

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**

2024

ACTIVITY REPORT

Project-Team

ERABLE

**European Research team in Algorithms
and Biology, formal and Experimental**

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

DOMAIN

Digital Health, Biology and Earth

THEME

Computational Biology

Inria

Contents

Project-Team ERABLE	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
3 Research program	3
3.1 Two main goals	3
3.2 Different research axes	4
4 Application domains	6
4.1 Biology and Health	6
5 Social and environmental responsibility	6
5.1 Footprint of research activities	6
5.2 Expected impact of research results	7
6 Highlights of the year	7
7 New software, platforms, open data	7
7.1 New software	7
7.1.1 AmoCoala	7
7.1.2 ASPefm	8
7.1.3 BrumiR	8
7.1.4 Caldera	9
7.1.5 Capybara	9
7.1.6 Cassis	9
7.1.7 Coala	9
7.1.8 Cycads	10
7.1.9 DBGWAS	10
7.1.10 Eucalypt	10
7.1.11 Fast-SG	10
7.1.12 Gobbolino-Touché	11
7.1.13 HgLib	11
7.1.14 KissDE	11
7.1.15 KisSplice	11
7.1.16 KisSplice2RefGenome	12
7.1.17 KisSplice2RefTranscriptome	12
7.1.18 MetExplore	12
7.1.19 MetHg	13
7.1.20 Mirinho	13
7.1.21 Momo	13
7.1.22 Moomin	14
7.1.23 MultiPus	14
7.1.24 paSAMcs	14
7.1.25 Pitufolandia	14
7.1.26 Sasita	15
7.1.27 Smile	15
7.1.28 Totoro	15
7.1.29 Wengan	15

8	New results	16
8.1	General comments	16
8.2	General theoretical result	16
8.2.1	Efficient enumeration of maximal split subgraphs and induced sub-cographs and related classes	16
8.2.2	Enumeration of minimal transversals of hypergraphs of bounded VC-dimension	17
8.3	Axis 1: (Pan)Genomics and transcriptomics in general	17
8.3.1	Connecting de Bruijn graphs	17
8.3.2	A unifying taxonomy of pattern matching in degenerate strings and founder graphs	18
8.3.3	Complexity and algorithms for Swap median and relation to other consensus problems	18
8.3.4	Identification and quantification of transposable element transcripts using Long-Read RNA-seq	18
8.4	Axis 2: Metabolism and (post)transcriptional regulation	19
8.4.1	Computing thermodynamically consistent Elementary Flux Modes with Answer Set Programming	19
8.4.2	Logic programming-based Minimal Cut Sets reveal consortium-level therapeutic targets for chronic wound infections	20
8.4.3	Modelling energy metabolism dysregulations in neuromuscular diseases – A case study of calpainopathy	20
8.4.4	Metabolic modelling links Warburg effect to collagen formation, angiogenesis and inflammation in the tumoral stroma	20
8.5	Axis 3: (Co)Evolution	21
8.5.1	Phylogenetic networks: Inferring such from multifurcating trees via cherry picking and machine learning	21
8.6	Axis 4: Health in general	21
8.6.1	Added value of whole-exome and RNA sequencing in advanced and refractory cancer patients with no molecular-based treatment recommendation based on a 90-gene panel	21
9	Partnerships and cooperations	22
9.1	International initiatives	22
9.1.1	Inria associate team not involved in an IIL or an international program	22
9.1.2	Participation in other International Programs	22
9.2	International research visitors	23
9.2.1	Visits of international scientists	23
9.2.2	Other international visits to the team	23
9.3	European initiatives	23
9.3.1	H2020 projects	23
9.4	National initiatives	24
9.4.1	PEPR-ANR	24
9.4.2	Others	24
10	Dissemination	25
10.1	Promoting scientific activities	25
10.1.1	Scientific events: organisation	25
10.1.2	Scientific expertise	26
10.1.3	Research administration	27
10.2	Teaching - Supervision - Juries	27
10.2.1	Teaching	27
10.2.2	Supervision	28
10.2.3	Juries	28
10.3	Popularization	28
10.3.1	Specific official responsibilities in science outreach structures	28
10.3.2	Productions (articles, videos, podcasts, serious games, ...)	28
10.3.3	Participation in Live events	28

10.3.4 Other science outreach relevant activities	29
11 Scientific production	29
11.1 Publications of the year	29

Project-Team ERABLE

Creation of the Project-Team: 2015 July 01

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

1 Team members, visitors, external collaborators

Research Scientists

- Marie-France Sagot [Team leader, INRIA, Senior Researcher]
- Solon Pissis [CWI, Researcher]
- Leen Stougie [CWI, Senior Researcher]
- Alain Viari [INRIA, Senior Researcher]

Faculty Members

- Roberto Grossi [UNIV PISE, Professor]
- Giuseppe Italiano [UNIV LUISS, Professor]
- Vincent Lacroix [UNIV LYON I, Associate Professor]
- Alberto Marchetti Spaccamela [SAPIENZA ROME, Professor]
- Arnaud Mary [UNIV LYON I, Associate Professor]
- Sabine Peres [UNIV LYON I, Professor]
- Nadia Pisanti [UNIV PISE, Associate Professor]
- Cristina Vieira [UNIV LYON I, Associate Professor]

PhD Students

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

Technical Staff

- François Gindraud [INRIA, Engineer]

Administrative Assistant

- Anouchka Ronceray [INRIA]

External Collaborators

- Laurent Jacob [CNRS]
- Susana Vinga [IST / ULISBOA]

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, although more recently the team has also been joined by members that come from biology including experimental.

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 as well as with Axis 5 on the methodological aspects, but since it gained more importance in the past few years, we decided to develop more these particular applications. Most of them started through collaborations with clinicians. Such applications are currently focused on 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, that 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 New software, 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 New software

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/>

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>

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

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 breakpoints.

