RESEARCH CENTRE

Inria Centre at the University of Bordeaux

IN PARTNERSHIP WITH:

CNRS, INRAE

2024

ACTIVITY REPORT

Project-Team

PLEIADE

Patterns of diversity and networks of function

IN COLLABORATION WITH: Laboratoire Bordelais de Recherche en Informatique (LaBRI), Biodiversité, Gènes & Communautés (BioGeCo)

DOMAIN

Digital Health, Biology and Earth

THEME

Computational Biology

Inria

Contents

Project-Team PLEIADE

Creation of the Project-Team: 2019 March 01

Keywords

Computer sciences and digital sciences

A3.1. – Data

A3.2. – Knowledge

A3.3.2. – Data mining

A3.3.3. – Big data analysis

A3.4. – Machine learning and statistics

A6.1. – Methods in mathematical modeling

A6.2. – Scientific computing, Numerical Analysis & Optimization

A8.2. – Optimization

A9.8. – Reasoning

Other research topics and application domains

B1.1.7. – Bioinformatics

B1.1.10. – Systems and synthetic biology

B3. – Environment and planet

1 Team members, visitors, external collaborators

Research Scientists

- David Sherman [Team leader, INRIA, Senior Researcher, HDR]
- Clemence Frioux [INRIA, Researcher]
- Simon Labarthe [INRAE, Researcher, HDR]
- Guilhem Sommeria-Klein [INRIA, Researcher, from Oct 2024]

PhD Students

- Chabname Ghassemi Nedjad [UNIV BORDEAUX]
- Maxime Lecomte [INRIA, until Apr 2024]
- Coralie Muller [INRIA, from Sep 2024]
- Mathilde Sola [INRAE, from Sep 2024]
- Sthyve Tatho [INRAE]

Technical Staff

- Leonard Brindel [INRIA, Engineer]
- Jean-Marc Frigerio [INRAE, Engineer]
- Coralie Muller [INRIA, Engineer, from Jul 2024 until Aug 2024]
- Coralie Muller [INRAE, from Feb 2024 until Jun 2024]
- Coralie Muller [INRIA, Engineer, until Jan 2024]
- Franck Salin [INRAE, Engineer]

Interns and Apprentices

- Juliette Audemard [INRIA, Apprentice, from Oct 2024]
- Juliette Audemard [INRAE, Intern, from Apr 2024 until Aug 2024]
- Juliette Gronier [INRIA, Intern, from Jun 2024 until Jul 2024]

Administrative Assistant

- Catherine Cattaert Megrat [INRIA]

External Collaborator

- Alain Franc [INRAE, HDR]

2 Overall objectives

Digital microbial ecology studies communities of microorganisms in delimited ecosystems, who interact through the production and consumption of metabolic goods. These interactions define *complex behaviors* that are much more that the sum of the individual behaviors of the community members, and arise from cooperation and competition between a diversity of organisms providing a diversity of beneficial and harmful functions.

Pleiade builds methodological tools for digital microbioal ecology, that elucidate the behavior of microbial communities.

- We measure the **diversity of organisms** by comparing DNA sequences in a sampled environment and building compact geometric representations.
- We identify the **diversity of functions** performed by these organisms, and represent them with genome-scale metabolic models and reasoning-compatible knowledge bases.
- We build numerical and discrete **spatio-temporal models** of community behavior.

Pleiade develops a synergistic, iterative combination of a mechanistic **community-based** strategy for deciphering the diversity in cultures and environmental samples, through metagenomic and metabolomic analysis of functional diversity and metabarcoding analysis of taxonomic diversity; and a phenomenological **function-based** strategy for contributing to *digital twins* of natural or designed communities through numerical models.

Shared methodologies needed to scale up to the complexity of biological systems, include *high-performance computing* (HPC); *machine learning*, including clustering, meta-modeling and classification for knowledge engineering; *machine reasoning*, specifically logical and rule-based methods used for model inference and network analysis. Logicial methods in particular promote *explainable* inference, since the rules are expressed in biological terms and are auditable by biologists, independently from the combinatorial and heuristic optimization techniques used to apply the rules.

Pleiade maintains strong collaborative relations with experimental biologists, and is committed to developing applications in ecology, evolution, biotechnology, and health. Team resources are dedicated to facilitating the adoption of our research by non specialist users, through development of reusable software, integration in HPC frameworks, improvement of web-based environments, and deployment of Jupyter, Galaxy, and Kubernetes interfaces.

