RESEARCH CENTRE

Inria Lyon Centre

IN PARTNERSHIP WITH:

Institut national des sciences appliquées de Lyon, Université Claude Bernard (Lyon 1), CNRS

2024 ACTIVITY REPORT

Project-Team BEAGLE

Artificial Evolution and Computational Biology

IN COLLABORATION WITH: Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS)

DOMAIN

Digital Health, Biology and Earth

THEME

Computational Biology

Inria

Contents

Project-Team BEAGLE

Creation of the Project-Team: 2013 January 01

Keywords

Computer sciences and digital sciences

A3.3. – Data and knowledge analysis

A3.3.2. – Data mining

A3.3.3. – Big data analysis

A6.1.1. – Continuous Modeling (PDE, ODE)

A6.1.3. – Discrete Modeling (multi-agent, people centered)

A6.1.4. – Multiscale modeling

A6.2.7. – High performance computing

A8.1. – Discrete mathematics, combinatorics

Other research topics and application domains

B1. – Life sciences

B1.1. – Biology

B1.1.2. – Molecular and cellular biology

B1.1.6. – Evolutionnary biology

B1.1.7. – Bioinformatics

B1.1.10. – Systems and synthetic biology

B1.1.11. – Plant Biology

B3. – Environment and planet

B3.1. – Sustainable development

B3.1.1. – Resource management

B3.5. – Agronomy

B3.6. – Ecology

B3.6.1. – Biodiversity

B9. – Society and Knowledge

B9.2. – Art

B9.2.1. – Music, sound

B9.2.4. – Theater

B9.9. – Ethics

1 Team members, visitors, external collaborators

Research Scientists

- Antonius Crombach [INRIA, Researcher]
- Eric Tannier [INRIA, Senior Researcher, until Nov 2024, HDR]

Faculty Members

- Guillaume Beslon [Team leader, INSA LYON, Professor, HDR]
- Carole Knibbe [INSA LYON, Professor, HDR]
- Christophe Rigotti [INSA LYON, Associate Professor, HDR]
- Jonathan Rouzaud-Cornabas [INSA LYON, Associate Professor]

Post-Doctoral Fellows

- Jean-Sebastien Beaulne [INRIA, Post-Doctoral Fellow, until Feb 2024]
- Hamza Chegraoui [INRIA, Post-Doctoral Fellow]

PhD Students

- Lisa Chabrier [INRIA, until Oct 2024, (from 01/01/2024 to 31/10/2024); INSA LYON (from 01/11/2014 to 31/12/2024)]
- Romain Galle [INRIA]
- Juliette Luiselli [INSA LYON]
- Arsene Marzorati [INRIA]
- Sofia Pacheco Garcia [INRIA, from Nov 2024]
- Thibaut Peyric [INRIA]

Technical Staff

- Mouhamad Al-Sayed Ali [INRIA, Engineer]
- David Parsons [INRIA, Engineer]

Interns and Apprentices

- Guillermo Benito Calvino [INRIA, Intern, from Mar 2024 until Aug 2024]
- Matthieu Deleglise [INRIA, Intern, from Jun 2024 until Aug 2024]
- Basile Gandon [INRIA, Intern, from Feb 2024 until Apr 2024]
- Gabin Joussot-Dubien [INRIA, Intern, from Jul 2024 until Sep 2024]
- Alexandre Layous [INRIA, Intern, from Apr 2024 until Aug 2024]
- Nils Ledin [INRIA, Intern, from Oct 2024]
- Sofia Pacheco Garcia [INRIA, Intern, from Mar 2024 until Aug 2024]

Administrative Assistant

- Sylvie Boyer [INRIA]

2 Overall objectives

2.1 An interface between biology and computer science

The expanded name for the BEAGLE research group is "Artificial Evolution and Computational Biology". Our aim is to position our research at the interface between biology and computer science and to contribute new results in biology by modeling biological systems. In other words we are making artifacts – from the Latin *artis factum* (an entity made by human art rather than by Nature) – and we explore them in order to understand Nature. The team is an INRIA Project-Team since January, 2014. It gathers researchers from INRIA, INSA, who are members of three different labs, the LIRIS [1], the LBBE [2], and CARMEN [3]. It is led by Prof. Guillaume Beslon (INSA-Lyon, LIRIS, Computer Science Dept.).

Our research program requires the team members to have skills in computer science but also in life sciences: they must have or develop a strong knowledge in biosciences to interact efficiently with biologists or, ideally, to directly interpret the results given by the models they develop. A direct consequence of this claim is that it is mandatory to restrict the domain of expertise in life sciences. This is why we focus on a specific scale, central in biology: the cellular scale. Indeed, we restrict our investigations on the cell, viewed as a dynamical system made of molecular elements. This specific scale is rich in open questions that deserve modeling and simulation approaches. We also focus on two different kinds of constraints that structure the cellular level: biophysical constraints and historical constraints. The cell is a system composed of molecules that physically interact and the spatio-temporal nature of these interactions is likely to strongly influence its dynamics. But the cell is also the result of an evolutionary process that imposes its own limits on what can evolve (or is the most likely to evolve) and what cannot (or is the less likely to evolve). A better understanding of what kind of systems evolution is the most likely to lead to in a given context could give us important clues for the analysis of extant biological systems.

2.2 An organization into two tools and three main axes

To study these two kinds of constraints we mainly rely on two specific tools: computational cellular biochemistry and evolution models. We use these tools to develop our "artifacts" and we compare their output with real data, either direct measurements collected by experimentalists or ancestral properties computationally inferred from their extant descendants. The team research is currently organized in four main research axes. The first two ones are methodologically-oriented: we develop general formalisms and tools for computational cellular biochemistry (research axis 1) and families of models to study the evolutionary process (research axis 2). Eventually the last axis aims at integrating the two tools, computational biochemistry and evolution, in what we call "Evolutionary Systems Biology" (research axis 3). The next three sections describe these three axes in more details. The biological questions described are not the sole topics tackled by the team. They are the ones that mobilize a substantial fraction of the researchers on the long run. Many other questions are tackled by individual researchers or even small groups. In the following these ones will be briefly described in their methodological context, i.e. in the two sections devoted to research axes 1 and 2.

2.3 A strategy

The scientific objective of the BEAGLE team is to develop a consistent set of concepts and tools – mainly based on computational science – to *in fine* contribute to knowledge discovery in systems biology. Our strategy is to develop strong interactions with life science researchers to become active partners of the

[1] Laboratoire d'Informatique en Image et Systèmes d'Information: UMR 5205 CNRS, INSA-Lyon, Univ. Claude Bernard Lyon 1, Univ. Louis Lumière Lyon 2, École Centrale de Lyon

[2] Laboratoire de Biometrie et Biologie Evolutive: UMR CNRS 5558, Univ. Claude Bernard Lyon 1.

[3] Laboratoire de Recherche en Cardiovasculaire, Métabolisme, Diabétologie et Nutrition: UMR U1060 INSERM, INSA-Lyon, INRA 1235, Univ. Claude Bernard Lyon 1.

biological discovery process. Thus, our aim as a team is not to be a computer science team interacting with biologists, nor to be a team of biologists using computer science tools, but rather to stay in the middle and to become a *trading zone* between biology and computer science. Our very scientific identity is thus fuzzy, melting components from both sciences. Indeed, one of the central claims of the team is that interdisciplinarity involves permanent exchanges between the disciplines. Such exchanges can hardly be maintained between distant teams. That's why the BEAGLE team tries to develop local collaborations with local scientists. That's also why BEAGLE also tries to organize itself as an intrinsically interdisciplinary group, gathering different sensitivities between biology and computer science inside the group. Our ultimate objective is to develop interdisciplinarity at the individual level, all members of the team being able to interact efficiently with specialists from both fields.