URL: <http://pbil.univ-lyon1.fr/software/Cassis/>

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

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/>

Contact: Blerina Sinimeri

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

7.1.8 Cycads

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...).

URL: <http://www.cycadsys.org/>

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

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/>

Contact: Blerina Sinimeri

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

7.1.11 Fast-SG

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>

Contact: Alex Di Genova

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

7.1.12 Gobbolino-Touché

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

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://kissplice.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

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+.

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

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

Contact: Vincent Lacroix

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

7.1.16 KisSplice2RefGenome

Keywords: Bioinformatics, NGS, Transcriptomics

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

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

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

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

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/>

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/>

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>

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/>

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 / PAPAPITUFO) are designed to solve the minimal precursor set problem, which consists in finding all minimal sets of precursors (usually, nutrients) in a metabolic network that are able to produce a set of target metabolites.

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/>

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>

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>

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>

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 2024.

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 by such results.

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 2024 are [1, 4, 5].

This year as was the case in 2023, there is an exception to that in the sense that we obtained results – theoretical – that have already been shown to be 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 [3] and on the preprint [19].

The second fact we want to call attention to is that in 2024, as was already the case for 2023 but things are now accelerating, represents a transition period for the ERABLE team. Indeed, due to the fact that in the next couple of years, various of the more senior members will retire (namely, Alberto Marchetti-Spaccamela, Leen Stougie, Alain Viari, and the team's leader Marie-France Sagot), there will be many changes in the overall composition of the team and in the scientific topics it continues to address.

Finally, before going to the presentation of the different results that ERABLE obtained in 2024, we would like to mention a special publication related to the participation of M.-F. Sagot as co-chair in the organisation of the Special Session "Bioinformatics in France" that was one part of the joint [International Conferences ISMB/ECCB](#) which took place in Lyon in 2023. This paper [7] (see also [ISMB/ECCB 2023 organization benefited from the strengths of the French bioinformatics community](#)) that we were invited by the International Society for Computational Biology (ISCB) to write after the joint conferences was published in *Bioinformatics Advances* from Oxford University Press.

8.2 General theoretical result

Two main general theoretical results have been obtained in 2024. The first one is already published, while the second has been submitted only. Both address enumeration problems which is one of the topics at the heart of ERABLE's scientific interests. We start by briefly presenting the results in the accepted paper, and then those in the submitted one.

8.2.1 Efficient enumeration of maximal split subgraphs and induced sub-cographs and related classes

Participants: Arnaud Mary.

This paper [3] is focused on algorithms that take as input an arbitrary graph G , and enumerate as output all the (inclusion-wise) maximal "subgraphs" of G which fulfill a given property Π . Several different properties Π were studied, and the notion of subgraph under consideration (induced or not) varied from one result to another. More precisely, we presented efficient algorithms to list all maximal split subgraphs, sub-cographs and some subclasses of cographs of a given input graph. All the algorithms presented run in polynomial delay. Moreover, for split graphs they only require polynomial space. In order to develop an algorithm for maximal split (edge-)subgraphs, we established a bijection between the maximal split subgraphs and the maximal independent sets of an auxiliary graph. For cographs and some subclasses thereof, the algorithms rely on a framework recently introduced by Conte and Uno called Proximity Search. Finally, we considered the extension problem, which consists in deciding if there exists a maximal induced subgraph satisfying a property Π that contains a set of prescribed vertices and

that avoids another set of vertices. We showed that this problem is NP-complete for every "interesting" hereditary property Π . We extended the hardness result to some specific edge version of the extension problem.

8.2.2 Enumeration of minimal transversals of hypergraphs of bounded VC-dimension

Participants: Arnaud Mary.

The problem considered in this paper [19] is the one of enumerating all minimal transversals (also called minimal hitting sets) of a hypergraph \mathcal{H} . An equivalent formulation of this problem known as the transversal hypergraph problem (or hypergraph dualization problem) is to decide, given two hypergraphs, whether one corresponds to the set of minimal transversals of the other. The existence of a polynomial-time algorithm to solve this problem is a long-standing open question. In a paper from 1996 by Fredman and Khachiyan, a first sub-exponential algorithm to solve the transversal hypergraph problem which runs in quasipolynomial time is presented, making it unlikely that the problem is (co)NP-complete. In our paper, currently submitted, we show that when one of the two hypergraphs is of bounded VC-dimension, the transversal hypergraph problem can be solved in polynomial time, or equivalently that if \mathcal{H} is a hypergraph of bounded VC-dimension, then there exists an incremental polynomial-time algorithm to enumerate its minimal transversals. This result generalises most of the previously known polynomial cases in the literature since they almost all consider classes of hypergraphs of bounded VC-dimension. As a consequence, the hypergraph transversal problem is solvable in polynomial-time for any class of hypergraphs closed under partial subhypergraphs. We also show that the proposed algorithm runs in quasi-polynomial time in general hypergraphs, and in polynomial time if the conformality of the hypergraph is bounded, which is one of the few known polynomial cases where the VC-dimension is unbounded.

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

We start by presenting the results obtained within this axis that are related to (pan)genomics in general and which include two papers published in 2024 and one preprint, and conclude by a paper on transcriptomics that was accepted in 2024 and is thus published.

8.3.1 Connecting de Bruijn graphs

Participants: Solon Pissis, Leen Stougie.

In this paper [16], the problem of making a de Bruijn graph (DBG), constructed from a collection of strings, weakly connected while minimizing the total cost of edge additions is studied. The input graph is a DBG that can be made weakly connected by adding edges (along with extra nodes if needed) from the underlying complete DBG. The problem arises from genome reconstruction, where the DBG is constructed from a set of sequences generated from a genome sample by a sequencing experiment. Due to sequencing errors, the DBG is never Eulerian in practice and is often not even weakly connected. We showed that for a DBG $G(V, E)$ of order k consisting of d weakly connected components, (1) making G weakly connected by adding a set of edges of minimal total cost is NP-hard, and (2) no PTAS exists for making G weakly connected by adding a set of edges of minimal total cost (unless the unique games conjecture fails). This result was complemented by showing that there does exist a polynomial-time $(2 - \frac{2}{d})$ -approximation algorithm for the problem.