3 Research program

Pleiade's mutually reinforcing strengths are a stable foundation of ecology and comparative genomics, and a novel synthesis of new methods extending our reach into microbial and planktonic communities and their dynamics. Our shared aim is to develop both: new challenges in microbial or planktonic communities leverage our solid expertise in foundational methods, and at the same time define new challenges for improving those foundations. We focus on reinforcing of a first set of *disciplinary* activities related to innovations in methods, be they in applied mathematics or in computer sciences, and a second set of *interdisciplinary* activities, to build advantageous assemblies of methods for understanding and managing biological systems of interest.

3.1 Research: A Geometric View of Diversity

Data analysis algorithms and tools must be revisited and scaled up. We mobilize both distributed algorithms and new algorithms, like random projection or column selection methods, to build point clouds in Euclidean spaces from massive data sets, and thus overcome the cubic complexity of computation of eigenvectors and eigenvalues of very large dense matrices. We also link distance geometry [64] with convex optimization procedures through matrix completion [43, 45].

Intercalibration: There is a considerable difference between supervised and unsupervised clustering: in supervised clustering, the result for an item i is independent from the result for an item $j \neq i$, whereas in unsupervised clustering, the result for an item i (e.g. the cluster it belongs to, and its composition)

depends on nearby items $j \neq i$. Which means that the result may change if some items are added to or subtracted from the sample. This raises the more global problem of how to merge two studies to yield a more comprehensive view of biodiversity?

See [53] for some of our recent work linking the distance geometry problem, nonlinear mapping, and weighted least-squares scaling.

Project-team positioning

This research topic is about metabarcoding, i.e. producing inventories through classification, or producing OTUs with clustering, as elementary bricks for diversity, in a way analgous to the role that species play in morphological or molecular based taxonomy. The number of reads produced by NGS facilities is a challenge for bioinformatics and data analysis downstream into these directions: most algorithms scale with the square or the cube of the number of involved reads, which are counted in the millions. Most approaches look for heuristics permitting to produce result within a reasonable time (see e.g. UCLUST or USEARCH as wrappers for many solutions). The Pleiade team has explored another path: relying on HPC (the possibilities of which are underestimated in the community of diversity studies) to derive exact algorithms without heuristics, for computing distances between reads and to analyse distance matrices with dimension reduction.

This has been made possible due to an involvement in scientific computing, associating algorithms, their implemntation and optimisation. Such a bridge between scientific computing and characterisation of the biodiversity is original and still an exploratory field. It is develped in collaboration with project teams CONCACE (ex-HIEPACS), STORM and TADAAM in Inria Bordeaux, through two technology development actions: one, ADT Gordon, which is currently being valorized by the redaction of an article, and the second, Diodon, under development.

Collaborations

- In collaboration with the Pasteur Institute in Cayenne and the INRA MIA Research Team in Toulouse, Pleiade is developing a stochastic model for simulation of metacommunities, in the framework of *patch occupancy* models. The objective is a better understanding of zoonose propagation, namely rabies through bat hosts in connection with disturbances of pristine forests in French Guiana, which have an impact on the exposure of human populations to wildlife that act as reservoirs of zoonoses.

- We have co-supervised with Anne Lavergne (Institut Pasteur de Guyane) a PhD student at IPG and University of Cayenne (Sourakhata Tirera, defense on December 17, 2021) on the drivers of the diversity of viromes of rodents and birds. One part is about bioinformatics in order to select viral fragment, which are a minority in shotgun sequencing of viromes, and a second part is about disentangling the role of the habitat and of the phylogeny of the host in viral diversity. This has led to manuscript [73]. A second one is in preparation to make available a new pipeline for dechimerisation of contings after shotgun sequencing and de nova assembly of reads into contigs (a process which is known to create many chimera).

- The Laboratory of excellence (Labex) CEBA promotes innovation in research on tropical biodiversity. It brings together a network of internationally-recognized French research teams, contributes to university education, and encourages scientific collaboration with South American countries. Pleiade has participated in three current international projects funded by CEBA:

  - *MicroBIOMES: Microbial Biodiversities*
  - *Neutrophyl: Inferring the drivers of Neotropical diversification*
  - *Phyloguianas: Biogeography and pace of diversification in the Guiana Shield*

- We collaborate with Institut Pasteur de Guyane at Cayenne for developing the domain of so-called Ecoviromics for some zoonoses in French Guiana. On top of co-supervizing a PhD student at IPG in Cayenne as mentioned above (deciphering the respective roles of host phylogeny and environmental variables in the virome of different hosts (bats, rodents, birds), this collaboration

has led to a participation in a synthesis paper on Ecology, evolution, and epidemiology of zoonotic and vector-borne infectious diseases in French Guiana [72].