3 Research program

3.1 Introduction

As stated above, the research topics of the BEAGLE Team are centered on the modeling and simulation of cellular processes. More specifically, we focus on two specific processes that govern cell dynamics and behavior: Biophysics and Evolution. We are strongly engaged into the integration of these level of biological understanding.

3.2 Research axis 1: Computational cellular biochemistry

Biochemical kinetics developed as an extension of chemical kinetics in the early 20th century and inherited the main hypotheses underlying Van't Hoff's law of mass action : a perfectly-stirred homogeneous medium with deterministic kinetics. This classical view is however challenged by recent experimental results regarding both the movement and the metabolic fate of biomolecules. First, it is now known that the diffusive motion of many proteins in cellular media exhibits deviations from the ideal case of Brownian motion, in the form of position-dependent diffusion or anomalous diffusion, a hallmark of poorly mixing media. Second, several lines of evidence indicate that the metabolic fate of molecules in the organism not only depends on their chemical nature, but also on their spatial organisation – for example, the fate of dietary lipids depends on whether they are organized into many small or a few large droplets. In this modern-day framework, cellular media appear as heterogeneous collections of contiguous spatial domains with different characteristics, thus providing spatial organization of the reactants. To improve our understanding of intracellular biochemistry, we study spatiotemporal biochemical kinetics using mathematical models and numerical simulations.

Specifically, with our biomedical partner lab (CarMeN) and our collaborators (Karolinska University Hospital, Stockholm), we investigate the molecular and cellular mechanisms that drive the fate of lipids as they are digested, absorbed, stored and mobilised in the body. Our goal is to build quantitative models of lipid fate based on data from in vitro digestion studies (enzymatic kinetics), from cellular cultures (gene expression data for genes involved in lipid metabolism or inflammation), from animal studies and from clinical studies (physiological data at the level of tissues and organs, like subcutaneous fat mass and visceral fat mass). We aim at taking into account the influence of the spatial organisation of dietary lipids (e.g. whether and how they are emulsified) on their fate in the body.

3.3 Research axis 2: Models for Molecular Evolution

We study the processes of genome evolution, with a focus on large-scale genomic events (rearrangements, duplications, transfers). We are interested in deciphering general laws which explain the organization of the genomes we observe today, as well as using the knowledge of these processes to reconstruct some aspects of the history of life. To do so, we construct mathematical models and apply them either in a "forward" way, *i.e.* observing the course of evolution from known ancestors and parameters, by simulation (*in silico experimental evolution*) or mathematical analysis (*theoretical biology*), or in a "backward" way, *i.e.* reconstructing ancestral states and parameters from known extant states (*phylogeny, comparative genomics*). Moreover we often mix the two approaches either by validating backwards reconstruction

methods on forward simulations, or by using the forward method to test evolutionary hypotheses on biological data.

3.4 Research axis 3: Evolutionary Systems Biology

This axis, consisting in integrating the two main biological levels we study, is a long-standing and long-term objective in the team. Over the last years, we have accomplished significant advances in this direction, mainly due to the evolution of the team staff and team projects. These novel developments allow us to give this axis a central place in the next (future) team. We have several middle term projects that integrate molecular data and evolution. First results were reported in 2019 with respect to an evolutionary perspective on chromatin-associated proteins. In this report, we discuss results derived from data analysis of single-cell multi-omics RNA expression data of different mouse brain regions. Other, ongoing projects include reverse engineering the regulatory networks of 'old' and 'young' brain regions through image bioinformatics and finding new therapeutic targets for lung tumours that evolve treatment resistance.

4 Application domains

4.1 Functional and Evolutionary Biology

We do not usually distinguish our research and its application domains. Our shared idea is that the research is oriented by a scientific question, which in the case of the Beagle team is a multidisciplinary one, most often of biological nature. We do not develop methodologies independently from this question and then look for applications. Instead we collectively work with other disciplines to solve a question, using our competencies.

In consequence the application domains are already listed in the description of our projects and goals. They concern functional and evolutionary biology, related to critical social questions as human or global health.

4.2 Implication domains

We still advocate for the "application domains" section of the activity report to be called "implication domains" to broaden its scope. Implication contains applications, but not conversely.

This could allow us and others to report for example on orientation activities of our research programs guided by a social demand rather than by an intrinsic dynamic of scientific evolution, a simple claim for "progress", or a social demand coming only from industry.

This could allow a better awareness of social and environmental issues, and integrate them in this section.

5 Social and environmental responsibility

5.1 Footprint of research activities

We gave several conferences and lectures on the rebound effects of optimising the efficiency of devices for environmental footprint attenuation (Rennes, Lyon)

5.2 Impact of research results

We have decided to handle this question as a research theme, in the Beagle team in 2024 and in SEMIS (Sciences, milieux, information, sociétés), one of the teams arising from Beagle and created in December 2024. The main focus of Semis is on social, politics and environmental impacts of science and technologies (see section 6.1).

6 Highlights of the year

6.1 Reorganization of the team

The Beagle team is arriving at the end of its 12 years lifecycle (the team has been created in 2013). As is often the case during such a period, this led to a reshuffling of the team, of team memberships and main team research topics. All Beagle members are activelly engaged in discussions for the creation of two new teams (note that part of the original Beagle team had already spined-off in 2022 following the creation of the AIStroSight team by Hugues Berry): BioTiC (Biologie Théorique et Computationnelle, that will be led by Guillaume Beslon) et Semis (Sciences, milieux, information, sociétés, led by Eric Tannier). Semis has been created as a new team in December 2024. For administrative reasons, the creation of the BioTiC team has been delayed waiting for most academic members of the team to move from the LIRIS Lab to the CITI Lab.

6.2 New team members

The reorganization of the team goes along with new members recruitment and several researchers have manifested their interest for the two new teams. We would like to highlight the fact that, in 2024, Beagle (or BioTiC/Semis) has attracted several researchers that joined (or will joint soon) the Lyon Inria center to reinforce its research activities:

Sandro Colizzi Enrico Sandro Colizzi is an theoretical biologist specializing in the evolution of biological novelty. Since 2021, he has been an associate researcher at the Sainsbury Laboratory at the University of Cambridge. His work focuses on the origins of life, microbial eco-evolutionary dynamics, and the transition to multicellularity, using mathematical models to understand the evolution of multi-level complexity. He earned his PhD in 2016 at Utrecht University, studying the evolution of symbolic information processing in biological systems. He has been recruited Inria CRCN in 2024 and will join the Beagle/BioTiC team in February 2025.

Pablo Jensen Formerly a physicist, Pablo Jensen has became prominent figure in interdisciplinary research, applying physical modeling techniques to understand social systems. This unique approach has led him to explore how concepts from physics can shed light on complex societal behaviors. He has authored several books, including “Pourquoi la société ne se laisse pas mettre en équations”, where he critically examines the limitations of mathematical models in understanding social phenomena. His work bridges the gap between the natural and social sciences, fostering dialogue on the role of science in society. Pablo Jensen is DR CNRS at the ENS-Lyon. He joined the Semis team in December 2024.

Clément Moulin-Frier Clément Moulin-Frier earned his PhD in the Engineering of Cognition, Interaction, Learning, and Creation from the University of Grenoble. Throughout his career, he has held various research positions, including at the SPECS laboratory at UPF in Barcelona and as a research scientist at Cogitai Inc. Since 2019, he is Inria CRCN in the Flower team (Inria Bordeaux). He is specialized on the study of “evolutionary transitions”, especially emergence of life, cultural evolution and the formation of social behaviors. He develop computational models of prebiotic evolution, language evolution and development. In 2024, he initiated discussions with the Beagle team (including a one-week visit in Lyon in June 2024) which led him to request a transfer to the inria center in Lyon to join the future BioTiC team. His transfer is scheduled for spring 2025.