A restricted version of the above problem was then considered, where we are asked to make G weakly connected by only adding directed paths between pairs of components. We showed that making G weakly connected by adding $d - 1$ such paths of minimal total cost can be done in $O(k|V|\alpha|V|) + |E|$ time, where $\alpha(\cdot)$ is the inverse Ackermann function. This improves on the $O(k|V|\log(|V|) + |E|)$ -time

algorithm proposed by Bernardini et al. in 2022 for the same restricted problem. Finally, we presented an ILP formulation of polynomial size for making G Eulerian with minimal total cost.

8.3.2 A unifying taxonomy of pattern matching in degenerate strings and founder graphs

Participants: Roberto Grossi, Nadia Pisanti.

Elastic Degenerate (ED) strings and Elastic Founder (EF) graphs are two versions of acyclic components of pangomes. Both ED strings and EF graphs (which we collectively name variable strings) extend the well-known notion of indeterminate string. Recent work has extensively investigated algorithmic tasks over these structures, and over several other variable strings notions that they generalise. Among such tasks, the basic operation of matching a pattern into a text, which can serve as a toolkit for many pangenomic data analyses using these data structures, deserves special attention.

In this paper [15], we: (1) highlight a clear taxonomy within both ED strings and EF graphs ranging through variable strings of all types, from the linear string up to the most general one; (2) investigate the problem $PvarT(X, Y)$ of matching a solid or variable pattern of type X into a variable text of type Y ; (3) using as a reference the quadratic conditional lower bounds that are known for $PvarT(solid, ED)$ and $PvarT(solid, EF)$, for all possible types of variable strings X and Y , we either prove the quadratic conditional lower bound for $PvarT(X, Y)$, or provide non-trivial, often sub-quadratic, upper bounds, also exploiting the above-mentioned taxonomy.

8.3.3 Complexity and algorithms for Swap median and relation to other consensus problems

Participants: Arnaud Mary.

The work presented in this preprint [18] that is currently submitted involved also a researcher from Brazil, Luís Felipe Ignácio Cunha, Associate Professor at the Fluminense Federal University of Brazil, who was a visitor at ERABLE for one month funded by Inria, and a student of Luís, Thiago Lopes.

The paper addresses a problem related to genome rearrangements. Genome rearrangements are events in which large blocks of DNA exchange pieces during evolution. The analysis of such events is a way for understanding evolutionary genomics, based on finding the minimum number of rearrangements to transform one genome into another. In a general scenario, more than two genomes are considered, thus leading to new challenges.

The Median problem consists in finding, given three permutations and a distance metric, a permutation s that minimizes the sum of the distances between s and each input. In this paper, the median problem over swap distances in permutations was studied, for which the computational complexity has been open for almost 20 years. We consider this problem through different approaches. We thus associate median solutions and interval convex sets, where the concept of graph convexity inspires the following investigation: Does a median permutation belong to every shortest path between one of the pairs of input permutations? We are able to partially answer this question, and as a by-product we solve a problem that had been open since a long time by proving that the Swap Median problem is NP-hard. Furthermore, using a similar approach, we show that the Closest problem, which seeks to minimize the maximum distance between the solution and the input permutations, is NP-hard even considering three input permutations. This gives a sharp dichotomy into the P vs. NP-hard approaches, since by considering two input permutations, the problem is easily solvable, and by considering any number of input permutations it is known to be NP-hard since 2007. In addition, we show that Swap Median and Swap Closest are APX-hard problems.

8.3.4 Identification and quantification of transposable element transcripts using Long-Read RNA-seq

Participants: Vincent Lacroix, Arnaud Mary, Cristina Vieira.

An initial version of this paper had been submitted in 2023 and presented in the Inria report of that year. The paper has since then been accepted [13].

Transposable elements (TEs) are repeated DNA sequences potentially able to move throughout the genome. In addition to their inherent mutagenic effects, TEs can disrupt nearby genes by donating their intrinsic regulatory sequences, for instance, promoting the ectopic expression of a cellular gene. TE transcription is therefore not only necessary for TE transposition per se but can also be associated with TE-gene fusion transcripts, and in some cases, be the product of pervasive transcription. Hence, correctly determining the transcription state of a TE copy is essential to apprehend the impact of the TE in the host genome.

Methods to identify and quantify TE transcription have mostly relied on short RNA-seq reads to estimate TE expression at the family level while using specific algorithms to discriminate copy-specific transcription. However, assigning short reads to their correct genomic location, and genomic feature is not trivial.

In this paper, full-length cDNA (TeloPrime, Lexogen) of *Drosophila melanogaster* gonads were retrieved and sequenced them using Oxford Nanopore Technologies. We showed that long-read RNA-seq can be used to identify and quantify transcribed TEs at the copy level. In particular, TE insertions overlapping annotated genes are better estimated using long reads than short reads. Nevertheless, long TE transcripts (> 4.5 kb) are not well captured. Most expressed TE insertions correspond to copies that have lost their ability to transpose, and within a family, only a few copies are indeed expressed. Long-read sequencing also allowed the identification of spliced transcripts for around 107 TE copies. Overall, this first comparison of TEs between testes and ovaries uncovers differences in their transcriptional landscape, at the subclass and insertion level.