3.2 Research: Community-scale Metabolic and Omics-based Modeling

Metabolism can be abstracted into sets of metabolic reactions, associated to the genome through gene-protein relationships, and connecting substrates to products thereby forming metabolic networks. **Genome-scale metabolic networks** (GSMNs) contain all the reactions predicted to occur in a organism according to its genomic contents. Combined with additional knowledge on the system, possibly other –omics data and mathematical models, GSMNs are used to **predict the behaviour of an organism or a community** of organisms.

A widely-used mathematical formalism for modeling GSMNs is **constraint-based modeling**, among which flux balance analysis (FBA) is the main representative [66]. Such methods permit a quantitative prediction of activity fluxes in metabolic networks while optimising an objective function and assuming steady-state of the system.

The emergent metabolism of microbial communities can also be **qualitatively modeled using a boolean approximation of metabolic dynamics**[8]. In this approach the behavior of the system is described by logical rules that activate a given reaction as soon as its substrates become available; numerical parameters such as stochiometry or enzyme kinetics are ignored in favor of graph topology and paths. The advantage is that such qualitative models, unlike quantitative methods such as FBA, do not require the assumption that the system is at steady-state and can model systems where cells are constantly growing or constantly reproducing.

Network expansion, introduced in [51] as a recursive traversal of the structure of a metabolic graph, lends itself to concise definition using *answer set programming* (ASP) [58] and thus to efficient implementation using SAT solvers [57]. In practice, using ASP for metabolic modeling makes it possible to define both the activation of metabolic reactions in different conditions, and the constraints and optimizations needed to find solutions in a combinatorically large state space.

We focus in particular on the key question of determining *minimal communities*, subsets of the organisms present in an environment that are sufficient to reproduce a chosen behavior [54]. The methodological goal here is to **identify key species in a community through use of ASP** to rapidly explore the search space and thus, through heuristic resolution of combinatorial problems, provide the guarantees an exhaustive search with a greatly reduced computational cost [4].

Functional and taxonomic diversities, beyond intrinsic specificities encoded in the genetic material, are also strongly shaped by their environment. Spatial nutritional niches, microbial interactions and abiotic constraints lead to **complex spatial structures in the microbial community** that impact its overall dynamics. PDE-based models of the microbiota in its environment allow including in the model these multiple mechanisms in order to decipher their influence on the community faith.

The main methodological developments in this area are related to **mathematical modeling** (in particular the correct level of simplification in the multi-physics description of the microbial environment), **model simplification** (asymptotic approximation), **inference from multi-omics data** (including dimension reduction, statistical learning) and **numerical developments** (in particular fast approximation of metabolic models with machine learning methods). Strong interactions with community-scale metabolic models are sought, specially for multi-omics inference and knowledge-based machine learning constraints.

The goal is to achieve **accurate models of microbial communities** that could be used as *digital twins* of controlled experiments in microbial ecology. Culturomic facilities allow for the acquisition of multi-omics time-series data in controlled conditions, which can be used to build and fit population dynamics models, and can be used in turn to explore numerically biological assumptions and to help in experimental planning and data analysis.

Project-team positioning

The team has expertise in the state of the art methods for metabolic modeling and in metabolic network reconstruction through earlier works [54, 55, 41, 68]. The team also masters ODE or PDE microbial population dynamics models including their complex environment [44, 60] and parameter inference

with experimental data [40]. We work with international or national teams (see below) for combinatorial problem solving (University of Potsdam, Germany), computational biology for health (Quadram Institute Bioscience, UK) and biological applications (Roscoff Biological Station, INRAE teams).

Other international teams working on the subject of community metabolic modeling include but is not restricted to the research groups of: Ines Thiele (U. Galway, Ireland), Kiran Patil (U. Cambridge, UK), Karoline Faust (U. Leuven, Belgium), Daniel Machado (Norwegian University of Science and Technology).