6.3 Awards

Guillaume Beslon has been awarded “Best reviewer” by the International Society for Artificial Life.

7 New software, platforms, open data

7.1 New software

7.1.1 aevol

Name: Artificial Evolution

Keywords: Evolution, Simulation

Functional Description: Aevol is a digital genetics model: populations of digital organisms are subjected to a process of selection and variation, which creates a Darwinian dynamics. By modifying the characteristics of selection (e.g. population size, type of environment, environmental variations) or variation (e.g. mutation rates, chromosomal rearrangement rates, types of rearrangements, horizontal transfer), one can study experimentally the impact of these parameters on the structure of the evolved organisms. In particular, since Aevol integrates a precise and realistic model of the genome, it allows for the study of structural variations of the genome (e.g. number of genes, synteny, proportion of coding sequences).

The simulation platform comes along with a set of tools for analysing phylogenies and measuring many characteristics of the organisms and populations along evolution.

An extension of the model (R-Aevol), integrates an explicit model of the regulation of gene expression, thus allowing for the study of the evolution of gene regulation networks.

News of the Year: In the context of the ANR Project "Evoluthon" we developed a new version of the aevol software that extends the binary code used in aevol into a 4-bases genetic code that respects the universal genetic code. Using this new version of the software, we have been able to simulate evolution along a speciation tree. The outcome of these simulations consists in 99 final genomic sequences that diverged for different amount of time. These 99 sequences have then been aligned using on-the-shelf bioinformatic software and the aligned sequences have been used to reconstruct the speciation tree. Comparison between the original simulated and the final reconstructed one shows that both trees diverge only on few branches. To the best of our knowledge, this result is the first attempt to reconstruct a phylogenetic tree from data generated by an artificial-life simulation. It simultaneously constitutes an important cross-validation step, both for the simulation software and for the inference method and opens the way to the simulation of demanding test sets using the 4-bases version of the software. This result has been published in the International Conference for Artificial Life [19]. Besides, in the context of Juliette Luiselli's PhD, an extension of the aevol model has been developed that extends the platform to eukaryotic genomes. This includes the possibility to simulate linear diploid chromosomes and sexual reproduction. Finally, a important code refactoring has been initiated to propose a unified software integrating all these elements.

URL: http://www.aevol.fr/

Contact: Guillaume Beslon

Participants: Romain Galle, Paul Banse, Guillaume Beslon, Marco Foley, Theotime Grohens, Juliette Luiselli, Jonathan Rouzaud-Cornabas, David Parsons

7.1.2 bioindication

Name: Bioindication

Keywords: Environment perception, Agroecology

Functional Description: Bioindication is a web platform designed to facilitate the reading of the landscape by users: identification of species living in a space, calculation of biodiversity indices, location, indicator values, suggestions of species or varieties to cultivate.

Release Contributions: First version

News of the Year: Bioindication is regularly used in its educational version, in two licence modules at Insa- Lyon. This involved around 20 teachers and several hundred students.

URL: http://bioindication.com

Contact: Eric Tannier

Participants: Arnaud Tilbian, David Parsons, Eric Tannier, Damien De Vienne, Jean-Sebastien Beaulne, Christophe Rigotti, Hugo Daudey, Julien Barnier, Simon Penel

7.1.3 TopShapLite

Keyword: Explainable Artificial Intelligence

Functional Description: TopShapLite is an efficient implementation of our TopShap algorithm for scikit-learn regression models. TopShap is an algorithm that computes the top-K absolute SHAP values, including possible ties, and their confidence intervals. TopShap is agnostic in the sense that it can be applied to any kind of models, and only uses the model as a black box. TopShap performs an iterative refinement of the set of top-K candidates by interleaving sampling operations, to improve SHAP value estimates, and pruning steps of the remaining candidates. The pruning is effective and leads to an important reduction of the number of calls to the model prediction function. TopShapLite is an implementation of TopShap based on NumPy vectorization and on batch-based sampling. It takes as input any scikit-learn regression model, but can also be used on any non-scikit-learn regressor object that has a method "predict" with the same signature.

TopShapLite has been used to evaluate deeply the TopShap algorithm and to compare it to the state-of-the-art alternative strategy to compute the top-K SHAP values in an agnostic setting. The evaluation shown, on a wide range of dataset geometry, that TopShap's pruning drastically reduces the search space, and led to an important gain in term of execution time. These results, and the TopShap algorithm itself, have been published in the journal IEEE Access [12]. The code to reproduce the experiments is publicly available as a Git repository at https://gitlab.inria.fr/topshap/topshap_and_experiments. TopShapLite was also the key software tools to develop and support a new method to detect rewiring events in gene regulation networks. This approach aims to detect such changes between cell types using only single-cell RNA sequencing data and has been design in the context of the PhD of Lisa Chabrier (CORDI-S PhD grant). Preliminary results have been presented in [18] and corresponding experiments are publicly available through the repository https://gitlab.inria.fr/topshap/rewiring_detection_workshop_dmbih_2024.

URL: https://gitlab.inria.fr/topshap/topshaplite_int

Contact: Christophe Rigotti

Participants: Christophe Rigotti, Sergio Peignier, Lisa Chabrier, Antonius Crombach

Partners: LIRIS, BF2I, Insa de Lyon

8 New results

8.1 Foreword

In terms of new results, 2024 has been a particular year. Indeed, it has seen the concretization of several research projects on which team members were working for several years. These results, sometimes gathered on a long period, have been submitted and accepted in renowned international scientific journals. This explains why some of the "new results" mentionned here partly overlap those presented in previous years.

8.2 Mammalian olfactory cortex retains molecular signatures of ancestral cell types

Participants: A. Crombach.

The repertoire of behavioural and cognitive abilities of mammals is thought to arise from the vast diversity of neuronal cell types and circuits of the cerebral cortex. To understand cortical circuit functions, it is thus essential to reconstruct the molecular logic driving the diversification of cell types across cortical areas. We performed single-nucleus transcriptome and chromatin accessibility analyses to compare the molecular identities of neurons across three- to six-layered cortical areas of adult mice and across tetrapod species. We found that, in contrast to the six-layered mouse neocortex, glutamatergic neurons of the three-layered mouse olfactory (piriform) cortex displayed a continuous rather than discrete variation in transcriptomic profiles. Surprisingly, subsets of glutamatergic cells with conserved transcriptomic profiles were distinguished by distinct, area-specific epigenetic states. Furthermore, we identified a prominent population of immature neurons in piriform cortex and observed that, in contrast to the neocortex, piriform cortex exhibited divergence between pyramidal cells in lab versus wild-derived mice. These results suggest a critical role for adult immature neurons in enhancing the adaptability of olfactory circuits. Finally, we showed that, unexpectedly, piriform neurons displayed marked transcriptomic similarities to cortical neurons in turtles, lizards, and salamanders. In summary, despite over 200 million years of co-evolution alongside the mammalian neocortex, olfactory cortex neurons seem to retain molecular signatures of ancestral cortical identity.

This work has been submitted to *Nature Neuroscience* in 2024. We expect publication at the very beginning of 2025.

8.3 Characterizing the fate of duplicated genes

Participants: Guillaume Beslon, Juliette Luiselli.