8.4 Axis 2: Metabolism and (post)transcriptional regulation

In 2024, the work of ERABLE concentrated more on metabolism, while no new results were obtained on (post)transcriptional regulation involving notably small RNAs. The team is however still interested in the latter topic, 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 mostly elementary flux modes (EFMs) and minimal cut set (MCSs) as these are powerful tools for the analysis of metabolic networks, allowing to better understand cellular functioning and to identify potential therapeutic targets. The following sections thus present our recent advances to study various aspects of these concepts, ranging from the thermodynamic analysis of metabolic pathways to the modelling of energetic dysfunctions in neuromuscular diseases, through the study of metabolic interactions in bacterial consortia and the exploration of the links between the Warburg effect and tumor stroma formation.

As may be seen, these works involved also health-related questions. We nevertheless decided to present them in this section, and to just mention them in Axis 4 below.

8.4.1 Computing thermodynamically consistent Elementary Flux Modes with Answer Set Programming

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 the

paper [17], we introduced the software ASPEFM 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. The addition of such constraints makes it possible to eliminate the pathways that lead to non-relevant EFMs. The computation of the EFMs of interest significantly reduces the computational time and saves space. Thermodynamic constraints in terms of the Gibbs energy of the reactions were also added, which constrain the metabolite concentrations within a chosen interval. These constraints are added as a theory propagator and reduce the enumeration during the computation. ASPEFM was applied to the central carbon metabolism of *Escherichia coli* and we showed that the Gibbs energy constraints suppress a large number of non-relevant EFMs.

8.4.2 Logic programming-based Minimal Cut Sets reveal consortium-level therapeutic targets for chronic wound infections

Participants: Sabine Peres.

Minimal Cut Sets (MCSs) identify sets of reactions which, when removed from a metabolic network, disable certain cellular functions. The traditional search for MCSs within genome-scale metabolic models (GSMMs) targets cellular growth, identifies reaction sets resulting in a lethal phenotype if disrupted, and retrieves a list of corresponding genes, mRNAs, or enzyme targets. Using the dual link between MCSs and Elementary Flux Modes (EFMs), the logic programming-based tool ASPEFM MCSs was able to compute MCSs of any size from GSMMs in acceptable running times. The software demonstrated better performance when computing large-sized MCSs than the mixed-integer linear programming methods. In the paper [10], we applied the new MCSs methodology to a medically-relevant consortium model of two cross-feeding bacteria, *Staphylococcus aureus* and *Pseudomonas aeruginosa*. The constraints of ASPEFM were used to bias the computation of MCSs toward exchanged metabolites that could complement lethal phenotypes in individual species. We found that interspecies metabolite exchanges could play an essential role in rescuing single-species growth, for instance inosine could complement lethal reaction knock-outs in the purine synthesis, glycolysis, and pentose phosphate pathways of both bacteria. Finally, MCSs were used to derive a list of promising enzyme targets for consortium-level therapeutic applications that cannot be circumvented via interspecies metabolite exchange.

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

Participants: Sabine Peres, Camille Siharath.

When deciphering metabolic adaptations, traditional biological experiments face challenges due to numerous enzymatic activities, needing modelling to anticipate pathway behaviours and guide research. The paper [14] thus aimed to implement a constraint-based modelling method of muscular energy metabolism, adaptable to individual situations, energy demands, and complex disease-specific metabolic alterations such as muscular dystrophy calpainopathy. Our calpainopathy-like model not only confirmed the ATP production defect under increasing energy demands, but suggested compensatory mechanisms through anaerobic glycolysis. However, excessive glycolysis indicates a need to enhance mitochondrial respiration, preventing excess lactate production common in several diseases. Our model suggested that moderate-intensity physiotherapy, known to improve aerobic performance and anaerobic buffering, combined with increased carbohydrate and amino acid sources, could be a potent therapeutic approach for calpainopathy.

8.4.4 Metabolic modelling links Warburg effect to collagen formation, angiogenesis and inflammation in the tumoral stroma

Participants: Sabine Peres.

Cancer cells are known to exhibit the Warburg effect, characterized by increased glycolysis and lactic acid production even in the presence of oxygen, along with elevated glutamine uptake. Within tumours, these cells are surrounded by collagen, immune cells, and neoangiogenesis. However, the relationship between collagen formation, neoangiogenesis, inflammation, and the Warburg effect remains to be clarified. Using a metabolic model of the human cell containing collagen, neoangiogenesis, and inflammation markers, we computed with our tool ASPEFM a subset of EFMs of biological relevance to the Warburg effect [11]. The EFM with the best linear regression fit to cancer cell line exometabolomics data was selected and showed that collagen production was possible directly *de novo* from glutamine uptake and without extracellular import of glycine and proline, which are the main constituents of collagen.

8.5 Axis 3: (Co)Evolution

As for (post)transcriptional regulation, the work of ERABLE in 2024 on (co)evolution was reduced although in this case, one paper was published on phylogenetic networks that we present below. Again here, the topic still interests ERABLE and will pick up again, involving notably Blerina Sinaimer and also Arnaud Mary and Susana Vinga, one of ERABLE's external members, as well as Marie-France Sagot if she obtains the emeritus. In particular, the submission of a European project that would be coordinated by Blerina Sinaimer, who is now Associate Professor at LUISS University in Rome, Italy, and ERABLE as one of the partners, is currently in preparation which includes (co)evolution as one of its main topics.

8.5.1 Phylogenetic networks: Inferring such from multifurcating trees via cherry picking and machine learning

Participants: Leen Stougie.

Building on previous work on binary trees, we present in the paper [2] the first algorithmic framework to heuristically solve the Hybridization problem for large sets of multifurcating trees whose sets of taxa may differ.

The Hybridization problem asks to reconcile a set of conflicting phylogenetic trees into a single phylogenetic network with the smallest possible number of reticulation nodes. This problem is computationally hard and previous solutions are limited to small and/or severely restricted data sets, for example, a set of binary trees with the same taxon set or only two non-binary trees with non-equal taxon sets.