Among Inria project-teams, Dyliss (Rennes) is the one with the closest research themes. Clémence Frioux did her PhD in this team (2015-2018) and stayed for an additional 6 months after the defense. As a result, the majority of her past contributions were done in collaboration with Dyliss members, and current collaborations persist through co-development and maintenance of software, applications etc.

Collaborations

- Quadram Institute Bioscience and Earlham Institute: meta-analyses of metagenomic cohorts for the human gut microbiota.
- University of Potsdam: answer set programming and combinatorial problem solving.
- Station Biologique de Roscoff: algae applications.
- INRAE BFP, STLO, MaIAGE, SAVE, Micalis, IEES: applications and methodological development in computational biology and mathematics.
- CEA. Bio-inspired digital sensors in the framework of the Pherosensor project.
- U. Besançon and U. Orléans. Modeling of the fluidic environment in the gut microbiota.
- U. Paris-Saclay/U.Evry. Machine learning with ANOVA-RKHS.
- Inria Bretagne Atlantique: systems biology and systems ecology.
- Biomathematica through a CIFRE PhD hosted at MaIAGE (INRAe) co-supervised by Pleiade.

3.3 Research: Bioinformatics, Genomes, and Knowledge Management

The heterogenous data generated in computational molecular biology and ecology are distinguished not only by their volume, but by the richness of the many levels of interpretation that biologists create. The same nucleic acid sequence can be seen as a molecule with a structure, a sequence of base pairs, a collection of genes, an allele, or a molecular fingerprint. To extract the maximum benefit from this treasure trove we must organize the knowledge in ways that facilitate extraction, analysis, and inference. Our focus has been on the efficient representation of relations between biological objects and operations on those representations, in particular heuristic analyses and logical inference.

Pleiade developes applications in comparative genomics of related organisms, using novel mathematical tools for representing compactly, at different scales of difference, comparisons between related genomes. New methods based on distance geometry will refine these comparisons. Compact representations can be stored, exchanged, and combined. They form the basis of novel simultaneous genome annotation methods, that can be linked directly to abductive inference methods for building functional models of the organisms and their communities.

Since a goal of Pleiade is to integrate diversity throughout the analysis process, it is necessary to incorporate **diversity as a form of knowledge** that can be stored in a knowledge base. Diversity can be represented using various compact representations, such as trees and quotient graphs storing nested sets of relations. Extracting structured representations and logical relations from integrated knowledge bases requires domain-specific query methods that can express forms of diversity.

Project-team positioning

Historically, Pleiade members have been pioneers in the development of large-scale eukaryote comparative genomics. We were involved since the late 1990's in the first genome sequencing of eukaryote microorganisms, co-authored 8 articles in the [59] special issue presenting the first large-scale comparative genomics study, and were in the first authors of the landmark Nature article [50] comparing five complete annotated genomes. Our articles in comparative genomics, particularly of the hemiascomycetous yeasts of biotechnological interest, have achieved thousands of citations and continue to do so decades later. A principle that we fought for [48, 70], that comparative genomics must be based on a systematic and mathematical comparison of the genomes, rather than on an opportunistic one-against-all comparison to the model organism *du jour*, is now considered standard practice. We also originally set the standard for web-based tools for comparative genomics, organized around the principle of an interaction design based on the questions asked by by biological user, rather than based on the organization of the underlying database as seen by a computer scientist[71, 70]. Pleiade capitalized on this experience in support of the research efforts describe above, and also through a wide network of collaborators in the biological sciences. Our current work applies the principles of comparative genomics to a series of smaller projects focused on collections of genomes of biotechnological or of health-related interest.

Collaborations

- Institut des Sciences du Vigne et du Vin (ISVV), U. Bordeaux
- Laboratoire de Microbiologie Fondamentale et Pathogénicité (LMFP), UMR 5234 CNRS U. Bordeaux
- Laboratory of Membrane Biology (LBM) UMR 5200 CNRS U. Bordeaux

4 Application domains

4.1 Molecular based systematics and taxonomy

Defining and recognizing the myriads of species occuring in the biosphere has been the focus of phenomenal energy over the past centuries and remains a major goal of Natural History. It is an iconic paradigm in pattern recognition (clustering has coevolved with numerical taxonomy many decades ago). Developments in evolution and molecular biology, as well as in data analysis, have over the past decades enabled a profound revolution, where species can be delimited and recognized by data analysis of sequences. We aim at proposing new tools, in the framework of E-science, which make possible (i) better exploration of the diversity in a given clade, and (ii) assignment of a place in these patterns for new, unknown organisms, using information provided by sets of sequences. This will require investment in data analysis, machine learning, and pattern recognition to deal with the volumes of data and their complexity.