We initiated a collaboration with Sherbrooke University (Manuel Lafond, Reza Kalhor, Sherbrooke University, Canada) and ISEM (Céline Scornavacca, Montpellier, France). The objective of this collaboration is to use the aevol platform to simulate gene duplication in order to help characterizing the evolutionary fate of the duplicates. Indeed, although gene duplication has a central role in evolution, little is known on the fates of the duplicated copies, their relative frequency, and on how environmental conditions affect them. Moreover, the lack of rigorous definitions concerning the fate of duplicated genes hinders the development of a global vision of this process. We proposed a new theoretical framework aiming at characterizing and formally differentiating the fate of duplicated genes. This new framework has been tested via aevol simulations. Our results show several patterns to confirm previous studies and exhibit new tendencies; this opens up new avenues to better understand the role of duplications as driver of evolution.

This research has been published the *Journal of Computational Biology* in September 2024 [13]. Note that the collaboration with Sherbrooke University has been reinforced by a long stay of Juliette Luiselli in Sherbrooke from April to July 2024.

8.4 Origin of genome streamlining

Participants: Juliette Luiselli, Jonathan Rouzaud-Cornabas, Guillaume Beslon.

Genome streamlining, i.e. genome size reduction, is observed in bacteria with very different life traits —including cyanobacteria and endosymbiotic bacteria— raising the question of its evolutionary origin. None of the hypotheses proposed in the literature is firmly established, mainly due to the many

confounding factors related to the diverse habitats of streamlined species. Computational models may help overcome these difficulties and rigorously test hypotheses. We used Aevol to test two main hypotheses: increase in either population size or mutation rate. Preevolved individuals were transferred into new conditions, characterized by either a population size increase, or a mutation rate increase. Both conditions lead to streamlining. However, the increased population size and mutation rate resulted in very different genome structures. Under increased population size, genomes have lost a significant fraction of non-coding sequences, but keep their coding genome size, resulting in densely packed genomes akin to cyanobacteria genomes. On the opposite, under increased mutation rates, genomes have lost both coding and non-coding sequences, akin to endosymbiotic bacteria genomes. In both cases, genome streamlining is largely driven by structural genomic variations and is due to an increased selection for robustness to structural genomic variants. However, under increased population size, selection for robustness is secondary to selection for fitness, hence the maintenance of coding sequences, while under increased mutation rate, selection for robustness outweighs selection for fitness, resulting in a loss of both coding and non-coding sequences.

This work has been published in *Genome Biology and Evolution* in December 2024 [15].

8.5 Origin of evolutionary bursts in viruses

Participants: Paul Banse, Guillaume Beslon.

Viruses are known to evolve by bursts, which are often triggered by exogenous factors such as environmental changes, antiviral therapies or spill-overs from reservoirs into novel host species. However, other types of events have been suggested to be able to trigger evolutionary burst: either fitness valley crossing or a neutral exploration of a fitness plateau until an escape mutant is found on a neutral ridge. In order to investigate the importance of these different causes of evolutionary burst, we used aevol to perform massive evolution experiments of viral-like genomes. We tested two conditions: after an "environmental" change or in constant conditions, this latter situation guaranteeing the absence of an exogenous triggering factor. As expected, an environmental change is almost systematically followed by an evolutionary burst. However, we show that bursts also occur, although much less frequently, in constant conditions. We analyze how many of these latter bursts are triggered by deleterious, neutral or beneficial mutations and we show that while bursts can occasionally be triggered by valley crossing or traveling along neutral ridge walking, many of them were triggered by chromosomal rearrangements, and in particular segmental duplications. Our results suggest that the difference in combinatorics between the different mutation types leads to punctuated evolutionary dynamics, with long periods of stasis occasionally interrupted by short periods of rapid evolution, akin to what is observed in virus evolution.

This work has been done in collaboration with Santiago Elena (professor at the Integrative Systems Biology, Valencia, Spain). It has been published in *Virus Evolution* in September 2024 [10].

8.6 Bridging the gap between artificial life and bioinformatics

Participants: David Parsons, Eric Tannier, Romain Gallé, Jonathan Rouzaud-Cornabas, Guillaume Beslon.

Computational Evolution is a large interdisciplinary research domain that is tackeld by (at least) two different communities: on the one hand, artificial life researchers use computational systems to understand emergent evolutionary processes and patterns such as complexity, robustness, evolvability and open-endedness; on the other hand, evolutionary bioinformatics researchers decipher patterns and processes in diverse domains of life on Earth using computa- tional methods based on biological data. Both communities use simulations of living organisms but with different aims, objects, and methods, resulting in disjoint research corpuses. In the context of the Evoluthon ANR research project, we proposed Aevol 4b, an artificial life evolution simulator, and show that the data it produces can be successfully and interestingly processed using bioinformatics methods: Using this model, we generated a large species

tree and showed that the tree can be recovered, at least partly, by two different bioinformatic pipelines without modifying them in any way. To the best of our knowledge, this is the first successful coupling of artificial life simulations with bioinformatic analysis, and we believe this result opens up significant opportunities for future exchanges between these two fields.

This work has been presented at the International Conference on Artificial Life in July 2024 in Copenhagen [19].

8.7 Mathematical modeling the evolution of non-coding sequences

Participants: Juliette Luiselli, Paul Banse, Jonathan Rouzaud-Cornabas, Guillaume Beslon.

Non-coding genome size evolution is poorly understood. While some fraction of non-coding DNA has arguably a regulatory function, a large part does not seem to have a detectable impact on any phenotypic trait. The abundance of non-functional DNA in genomes, observed across the Tree of Life, challenges a purely adaptationist explanation. Several non-adaptive theories have been proposed to explain its presence and identify its determinants, emphasizing either the mutational processes or the mutational hazard entailed by non-coding and non-functional DNA. However, those theories have not yet been integrated into a single framework, and the exact nature of the mutational hazard is not yet fully understood. We have developped a simple mathematical model of genome size evolution. The model shows how the non-coding fraction of the genome is shaped by two factors: unavoidable biases in the neutrality of the different mutation types (adding base pairs is more likely to be neutral than removing some), and the robustness selection imposed by the mere existence of structural mutations (larger genomes are more prone to double-strand breaks that can initiate structural mutations, imposing a second-order selection on robustness). Together, these two factors ensure the existence of an equilibrium non-coding fraction. We show that this equilibrium depends solely on mutation biases and the product of population size and mutation rate. This work has been presented in several seminars and conferences and a publication is now ready to be submitted.

9 Partnerships and cooperations

9.1 International initiatives

As participant on the NIH R01 grant "Gene regulatory network control of olfactory cortex cell type specification" (2023 – 2028), Anton Crombach was awarded funding for a PhD student. This is a 5 year grant led by Alexander Fleischmann (Brown University, USA) with 2 partners, Ritambhara Singh (Brown University) and Anton Crombach. Total amount funded: $500k.

Participants: Antonius Crombach.

9.2 International research visitors

9.2.1 Visits of international scientists

Participants: Ruggero Pensa.

Status Researcher

Institution of origin: University of Torino

Country: Italy

Dates: March-April 2024 (one month)

Context of the visit: During his sabbatical from the university of Torino, Ruggero Pensa visited us twice (one month in 2023, one month in 2024) to develop co-clusturing approaches in the context of Lisa Chabrier's PhD.

Mobility program/type of mobility: sabbatical

Participants: Sandro Colizzi.

Status Researcher

Institution of origin: Cambridge University

Country: UK

Dates: Junuary 2024

Context of the visit: Sandro Colizzi visited the team to present his work and initiate collaborations with a view to his application as CRCN Inria (Sandro has been hired CRCN in june 2024 and will join the team in February 2025).

Mobility program/type of mobility: Research stay

Participants: Sofia Pacheco-Garcia.

Status intern (master/eng)

Institution of origin: University of Padua

Country: Italy

Dates: March-August 2024 (6 months)

Context of the visit: Studying at Padova University (Italy), Sofia Pacheco-Garcia joined the team for half a year to do her Master's thesis project. She studied how to couple a neuronal network model with a gene regulatory network model. To this end, she formulated several options, implemented them, and evaluated the behaviour of these mathematical models in the case of habituation, the simplest form of neuronal learning. She graduated in early September 2024.

Mobility program/type of mobility: internship

Participants: Nils Ledin.

Status intern (master/eng)

Institution of origin: Chalmers University

Country: Sweden

Dates: October 2024-January 2025 (4 months)

Context of the visit: A sweden student engaged in the UNITECH exchange program, Nils joined the beagle team for a master's internship, to learn about computational modeling in biology and evolution. More specifically, he has worked on the development and analysis of a variant of Fisher's geometric model to enable the study of gene duplication in this model.

Mobility program/type of mobility: internship

9.2.2 Visits to international teams

Participants: Juliette Luiselli.

Visited institution: Sherbrook University

Country: Canada

Dates: April-July 2024

Context of the visit: Beagle initiated a collaboration with Manuel Lafond (Univ. Sheerbrooke) in 2023 to study the fate of duplicated genes. To foster the collaboration and initiate new research avenues, Juliette Luiselly visited Manuel Lafond for a 3 months visit. She also took advantage of the visit to take part in two international conferences (Evolution and International Conference on Computational Biology) in Montreal in July 2024. This long stay, combining research visits and participation in international conferences on the same transatlantic flight, exemplifies the Beagle team's policy of reducing its carbon footprint. Indeed, our aim is to find – wherever possible – an effective compromise between maintaining international exchanges, training young researchers and reducing greenhouse gas emissions.

Mobility program/type of mobility: Research stay

9.3 National initiatives

9.3.1 ANR Evoluthon

Summary: The ANR project Evoluthon (Artificial Life as a Testbed for Molecular Evolution, 2019-2025) is coordinated by Eric Tannier. It involves the following partners: Inria Lyon (Beagle team) and Laboratoire de Biométrie et Biologie Évolutive (LBBE, UMR CNRS 5558). The aim of the project is to adapt the Aevol program, initially developed for artificial life research, to study molecular evolutionary processes.

Participants: Eric Tannier(project leader) , Guillaume Beslon, Jonathan Rouzaud-Cornabas. .

9.3.2 ANR Flores

Summary: The ANR project Flores (2024-2025) is a collaboration with Inria Rennes (lead by Simon Castellan), University Paris-Saclay and Tela Botanica, a botanical network association. It corresponds to the call "with and for society" of the ANR (sciences avec et pour la société) and its aim is to explore the relations between humans, technology and nature through a plant identification tool that develops the "attention" of the user to the plant morphology.

Participants: Eric Tannier. .

9.3.3 ANR Flowers

Summary: The ANR project Flowers (2024-2027) is a collaboration between the ecobio lab in Rennes (lead by Sylvain Glémin), the Isem in Montpellier. Its aim is to understand the evolutionary success of some flowering plants by ecological, morphologcial, of functional characteristics.

Participants: Eric Tannier. .

9.3.4 ANR NeGA

Summary: TThe ANR project NeGA (Influence of Effective Population Size on Animal Genome Architecture), running from 2020 to 2024, aims to empirically test Michael Lynch's hypothesis that non-adaptive forces drive the complexity of eukaryotic genomes. By comparing closely related animal species with varying effective population sizes (Ne), the project investigates how Ne influences genome size evolution, transposable element dynamics, gene structure (including intron number and size), and transcription complexity (such as alternative transcript diversity). This research involves generating genomic resources through advanced sequencing technologies, annotating transposable elements, estimating polymorphism patterns, reconstructing historical Ne, and the development of computational models. The project involves the following partners: Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés (LEHNA, UMR CNRS 5023), Laboratoire de Biométrie et Biologie Évolutive (LBBE, UMR CNRS 5558), Institut des Sciences de l'Évolution de Montpellier (UMR CNRS 5554) and the Inria Beagle Team.

Participants: Guillaume Beslon, Jonathan Rouzaud-Cornabas, Juliette Luiselli. .

9.4 PEPR Digital Agro-ecology, flagship "Coeditag"

.

Coeditag is a social science project on the dynamics and consequences of the usage of digital technology in the agricultural world. It is lead by Pierre Gasselin in INRAE Montpellier, and we participate with a co-supervised Ph-D student starting in the Semis team in 2024.

Participants: Eric Tannier. .

9.5 PEPR Digital Agro-ecology, flagship "Cobreeding"

Cobreeding is an agronomic project exploring the genetic bases of the diversity of breeding animals and crops. It is lead by Florence Phocas, at INRAE Paris, and we participate as the Inria partner, with a project on "envirotyping" related to bioindication.com.

Participants: Eric Tannier. .

9.6 PEPR Santé Numerique, Flagship "AI4scMed"

In the context of the flagship MultiScale AI for SingleCell-Based Precision Medicine (AI4scMed), we lead a project to develop methodology to analyze single-cell data.

Participants: Antonius Crombach, Thibaut Peyric. .

9.7 Action Exploratoire ExODE

In biology, the vast majority of systems can be modeled as ordinary differential equations (ODEs). Modeling more finely biological objects leads to increase the number of equations. Simulating ever larger systems also leads to increasing the number of equations. Therefore, we observe a large increase in the size of the ODE systems to be solved. A major lock is the limitation of ODE numerical resolution so ware (ODE solver) to a few thousand equations due to prohibitive calculation time. The AEx ExODE tackles this lock via 1) the introduction of new numerical methods that will take advantage of the mixed precision

that mixes several floating number precisions within numerical methods, 2) the adaptation of these new methods for next generation highly hierarchical and heterogeneous computers composed of a large number of CPUs and GPUs. For the past year, a new approach to Deep Learning has been proposed to replace the Recurrent Neural Network (RNN) with ODE systems. The numerical and parallel methods of ExODE will be evaluated and adapted in this framework in order to improve the performance and accuracy of these new approaches.

Participants: Jonathan Rouzaud-Cornabas, Mouhamad Al-Sayed Ali, Arsene Marzorati.

9.8 Other National Initiatives

The "Institut National du Cancer" funds the project CLAIRE, a collaboration of Anton Crombach with Sandra Ortiz-Cuaran (head), Virginie Marcel, and Gabriel Ichim from the Cancer Research Centre of Lyon (CRCL) / Centre Léon Bérard (CLB). This is a three-year grant of 526 k€, including a postdoc position for the Beagle team. Duration: November 2022 – November 2025.

Participants: Antonius Crombach, Hamza Chegraoui. .

9.9 Regional initiatives

9.9.1 Fédération Informatique de Lyon (FIL)

Guillaume Beslon participate to a collaborative project granted by the FIL. EvoluNet aims at fostering a collaboration with Emmanuel Roux (CREATIS) to develop algorithms allowing to evolve simultaneously the weigths and the architecture of deep neural networks. 10keuros.

Participants: Guillaume Beslon. .

9.9.2 Institut Rhône-Alpin des Systèmes Complexes (IXXI)

Guillaume Beslon participate to a collaborative project granted by IXXI. This project aims at fostering a collaboration with Nicolas Lartillot (CREATIS) to study the evolution of genome architecture. 5keuros.

Participants: Guillaume Beslon, Juliette Luiselli. .

9.9.3 Shapemed

Eric Tannier is one of the two leaders of a project "planetary health" with Anne-Laure Fougère, from university of Lyon 1, funded by "Shapemed" (university of Lyon 1). 150keuros.