The new heuristics we introduced combine the cherry-picking technique, recently proposed to solve the same problem for binary trees, with two carefully designed machine-learning models. We demonstrated that our methods are practical and produce qualitatively good solutions through experiments on both synthetic and real data sets.

8.6 Axis 4: Health in general

As indicated in Axis 2 above, some of the work on metabolism developed and published in 2024 concerned health-related questions. These include the papers [10], [11], and [14].

Besides those papers, there are four 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 three other ones [8, 9, 12].

8.6.1 Added value of whole-exome and RNA sequencing in advanced and refractory cancer patients with no molecular-based treatment recommendation based on a 90-gene panel

Participants: Alain Viari.

The objective of the paper [6] was to determine the added value of comprehensive molecular profile by whole-exome and RNA sequencing (WES/RNA-Seq) in advanced and refractory cancer patients who had no molecular-based treatment recommendation (MBTR) based on a more limited targeted gene panel (TGP) plus array-based comparative genomic hybridization (aCGH). **Materials and Methods** In this retrospective analysis, we selected 50 patients previously included in the PROFILER trial (NCT01774409) for which no MBT could be recommended based on a targeted 90-gene panel and aCGH. For each patient, the frozen tumor sample mirroring the FFPE sample used for TGP/aCGH analysis were processed for WES and RNA-Seq. Data from TGP/aCGH were reanalyzed, and together with WES/RNA-Seq, findings were simultaneously discussed at a new molecular tumor board (MTB). **Results** After exclusion of variants of unknown significance, a total of 167 somatic molecular alterations were identified in 50 patients (median: 3 [1-10]). Out of these 167 relevant molecular alterations, 51 (31%) were common to both TGP/aCGH and WES/RNA-Seq, 19 (11%) were identified by the TGP/aCGH only and 97 (58%) were identified by WES/RNA-Seq only, including two fusion transcripts in two patients. A MBTR was provided in 4/50 (8%) patients using the information from TGP/aCGH versus 9/50 (18%) patients using WES/RNA-Seq findings. Three patients had similar recommendations based on TGP/aCGH and WES/RNA-Seq. **Conclusions** In advanced and refractory cancer patients in whom no MBTR was recommended from TGP/aCGH, WES/RNA-Seq allowed to identify more alterations which may in turn, in a limited fraction of patients, lead to new MBTR.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Inria associate team not involved in an IIL or an international program

Capoeira

Participants: Giuseppe Italiano, Vincent Lacroix, Alberto Marchetti-Spaccamela, Arnaud Mary, Marie-France Sagot, Blerina Sinimeri, Leen Stougie.

Title: Computational Approaches with the Objective to Explore intra and cross-species Interactions and their Role in All domains of life

Duration: 2020 - 2022, extended to 2024 because of the pandemic

Coordinators: Marie-France Sagot (ERABLE) and André Fujita (Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil)

Webpage: [Capoeira](#)

9.1.2 Participation in other International Programs

Ahimsa

Participants: Arnaud Mary, Marie-France Sagot, Blerina Sinimeri.

Title: Alternative approach to Investigating and Modelling Sickness and health.

Coordinators: M.-F. Sagot (ERABLE), A. Ávila (Instituto de Biologia Molecular do Paraná - Fiocruz-PR, Curitiba, Paraná, Brazil).

Additional ERABLE participants: M. Ferrarini.

Type: Capes-Cofecub (2020-2022, extended until 2024).

Webpage: [Ahimsa](#)

9.2 International research visitors

9.2.1 Visits of international scientists

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

Visit of Renata Wassermann, Associate Professor, University of São Paulo (USP), Brazil, from August 31 to September 7 to Erable in Lyon. This had various objectives all related with the problem of the representation of genome-scale metabolic networks (also called GEMs). 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.

Visit of A. Silber and G. T. Montanaro to Erable in Lyon from October 14 to October 24. The purpose of such visit was to discuss with the members of Erable but also with the persons who were visiting Erable at the time, including N. Beerenwinkel from ETH-Zürich and three of his PhD students in the context of the H2020 Twinning project Olissipo, Jörg Stelling also from ETH-Zürich, and Lucas Gentil Azevedo from Fiocruz, Bahia, Brazil (see below).

9.2.2 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

9.3 European initiatives

9.3.1 H2020 projects

Olissipo

Participants: Giuseppe Italiano, Vincent Lacroix, Alberto Marchetti-Spaccamela, Arnaud Mary, Marie-France Sagot, Blerina Sinimeri, Leen Stougie, Alain Viari.

Title: Fostering Computational Biology Research and Innovation in Lisbon.

Coordinator: Susana Vinga, INESC-ID, Instituto Superior Técnico, Lisbon.

Other participants: Inria EPI ERABLE, the Swiss Federal Institute of Technology (ETH Zürich) in Switzerland, and the European Molecular Biology Laboratory (EMBL) in Germany.

ERABLE coordinator: Marie-France Sagot.

Type: H2020 Twinning.

Comments: Due to the Covid-19, the start of this project was delayed until January 1st, 2021. For the same reason, although it should have lasted until the end of 2023, it was extended until the end of June 2024.

Webpages: [Olissipo-Erable](#) and [Olissipo](#)

9.4 National initiatives

9.4.1 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.

Web page: Not available.

9.4.2 Others

MITOTIC

Participants: Sabine Peres.

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

Coordinator: Sabine Peres.

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

Duration: 2021-2024.

Web page: Not available.

Notice that, besides the project above, were included here also national projects of our members from Italy and the Netherlands when these have no other partners than researchers from the same country. These concern the following:

Networks

Participants: Solon Pissis, Leen Stougie.

Title: Networks.

Coordinator: Michel Mandjes, University of Amsterdam.

Type: NWO Gravity Program.

Duration: 2014-2024.

Web page: [Networks](#).

Optimal

Participants: Leen Stougie.

Title: Optimization for and with Machine Learning.

Coordinator: Dick den Hertog.

Type: NWO ENW-Groot Program.

Web page: Not available.

10 Dissemination

10.1 Promoting scientific activities

10.1.1 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)*.
- Solon Pissis was co-chair of the Program Committee of *Alenex* and *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

- Sabine Peres was co-organiser of the meeting of the GT **BIOSS**, May 27-28, 2024.
- Marie-France Sagot was co-organiser of the **First Edition of the Workshop Phylogeny and Cophylogeny: Tree for a Tango**, held virtually, Nov 4-5, 2024; and co-organiser of the **Fourth Edition of the Workshop Metabolism and mathematical models: Two for a tango**, held virtually, Nov 19-20, 2024.

Member of the conference program committees

- Vincent Lacroix was a member of the Program Committee of *JOBIM* and *SeqBim*.
- Arnaud Mary was a member of the Program Committee of *ISMB*.
- Sabine Peres was a member of the Program Committee of *CMSB*.
- 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*, see *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*.

10.1.2 Scientific expertise

Giuseppe Italiano is since 2024 President of the European Association for Theoretical Computer Science (EATCS). He is since 2023 Scientific Co-Director of the Master in Artificial Intelligence, 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.

Vincent Lacroix is since 2022 member of the Advisory Committee Section 67-68 of the University Lyon 1, and internal member of the direction/administration of the E2M2 doctoral school of the University of Lyon 1.

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 is since 2014 member of the Scientific Advisory Board of CWI, and since 2022 member of the Scientific Advisory Board of the Dept. of Computational Biology at the Univ. of Lausanne, Switzerland. Since 2022 also, she is 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. From 2020 to 2024, she was member of the Review Committee for the Human Frontier Science Program. Finally, since 2023, she is member of the office ("bureau") of the SFBI ("Société Française de BioInformatique").

Alain Viari is member of the scientific advisory board of IRT (Institut de Recherche Technologique) BioAster and of Centre Léon Bérard. He also coordinates together with J.-F. Deleuze (CNRGH-Evry) the Research and Development part (CReflX) of the "Plan France Médecine Génomique 2025".

Cristina Vieira is member of the "Conseil National des Universités" (CNU) 67 ("Biologie des Populations et Écologie"), and was from 2017 to 2024 member of the "Conseil de la Faculté des Sciences et Technologies (FST)" of the University Lyon 1.

10.1.3 Research administration

Marie-France Sagot is since 2021 member of the "Conseil Scientifique (COS)" and of the "COMité des Moyens Incitatifs (COMI)" for Inria Lyon.

10.2 Teaching - Supervision - Juries

10.2.1 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).

Vincent Lacroix is co-responsible for the M1 master in bioinformatics (together with Arnaud Mary) and responsible for 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 2024. Since 2021, he is also involved in the group who proposed a new course called Climate and Transitions, mandatory for L1 students in Science at University Lyon1 (1500 students). Most of the course is a MOOC (see [here](#)), but there are also 4 occasions where teachers and students discuss the topics covered by the course with various group activities described briefly [here](#). Since 2023, the course is also proposed as an optional one for students at Université Lyon 2.

Arnaud Mary is co-responsible for the M1 master in bioinformatics (together with Vincent Lacroix) and for two courses of the Bioinformatics Curriculum at the University (M1: Object Oriented Programming, M2: new course on Advanced Algorithms for Bioinformatics). He taught 198 hours in 2024.

Sabine Peres is 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).

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.

Moreover, Emma Crisci and Sasha Darmon participate in the teaching via the "Aide Complémentaire aux Enseignements" at a rate of 64h per 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 and 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).

10.2.2 Supervision

The following are the PhDs in progress or which ended in 2024:

- Emma Crisci, University Lyon 1 & Inria (supervisor: Sabine Peres, together with Arnaud Mary)
- Sasha Darmon, University Lyon 1 (supervisor: Vincent Lacroix, together with Arnaud Mary)
- Esteban Gabory, CWI (supervisor: Solon Pissis)
- Pierre Gerenton, University Lyon 1 (supervisor Bastien Boussau, together with Vincent Lacroix)
- Moses Njagi Mwaniki, Università di Pisa (supervisor: Nadia Pisanti)
- Camille Siharath, University Lyon 1 (supervisors: Sabine Peres and Olivier Biondi, University Evry Paris-Saclay)
- Michelle Sweering, CWI (co-supervisors: Solon Pissis and Leen Stougie)

10.2.3 Juries

The following are the PhD and HDR juries to which members of ERABLE participated in 2024:

- Vincent Lacroix : Member of the the HDR committee of Sylvain Foissac, University Toulouse Midi-Pyrénées, June 2024; Member of the HDR committee of Anamaria Necsulea, University Lyon 1, January 2024.
- Sabine Peres : President of the PhD jury of Florian Bénitière, University Lyon 1, May 2024; Reviewer of the PhD of Kate E. Meeson, University of Manchester, July 2024; Reviewer of the PhD of Delphine Nègre, University of Nantes, December 2024.
- Marie-France Sagot : Reviewer of the HDR of Sylvain Foissac, University Toulouse Midi-Pyrénées, June 2024; Member of the HDR committee of Nicolas Bousquet, University of Lyon, September 2024; Reviewer of the HDR of Philippe Gambette, University of Paris-Est Sup, October 2024; Reviewer of the PhD of Francesco Strozzi, Conservatoire National des Arts et Métiers (CNAM), September 2024; Reviewer of the PhD of Tomas Caetano, University Toulouse 3 Paul Sabatier, October 2024.

10.3 Popularization

10.3.1 Specific official responsibilities in science outreach structures

Sasha Darmon is president of the science popularisation association **Démesures**, whose mission is to promote critical thinking in the sciences and make research more accessible to all.

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

Sasha Darmon developed two new activities for the association **Démesures** on AI-generated content, explaining how it works, how to recognize it, and what to watch out for with these emerging technologies.