One example of this project is about the diversity of trees in Amazonian forest, in collaboration with botanists in French Guiana. Protists (unicellular Eukaryots) are by far more diverse than plants, and far less known. Molecular exploration of Eukaryotes diversity is nowadays a standard in biodiversity studies. Data are available, through metagenomics, as an avalanche and make molecular diversity enter the domain of Big Data. Hence, an effort will be invested, in collaboration with other INRIA teams (GENSCALE, CONCACE) for porting to HPC algorithms of pattern recognition and machine learning, or distance geometry, for these tools to be available as well in metagenomics. This will be developed first on diatoms (unicellular algae) in collaboration with INRAE team at Thonon and University of Uppsala, on pathogens of tomato and grapewine, within an existing network, and on bacterial communities, in collaboration with University of Pau. For the latter, the studies will extend to correlations between molecular diversity and sets of traits and functions in the ecosystem.

4.2 Genome and transcriptome annotation

Sequencing genomes and transcriptomes provides a picture of how a biological system can function, or does function under a given physiological condition. Simultaneous sequencing of a group of related

organisms is now a routine procedure in biological laboratories for studying a behavior of interest, and provides a marvelous opportunity for building a comprehensive knowledge base of the relations between genomes[1, 12]. Key elements in mining these relations are: classifying the genes in related organisms and the reactions in their metabolic networks, recognizing the patterns that describe shared features, and highlighting specific differences.

Pleiade develops applications in comparative genomics of related organisms, using new mathematical tools for representing compactly, at different scales of difference, comparisons between related genomes. New methods based on computational geometry refine these comparisons. Compact representations can be stored, exchanged, and combined. They will form the basis of new simultaneous genome annotation methods, linked directly to abductive inference methods for building functional models of the organisms and their communities.

Our ambition in biotechnology is to permit the design of synthetic or genetically selected organisms at an abstract level, and guide the modification or assembly of a new genome. Our effort is focused on two main applications: genetic engineering and synthetic biology of oil-producing organisms (biofuels in CAER, palm oils), and improving and selecting starter microorganisms used in winemaking (collaboration with the ISVV and the BioLaffort company).

4.3 Community ecology

Community assembly models how species can assemble or diassemble to build stable or metastable communities. It has grown out of inventories of countable organisms. Using *metagenomics* one can produce molecular based inventories at rates never reached before. Most communities can be understood as pathways of carbon exchange, mostly in the form of sugar, between species. Even a plant cannot exist without carbon exchange with its rhizosphere. Two main routes for carbon exchange have been recognized: predation and parasitism. In predation, interactions–even if sometimes dramatic–may be loose and infrequent, whereas parasitism requires what Claude Combes has called intimate and sustainable interactions [47]. About one decade ago, some works [69] have proposed a comprehensive framework to link the studies of biodiversity with community assembly. This is still incipient research, connecting community ecology and biogeography.

We aim at developping graph-based models of co-occurence between species from NGS inventories in metagenomics, i.e. recognition of patterns in community assembly, and as a further layer to study links, if any, between diversity at different scales and community assemblies, starting from current, but oversimplified theories, where species assemble from a regional pool either randomly, as in neutral models, or by environmental filtering, as in niche modeling. We propose to study community assembly as a multiscale process between nested pools, both in tree communities in Amazonia, and diatom communities in freshwaters. This will be a step towards community genomics, which adds an ecological flavour to metagenomics.

Next-generation sequencing technologies are now an essential tool in population and community genomics, either for making evolutionary inferences or for developing SNPs for population genotyping analyses. Two problems are highlighted in the literature related to the use of those technologies for population genomics: variable sequence coverage and higher sequencing error in comparison to the Sanger sequencing technology. Methods are developed to develop unbiased estimates of key parameters, especially integrating sequencing errors [65]. An additional problem can be created when sequences are mapped on a reference sequence, either the sequenced species or an heterologous one, since paralogous genes are then considered to be the same physical position, creating a false signal of diversity [56]. Several approaches were proposed to correct for paralogy, either by working directly on the sequences issued from mapped reads [56] or by filtering detected SNPs. