. 2197. doi: [10.1038/s41467-025-57305-8](https://doi.org/10.1038/s41467-025-57305-8). URL: <https://hal.science/hal-04987274> (cit. on p. 25).
- [13] O. Trédan, D. Pouessel, N. Penel, S. Chabaud, C. Gomez-Roca, J.-P. Delord, D. Pannier, M. Brahmi, M. Fabbro, M.-E. Garcia, D. Larriau-Ciron, I. Ray-Coquard, M. Viala, A. Italiano, D. Tosi, P. Cassier, A. Dufresne, V. Attignon, S. Boyault, I. Treilleux, A. Viari, D. Pérol and J. Y. Blay. ‘Broad versus limited gene panels to guide treatment in patients with advanced solid tumors: a randomized controlled trial’. In: *Nature Medicine* 31.5 (7th Apr. 2025), pp. 1502–1508. doi: [10.1038/s41591-025-03613-x](https://doi.org/10.1038/s41591-025-03613-x). URL: <https://hal.science/hal-05112043> (cit. on p. 25).

International peer-reviewed conferences

- [14] G. Bernardini, H. Chen, A. Conte, R. Grossi, V. Guerrini, G. Loukides, N. Pisanti and S. P. Pissis. ‘Indexing Strings with Utilities’. In: *ICDE 2025 - 41st IEEE International Conference on Data Engineering*. Hong Kong, China: IEEE, 8th Apr. 2025, pp. 2782–2795. doi: [10.1109/ICDE65448.2025.00209](https://doi.org/10.1109/ICDE65448.2025.00209). URL: <https://inria.hal.science/hal-05435053> (cit. on p. 19).
- [15] L. Oettershagen, A. L. Konstantinidis and G. F. Italiano. ‘An Edge-Based Decomposition Framework for Temporal Networks’. In: *ACM WSDM 2025 - 18th ACM International Conference on Web Search and Data Mining*. Hannover, Germany: ACM, 10th Mar. 2025, pp. 735–743. doi: [10.1145/3701551.3703556](https://doi.org/10.1145/3701551.3703556). URL: <https://inria.hal.science/hal-05440039> (cit. on p. 19).

Edition (books, proceedings, special issue of a journal)

- [16] *The Tape Reconfiguration Problem and Its Consequences for Dominating Set Reconfiguration*. European Symposium on Algorithms. Varsovie, Poland: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. DOI: [10.4230/LIPICS.ESA.2025.29](https://doi.org/10.4230/LIPICS.ESA.2025.29). URL: <https://hal.science/hal-05369358> (cit. on p. 20).
- [17] *Proceedings of the 23rd International Conference on Computational Methods in Systems Biology, CMSB 2025*. CMSB 2025 - 23rd International Conference on Computational Methods in Systems Biology. Vol. Lecture Notes in Computer Science. Proceedings of the 23rd International Conference on Computational Methods in Systems Biology, CMSB 2025. LNCS-15959. Lyon, France: Springer Nature Switzerland; Springer, 2025. DOI: [10.1007/978-3-032-01436-8](https://doi.org/10.1007/978-3-032-01436-8). URL: <https://inria.hal.science/hal-05396349> (cit. on p. 28).

Other scientific publications

- [18] S. Darmon, A. Mary and V. Lacroix. ‘Models and Algorithms for Managing Repeats in the De Novo Assembly of Transcriptomes Models and Algorithms for Managing Repeats in the De Novo Assembly of Transcriptomes’. In: JOBIM 2025. Talence (Université de Bordeaux 1), France, 10th July 2025. URL: <https://hal.inrae.fr/hal-05214723> (cit. on p. 21).
- [19] P. Gérenton and P. Veber. ‘Search for photosynthesis-related protein sites through chloroplast phylogenomics’. In: ALPHY/AIEM 2025. Ed. by V. Lacroix and B. Boussau. Villeurbanne, France, 4th Feb. 2025. URL: <https://hal.science/hal-04958619> (cit. on p. 25).