10.3.3 Participation in Live events

Emma Crisci participated in two doctoral and high school student meetings, organised over two days. The goal was to introduce the professions in research and to explain her PhD work on metabolism in a more accessible way. Emma needed to create a paper resource to make the meeting more engaging.

Sasha Darmon organised, set up, and led a science outreach stand for a weekend at the **Geek Touch / Japan Touch Haru** convention which took place in May 2024. The goal was to engage visitors, who had initially come for different events, and introduce them to science through fun activities, so they would leave with a broader and more accessible view of it. We focused on cognitive biases and critical thinking, essential in both science and everyday life.

10.3.4 Other science outreach relevant activities

Sasha Darmon was Laureate of the “Sciences en Bulles 2025” project, which involved popularising his PhD in a french comic book form, set for publication at the 2025 Fête de la Science.

Vincent Lacroix participated 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.

11 Scientific production

11.1 Publications of the year

International journals

- [1] G. Bernardini, E. Gabory, S. P. Pissis, L. Stougie, M. Sweering and W. Zuba. ‘Elastic-Degenerate String Matching with 1 Error or Mismatch’. In: *Theory of Computing Systems* (16th Sept. 2024). DOI: [10.1007/s00224-024-10194-8](https://doi.org/10.1007/s00224-024-10194-8). URL: <https://inria.hal.science/hal-04747458> (cit. on p. 16).
- [2] G. Bernardini, L. van Iersel, E. Julien and L. Stougie. ‘Inferring phylogenetic networks from multifurcating trees via cherry picking and machine learning’. In: *Molecular Phylogenetics and Evolution* 199 (Oct. 2024). DOI: [10.1016/j.ympev.2024.108137](https://doi.org/10.1016/j.ympev.2024.108137). URL: <https://inria.hal.science/hal-04747453> (cit. on p. 21).
- [3] C. Brosse, A. Lagoutte, V. Limouzy, A. Mary and L. Pastor. ‘Efficient enumeration of maximal split subgraphs and induced sub-cographs and related classes’. In: *Discrete Applied Mathematics* 345 (15th Mar. 2024), pp. 34–51. DOI: [10.1016/j.dam.2023.10.025](https://doi.org/10.1016/j.dam.2023.10.025). URL: <https://insep.hal.science/hal-04058748> (cit. on p. 16).
- [4] T. Calamoneri, M. Lafond, A. Monti and B. Sinaimeri. ‘On Generalizations of Pairwise Compatibility Graphs’. In: *Discrete Mathematics and Theoretical Computer Science* vol. 26:3. Graph Theory (6th Oct. 2024). DOI: [10.46298/dmtcs.12295](https://doi.org/10.46298/dmtcs.12295). URL: <https://inria.hal.science/hal-04803957> (cit. on p. 16).
- [5] P. Charalampopoulos, H. Chen, P. Christen, G. Loukides, N. Pisanti, S. P. Pissis and J. Radoszewski. ‘Pattern Masking for Dictionary Matching: Theory and Practice’. In: *Algorithmica* 86 (6th Mar. 2024), pp. 1948–1978. DOI: [10.1007/s00453-024-01213-8](https://doi.org/10.1007/s00453-024-01213-8). URL: <https://inria.hal.science/hal-04747475> (cit. on p. 16).
- [6] A. Dufresne, V. Attignon, A. Ferrari, L. Tonon, S. Boyault, S. Tabone-Eglinger, P. Cassier, O. Trédan, N. Corradini, A. Vinceneux, A. Swalduz, A. Viari, S. Chabaud, D. Pérol, J.-Y. Blay and P. Saintigny. ‘Added value of whole-exome and RNA sequencing in advanced and refractory cancer patients with no molecular-based treatment recommendation based on a 90-gene panel’. In: *Cancer Medicine* 13.7 (2024), e7115. DOI: [10.1002/cam4.7115](https://doi.org/10.1002/cam4.7115). URL: <https://hal.science/hal-04646005> (cit. on p. 22).
- [7] A.-S. Fiston-Lavier, S. Dérozier, G. Perrière and M.-F. Sagot. ‘ISMB/ECCB 2023 organization benefited from the strengths of the French bioinformatics community’. In: *Bioinformatics Advances* 4.1 (1st Jan. 2024), vbae040. DOI: [10.1093/bioadv/vbae040](https://doi.org/10.1093/bioadv/vbae040). URL: <https://hal.inrae.fr/hal-04684148> (cit. on p. 16).
- [8] C. Hernandez, C. Verzeroli, A. A. Roca Suarez, A. J. Farca Luna, L. Tonon, R. Esteban-Fabro, R. Pinyol, M.-I. Plissonnier, I. Chicherova, A. Dubois, P. Bellaud, M. Seffals, B. Turlin, A. Fautrel, G. Ichim, M. Rivoire, G. Passot, Z. M. Jílková, T. Decaens, A. Viari, B. Testoni, S. Rebouissou, J. M. Llovet, F. Zoulim and R. Parent. ‘Hepatocellular carcinoma hosts cholinergic neural cells and tumoral hepatocytes harboring targetable muscarinic receptors’. In: *JHEP reports : innovation in hepatology* 7.1 (2024), p. 101245. DOI: [10.1016/j.jhepr.2024.101245](https://doi.org/10.1016/j.jhepr.2024.101245). URL: <https://hal.science/hal-04876113> (cit. on p. 21).

- [9] G. Lalle, R. Lautraite, K. Bouherrou, M. Plaschka, A. Pignata, A. Voisin, J. Twardowski, M. Perrin-Niquet, P. Stéphan, S. Durget, L. Tonon, M. Ardin, C. Degletagne, A. Viari, L. Belgarbi Dutron, N. Davoust, T. Postler, J. Zhao, C. Caux, J. Caramel, S. Dalle, P. Cassier, U. Klein, M. Schmidt-Supprian, R. Liblau, S. Ghosh and Y. Grinberg-Bleyer. ‘NF- κ B subunits RelA and c-Rel selectively control CD4+ T cell function in multiple sclerosis and cancer’. In: *Journal of Experimental Medicine* 221.6 (3rd June 2024). DOI: [10.1084/jem.20231348](https://doi.org/10.1084/jem.20231348). URL: <https://hal.science/hal-04567979> (cit. on p. 21).
- [10] M. Mahout, R. Carlson, L. Simon and S. Peres. ‘Logic programming-based Minimal Cut Sets reveal consortium-level therapeutic targets for chronic wound infections’. In: *npj Systems Biology and Applications* 10.1 (2024), p. 34. DOI: [10.1038/s41540-024-00360-6](https://doi.org/10.1038/s41540-024-00360-6). URL: <https://hal.science/hal-04610153> (cit. on pp. 20, 21).
- [11] M. Mahout, L. Schwartz, R. Attal, A. Bakkar and S. Peres. ‘Metabolic modelling links Warburg effect to collagen formation, angiogenesis and inflammation in the tumoral stroma’. In: *PLoS ONE* 19.12 (2024), e0313962. DOI: [10.1371/journal.pone.0313962](https://doi.org/10.1371/journal.pone.0313962). URL: <https://hal.science/hal-04873696> (cit. on p. 21).
- [12] M. Pagliuca, J. Havas, E. Thomas, Y. Drouet, D. Soldato, M. A. Franzoi, J. Ribeiro, C. Chiodi, E. Gillanders, B. Pistilli, G. Menvielle, F. Joly, F. Lerebours, O. Rigal, T. Petit, S. Giacchetti, F. Dalenc, J. Wassermann, O. Arsene, A. L. Martin, S. Everhard, O. Tredan, S. Boyault, M. de Laurentiis, A. Viari, J. F. Deleuze, A. Bertaut, F. André, I. Vaz-Luis and A. Di Meglio. ‘Long-term behavioral symptom clusters among survivors of early-stage breast cancer: Development and validation of a predictive model’. In: *JNCI: Journal of the National Cancer Institute* (9th Sept. 2024). DOI: [10.1093/jnci/djae222](https://doi.org/10.1093/jnci/djae222). URL: <https://hal.science/hal-04765650> (cit. on p. 21).
- [13] R. Rebollo, P. Gerenton, E. Cumunel, A. Mary, F. Sabot, N. Burlet, B. Gillet, S. Hughes, D. Oliveira, C. Goubert, M. Fablet, C. Vieira and V. Lacroix. ‘Identification and quantification of transposable element transcripts using Long-Read RNA-seq in Drosophila germline tissues’. In: *Peer Community Journal* 4.e89 (10th Sept. 2024). DOI: [10.24072/pcjournal.457](https://doi.org/10.24072/pcjournal.457). URL: <https://hal.science/hal-04285238> (cit. on p. 19).
- [14] C. Siharath, O. Biondi and S. Peres. ‘Modelling energy metabolism dysregulations in neuromuscular diseases: A case study of calpainopathy’. In: *Heliyon* 10.24 (2024), e40918. DOI: [10.1016/j.heliyon.2024.e40918](https://doi.org/10.1016/j.heliyon.2024.e40918). URL: <https://hal.science/hal-04845028> (cit. on pp. 20, 21).

International peer-reviewed conferences

- [15] R. Ascone, G. Bernardini, A. Conte, M. Equi, E. Gabory, R. Grossi and N. Pisanti. ‘A Unifying Taxonomy of Pattern Matching in Degenerate Strings and Founder Graphs’. In: WABI 2024 - Workshop on Algorithms in Bioinformatics. London, United Kingdom: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. DOI: [10.4230/LIPIcs.WABI.2024.14](https://doi.org/10.4230/LIPIcs.WABI.2024.14). URL: <https://inria.hal.science/hal-04747535> (cit. on p. 18).
- [16] G. Bernardini, H. Chen, I. L. Gørtz, C. Krogh, G. Loukides, S. Pissis, L. Stougie and M. Sweering. ‘Connecting de Bruijn Graphs’. In: CPM 2024 - 35th Annual Symposium on Combinatorial Pattern Matching. Fukuoka, Japan: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. DOI: [10.4230/LIPIcs.CPM.2024.6](https://doi.org/10.4230/LIPIcs.CPM.2024.6). URL: <https://inria.hal.science/hal-04747471> (cit. on p. 17).
- [17] E. Crisci, M. Mahout and S. Peres. ‘Computing Thermodynamically Consistent Elementary Flux Modes with Answer Set Programming’. In: *Computational Methods in Systems Biology, 22nd International Conference, CMSB 2024, Pisa, Italy, September 16–18, 2024, Proceedings*. CMSB 2024 - 22nd International Conference Computational Methods in Systems Biology. Vol. LNCS-14971. Lecture Notes in Computer Science. Pise, Italy: Springer Nature Switzerland, 19th Sept. 2024, pp. 80–88. DOI: [10.1007/978-3-031-71671-3_7](https://doi.org/10.1007/978-3-031-71671-3_7). URL: <https://hal.science/hal-04707880> (cit. on p. 20).

Reports & preprints

- [18] L. Cunha, T. Lopes and A. Mary. *Complexity and algorithms for Swap median and relation to other consensus problems*. 2024. DOI: [10.48550/arXiv.2409.09734](https://doi.org/10.48550/arXiv.2409.09734). URL: <https://hal.science/hal-04886139> (cit. on p. 18).

-
- [19] A. Mary. *Enumeration of minimal transversals of hypergraphs of bounded VC-dimension*. 2024. DOI: [10.48550/arXiv.2407.00694](https://doi.org/10.48550/arXiv.2407.00694). URL: <https://hal.science/hal-04886143> (cit. on pp. 16, 17).