Participants: Eric Tannier. .

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

Member of the organizing committees

Eric Tannier co-organized the IXXI days on November 21th, 2024, with the general theme "futures".

Eric Tannier was a member of the organizing committee of the "journées des sciences engagées et reliées", Lyon.

10.1.2 Scientific events: selection

Eric Tannier was a reviewer for ISMB 2024.

Guillaume Beslon was a reviewer for ALife 2024.

10.1.3 Journal

Member of the editorial boards

Eric Tannier is a member of the editorial board of "PCI Evolutionary Biology" and "PCI Mathematical and Computational Biology", and was the editor of several articles in 2024.

Reviewer - reviewing activities

Antonius Crombach reviewed for Developmental Biology and PLoS ONE.

Guillaume Beslon reviewed for Artificial Life

10.1.4 Invited talks

Carole Knibbe gave an invited seminar at the international workshop Matidays 2024 – Mathematical models of lipid transport and storage ("A quantitative model of long-chain fatty acid intestinal uptake, or quantitative modelling as a tool to integrate decades of heterogeneous experimental data"). Paris, November 2024.

Guillaume Beslon gave an invited seminar to the international TORC Network ("Emergence of Supercoiling-Mediated Regulatory Networks through Bacterial Chromosome Organization"), Online seminar, January 2024

Guillaume Beslon gave an invited seminar at the TIMC-IMAG Lab ("Computational and mathematical evidence of spontaneous regulation of non-coding sequences in genomes"), Grenoble, May 2024

Antonius Crombach was invited to talk about "Molecular characterizations of the mouse olfactory system" at SBRI, Lyon, February 2024

Eric Tannier, "Rebound effects in bioinformatics", seminar organized by the students of the biochemistry master of the university of Lyon

Eric Tannier, "Rebound effects in bioinformatics", IRISA, Rennes, 2024

Eric Tannier, "Vies des ateliers sciences environnements sociétés", journées scientifiques Inria 2024

Eric Tannier, "Se réapproprier la production de connaissance", conférence Archipel 2024

10.1.5 Leadership within the scientific community

An article "Un atelier pour donner du sens" was published in the online media "L'âge de faire" at the occasion of a theme "d'autres voies pour la recherche", on the science environment societies workshops that we contributed to organize.

10.1.6 Research administration

- Eric Tannier was an elected member of the administration council of Inria up to october 2024
- Eric Tannier is a member of the "FSS" Formation spécialisée de site of the Lyon Inria Research Center, from 2024
- Eric Tannier is a member of the scientific council of IXXI, Complex system institute, Lyon
- Eric Tannier is a member of the scientific council of the Lyon science shop.
- Guillaume Beslon, Juliette Luiselli and Lisa Chabrier are member of the Comité de Centre of the INRIA Center of Lyon
- Guillaume Beslon is a member of the COMI (Comité des Moyens Incitatifs) of the INRIA Center of Lyon
- Jonathan Rouzaud-Cornabas is the president of the CUMI (Comité des Utilisateurs des Moyens Informatiques) of the INRIA Center of Lyon
- Jonathan Rouzaud-Cornabas is the representive for the Inria Center of Lyon in the CUMC (Comité des Utilisateurs des Moyens Calcul) at INRIA

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

- License: Jonathan Rouzaud-Cornabas, Computer Architecture, 100h, L3, Computer Science Department, INSA Lyon
- Master: Jonathan Rouzaud-Cornabas, High Performance Computing, 60h, M2, Computer Science Department, INSA Lyon
- Master: Jonathan Rouzaud-Cornabas, High Performance Computing, 40h, M2, Biosciences Department, INSA Lyon
- Master: C.Knibbe,"Why use modelling in nutrition research", 2h CM, M2, master "Cardiovascular, metabolic and nutritional regulations" of Lyon 1 University.
- Licence: C.Knibbe, Algorithmics and Python programming, 48 h eqTD, L3,Bioinformatics and Modelling program of INSA-Lyon
- Licence: C.Knibbe, Introduction to automatic data processing, 16 h eqTD, L3, Biosciences program of INSA-Lyon
- Master: C.Knibbe,Careers in bioinformatics and modelling, 20 h eqTD, M1, Bioinformatics and Modelling program of INSA-Lyon
- License: David Parsons, Linux - Local and Remote, 22h, L3, Biosciences Department, INSA Lyon
- Master : David Parsons, Software Development, 36h, M1, Biosciences Department, INSA Lyon
- Licence: Christophe Rigotti, Object-Oriented Programming and Graphical User Interfaces, 86h, L2, Department 1er cycle of INSA-Lyon.

- Licence: Christophe Rigotti, Simulation of Chemical Reactions, 26h, L2, Department 1er cycle of INSA-Lyon.
- Licence: Christophe Rigotti, Numerical Modelling for Engineering, 60h, L2, Department 1er cycle of INSA-Lyon.
- Master: Christophe Rigotti, Data Mining, 55h, M1, Bioinformatics and Modeling Department, and Civil Engineering Department of INSA-Lyon.
- Master: Eric Tannier, Research Ethics, 8h, M2,Bioinformatics UCBL
- Doctorat: Eric Tannier, Research Ethics, 8h, University of Lyon.
- Licence: Guillaume Beslon, Computer Architecture, 100h, L3, Computer Science Department, INSA-Lyon
- Master: Guillaume Beslon, Computational Science, 25h, M2, Computer Science Department, INSA-Lyon
- Licence: Guillaume Beslon, Stage Lighting, 25h, L2, Humanities Department, INSA-Lyon
- Licence: Juliette Luiselli, Introduction to automatic data analysis, 32h, L3, Biosciences Department, INSA-Lyon
- Licence: Thibaut Peyric, Introduction to automatic data analysis, 16h, L3, Biosciences Department, INSA-Lyon
- **E-learning**
  - MOOC: Eric Tannier, member of the pedagogical team of the Research Ethics MOOC, FUN, released 2018, still online, Ph-D candidates, 3000 registered participants at each session.
  - Online ethic courses: Eric Tannier, 2 videos on research ethics on vimeo, uploaded in 2020 to diversify distant courses.

10.2.2 Supervision

- PhD in progress (CORDI-S PhD grant): Lisa Chabrier, “Efficient approximation method for local explanation of machine learning models, applied to the inference of local activity of gene regulatory networks”, supervised by Anton Crombach, Christophe Rigotti and Sergio Peignier (BF2I UMR203 - Biologie Fonctionnelle Insectes et Interactions), started October 2021. Manuscript under reviewer. Defense planned 9th April 2025.
- PhD in progress (PEPR Santé Numérique): Thibaut Peyric, “Single-cell multi-omics data integration for gene regulatory network inference”, supervised by Anton Crombach and Thomas Guyet (AIstrosight), started November 2023.
- Postdoc in progress (Institut du Cancer, project CLAIRE): Hamza Chegraoui, “Molecular mechanisms of cancer cell adaptation to targeted therapies: novel insights into the biology of drug-tolerance.”, supervised by Anton Crombach and Sandra Ortiz-Cuaran (CRCL), started November 2023.
- PhD in progress (NIH R01 grant): Sofia Pacheco-Garcia, "Odors and gene regulation" , supervised by Anton Crombach and Guillaume Beslon, started November 2024.
- PhD in progress (Allocation Doctorale ENS Ulm): Juliette Luiselli, "How chromosomal rearrangements shape genomes : a computational and mathematical modelling study", supervised by Guillaume Beslon and Nicolas Lartillot (LBBE), started October 2022

- PhD in progress (Inria PhD grant): Romain Gallé, "Impact et évolution des modèles de programmation et des systèmes d'exécution pour les machines à mémoire hétérogène : application à la biologie computationnelle.", supervised by Thierry Gauthier (LIP UMR 5668 - Inria Avalon) and Jonathan Rouzaud-Cornabas, started October 2023.
- PhD in progress (AEx ExODE PhD grant): Arsene Marzorati, "Scaling the solving of Ordinary Differential Equation for Computational Biology (and Deep Learning)"", supervised by Samuel Bernard (ICJ UMR 5208 - Inria Musics) and Jonathan Rouzaud-Cornabas, started October 2022.
- Postdoc in progress: Mouhama Al Sayed Ali (AEx ExODE senior engineer), "Scaling the solving of Ordinary Differential Equation for Computational Biology (and Deep Learning)", supervised by Samuel Bernard (ICJ UMR 5208 – Inria Dracula) and Jonathan Rouzaud-Cornabas, started July 2023.

10.3 Popularization

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

Eric Tannier is the scientific advisor of the construction of a serious game on the environmental footprint of digital industry dedicated to high school pupils, with the association "arbre de la connaissance"

10.3.2 Internships for young pupils

The Beagle team is accustomed to welcoming young students (3ieme and seconde) wishing to discover the world of research. This year, we welcomed two teenagers for short stays dedicated to research discovery:

- Eloise Coursin, a first year middle-school student, for her end-of-year internship. During this two weeks internship, she led a research project on the simulation of genome evolution.
- Younes Berkane, for his end of collège internship (one week) in the context of ZEP actions.

10.3.3 Participation in Live events

- Guillaume Beslon gave two lectures in the "Université Ouverte de Lyon". April 2024 ("Que peut-on apprendre d'une épidémie en 25 lignes de code ?") ; May 2024 ("L'évolu(on : hasard ou nécessité ? La « réponse » de la simula2on informa2que")
- Guillaume Beslon gave an interdisciplinary lecture with Dominique Schneider at the Université Grenoble Alpes during the Citizen-Campus initiative (February 2024)
- Eric Tannier, "Les rôles de la recherche scientifique dans l'anthropocène", université ouverte de Lyon, 2024
- Juliette Luiselli, scientific speed dating with bachelor female students, partnership with INRIA and ENS Lyon, journée Filles & info-maths, November 2024
- Juliette Luiselli, scientific speed-dating in the science festival "les Échappées inattendues" organized by the CNRS, November 2024
- Juliette Luiselli, "scientific carreers for womens", Lycée Condorcet, St-Priest with the association "Femmes & Sciences", December 2024

10.3.4 Others science outreach relevant activities

Eric Tannier participated to the organization of a round table on the social engagement of academic researchers in collaboration with the Lyon Science Shop and the "Fabrique des questions simples" (June 2024) link

11 Scientific production

11.1 Major publications

[1] P. Banse, S. F. Elena and G. Beslon. 'Innovation in viruses: fitness valley crossing, neutral landscapes, or just duplications?' In: *Virus Evolution* 10 (20th Sept. 2024). DOI: 10.1093/ve/veae078. URL: https://hal.science/hal-04801559.

[2] P. Banse, J. Luiselli, D. P. Parsons, T. Grohens, M. Foley, L. Trujillo, J. Rouzaud-cornabas, C. Knibbe and G. Beslon. 'Forward-in-time simulation of chromosomal rearrangements: The invisible backbone that sustains long-term adaptation'. In: *Molecular Ecology* (11th Dec. 2023). DOI: 10.1111/mec.17234. URL: https://hal.science/hal-04350147.

[3] A. Davín, E. Tannier, T. A. Williams, B. Boussau, V. Daubin and G. Szöllősi. 'Gene transfers can date the tree of life'. In: *Nature Ecology & Evolution* 2.5 (May 2018), pp. 904–909. DOI: 10.1038/s41559-018-0525-3. URL: https://hal.science/hal-01913812.

[4] J. Lehman, J. Clune, D. Misevic, C. Adami, J. Beaulieu, P. J. Bentley, S. Bernard, G. Beslon, D. M. Bryson, N. Cheney, A. Cully, S. Doncieux, F. C. Dyer, K. O. Ellefsen, R. Feldt, S. Fischer, S. Forrest, A. Frenoy, C. Gagneé, L. Le Goff, L. M. Grabowski, B. Hodjat, L. Keller, C. Knibbe, P. Krcah, R. E. Lenski, H. Lipson, R. MacCurdy, C. Maestre, R. Miikkulainen, S. Mitri, D. E. Moriarty, J.-B. Mouret, A. D. Nguyen, C. Ofria, M. Parizeau, D. Parsons, R. T. Pennock, W. F. Punch, T. S. Ray, M. Schoenauer, E. Shulte, K. Sims, K. O. Stanley, F. Taddei, D. Tarapore, S. Thibault, W. Weimer, R. Watson and J. Yosinksi. 'The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities'. In: *Artificial Life* 26.2 (June 2020), pp. 274–306. DOI: 10.1162/artl_a_00319. URL: https://hal.inria.fr/hal-01735473.

[5] Q. Li, C. Hagberg, H. Silva Cascales, S. Lang, M. Hyvönen, F. Salehzadeh, P. Chen, I. Alexandersson, E. Terezaki, M. Harms, M. Kutschke, N. Arifen, N. Krämer, M. Aouadi, C. Knibbe, J. Boucher, A. Thorell and K. Spalding. 'Obesity and hyperinsulinemia drive adipocytes to activate a cell cycle program and senesce'. In: *Nature Medicine* 27.11 (Nov. 2021), pp. 1941–1953. DOI: 10.1038/s41591-021-01501-8. URL: https://hal.inria.fr/hal-03479060.

[6] C. Rocabert, G. Beslon, C. Knibbe and S. Bernard. 'Phenotypic noise and the cost of complexity'. In: *Evolution - International Journal of Organic Evolution* (Aug. 2020). DOI: 10.1111/evo.14083. URL: https://hal.archives-ouvertes.fr/hal-02920356.

[7] T. Tricou, E. Tannier and D. M. de Vienne. 'Ghost lineages can invalidate or even reverse findings regarding gene flow'. In: *Plos Biology* 20.9 (14th Sept. 2022), e3001776. DOI: 10.1371/journal.pbio.3001776. URL: https://hal.science/hal-03781025.

[8] S. Zeppilli, T. Ackels, R. Attey, N. Klimpert, K. Ritola, S. Boeing, A. Crombach, A. Schaefer and A. Fleischmann. 'Molecular characterization of projection neuron subtypes in the mouse olfactory bulb'. In: *eLife* 10 (22nd July 2021). DOI: 10.7554/eLife.65445. URL: https://hal.science/hal-03473445.

11.2 Publications of the year

International journals

[9] C. Alamichel, J. Calvo, E. Hingant, S. Latrach, N. Quiblier and R. Yvinec. 'Modeling compartmentalization within intracellular signaling pathway'. In: *ESAIM: Proceedings and Surveys* 77 (18th Nov. 2024), pp. 100–122. DOI: 10.1051/proc/202477100. URL: https://hal.science/hal-04098543.

[10] P. Banse, S. F. Elena and G. Beslon. 'Innovation in viruses: fitness valley crossing, neutral landscapes, or just duplications?' In: *Virus Evolution* 10 (20th Sept. 2024). DOI: 10.1093/ve/veae078. URL: https://hal.science/hal-04801559 (cit. on p. 10).

[11] L. Blum Moyse and H. Berry. 'A Coupled Neural Field Model for the Standard Consolidation Theory'. In: *Journal of Theoretical Biology* 588 (2024), p. 111818. URL: https://inria.hal.science/hal-04527872.

[12] L. Chabrier, A. Crombach, S. Peignier and C. Rigotti. 'Effective pruning for top-k feature search on the basis of SHAP values'. In: *IEEE Access* 12 (2024), pp. 163079–163092. DOI: 10.1109/ACCESS.2024.3489958. URL: https://hal.science/hal-04549416 (cit. on p. 8).

[13] R. Kalhor, G. Beslon, M. Lafond and C. Scornavacca. 'A Rigorous Framework to Classify the Post-Duplication Fate of Paralogous Genes (extended version)'. In: *Journal of Computational Biology* (2024). URL: https://hal.science/hal-04608855. In press (cit. on p. 9).

[14] J. Luiselli, I. Overcast, A. Rominger, M. Ruffley, H. Morlon and J. Rosindell. 'Detecting the ecological footprint of selection'. In: *PLoS ONE* 19.6 (7th June 2024), e0302794. DOI: 10.1371/journal.pone.0302794. URL: https://hal.science/hal-04753991.

[15] J. Luiselli, J. Rouzaud-Cornabas, N. Lartillot and G. Beslon. 'Genome Streamlining: Effect of Mutation Rate and Population Size on Genome Size Reduction'. In: *Genome Biology and Evolution* 16 (14th Nov. 2024). DOI: 10.1093/gbe/evae250. URL: https://hal.science/hal-04905734 (cit. on p. 10).

[16] C. Rocabert, S. Fenet, B. Kaufmann and J. Gippet. 'Accounting for the topology of road networks to better explain human-mediated dispersal in terrestrial landscapes'. In: *Ecography* 2024.3 (Mar. 2024), e07068:1–13. DOI: 10.1111/ecog.07068. URL: https://hal.science/hal-04449314.

[17] T. Tricou, E. Tannier and D. de Vienne. 'Response to "On the impact of incomplete taxon sampling on the relative timing of gene transfer events"'. In: *PLoS Biology* 22.3 (2024), e3002557. DOI: 10.1371/journal.pbio.3002557. URL: https://hal.science/hal-04515190.

International peer-reviewed conferences

[18] L. Chabrier, A. Crombach, S. Peignier and C. Rigotti. 'Detecting gene regulatory network rewiring as changes in marker regulatory associations using only single-cell gene expression data'. In: IEEE 12th Data Mining in Biomedical Informatics and Healthcare Workshop (DMBIH'24) at ICDM'24. Abu Dhabi, United Arab Emirates, 2024, p. 10. URL: https://inria.hal.science/hal-04808464 (cit. on p. 8).

[19] H. Daudey, D. P. Parsons, D. P. Parsons, E. Tannier, V. Daubin, B. Boussau, V. Liard, R. Gallé, J. Rouzaud-Cornabas and G. Beslon. 'Aevol 4b: Bridging the gap between artificial life and bioinformatics'. In: ALIFE 2024 - International Conference for Artificial Life. ALIFE 2024: Proceedings of the 2024 Artificial Life Conference. Copenhague, Denmark: MIT Press; MIT Press Direct, 2024, pp. 41–49. DOI: 10.1162/isal_a_00716. URL: https://hal.science/hal-04667220 (cit. on pp. 7, 11).

[20] R. G. Pensa, A. Crombach, S. Peignier and C. Rigotti. 'Combining SHAP-driven Co-clustering and Shallow Decision Trees to Explain XGBoost'. In: *LNAI Springer*. 27th International Conference Discovery Science (DS'24). Pise, Italy, 2024, pp. 1–16. URL: https://hal.science/hal-04809871.

[21] C. Pothier, E. Neyrinck, C. Lasserre, B. Rousset, C. Rigotti, M.-P. Doin and P. Durand. 'Exploring an unsupervised data mining approach to extract aseismic slip events along the Izmit section of the North Anatolian Fault, based on FLATSIM time series'. In: *BRGM (online)*. 6th Symposium on Measurement of Earth surface Deformations with Earth Observation (MDIS'24). Orléans, France, 2024, p. 2. URL: https://inria.hal.science/hal-04808572.

National peer-reviewed Conferences

[22] L. Chabrier, A. Crombach, S. Peignier and C. Rigotti. 'Re_actShap : détection de rebranchements des réseaux de régulation d'expression génique à l'aide des valeurs SHAP'. In: *Actes EGC'25*. 25ième conference sur l'Extraction et Gestion des Connaissances (EGC), session demonstrations. Strasbourg, France, 2025, p. 8. URL: https://inria.hal.science/hal-04877547.

Conferences without proceedings

[23] L. Chabrier, A. Crombach, S. Peignier and C. Rigotti. 'Unraveling regulatory associations using explainable AI'. In: Journée GT Machine learning pour la génomique LEGO à JOBIM 2024. Toulouse, France, 25th June 2024. URL: https://hal.science/hal-04783821.

[24] J. Luiselli. 'Effect of mutation rate and population size on genome size variation'. In: International Congress on Ecology and Evolution. Lyon, France, 23rd Oct. 2024. URL: https://hal.science/hal-04754007.

[25] J. Luiselli. 'Forward-in-time simulation of chromosomal rearrangements: The invisible backbone that sustains long-term adaptation'. In: Evolution. Montréal, Canada, 26th July 2024. URL: https://hal.science/hal-04754017.

[26] A. Marzorati, M. Al-Sayed Ali, S. Bernard and J. Rouzaud-Cornabas. 'Mixed precision and local error in ordinary differential equations'. In: Congrés National d'Analyse Numérique 2024. Ile de Ré, France, 27th May 2024. URL: https://inria.hal.science/hal-04953537.

[27] M. Al-Sayed Ali, A. Marzorati, S. Bernard and J. Rouzaud-Cornabas. 'Mixed precision numerical methods for solving large systems of ordinary differential equations'. In: Congrés National d'Analyse Numérique 2024. Ile de ré, France, 27th May 2024. URL: https://inria.hal.science/hal-04953553.

Reports & preprints

[28] N. Méger, T. Nguyen, C. Rigotti, C. Pothier and E. Trouvé. *Information gain-based selection of sequential patterns extracted from partial unimodal probabilistic bases of sequences*. 2024-1. Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC) - Polytech Annecy-Chambéry, 20th Feb. 2024, p. 85. URL: https://hal-lara.archives-ouvertes.fr/hal-04472843.

[29] N. Quiblier, J.-M. Rye, P. Leclerc, H. Truong, A. Hannou, L. Héliot and H. Berry. *Enhancing Fluorescence Correlation Spectroscopy with machine learning to infer anomalous molecular motion*. 8th Jan. 2025. URL: https://inria.hal.science/hal-04650578.

Other scientific publications

[30] L. Chabrier, C. Leonce, R. Schneider, C. Dégletagne, P. Saintigny, A. Crombach and S. Ortiz-Cuaran. 'Regulatory network inference of drug-tolerant persister cells'. In: AACR (American Association for Cancer Research) annual meeting - poster session. San Diego (California), United States, 5th Apr. 2024. URL: https://hal.science/hal-04879042.

[31] L. Chabrier, C. Leonce, R. Schneider, C. Dégletagne, P. Saintigny, A. Crombach and S. Ortiz-Cuaran. 'Regulatory network inference of lung adenocarcinoma drug-tolerant persister cells'. In: Computational Biology of Cancer – BioSyL 2024 (COBICA) - poster session. Grenoble, France, 18th Mar. 2024. URL: https://hal.science/hal-04879089.

[32] T. Peyric, A. Crombach and T. Guyet. 'Single-cell multi-omics data integration powered by PCA-like autoencoders'. In: JOBIM 2024. Toulouse, France, 25th June 2024. URL: https://hal.science/hal-04750060.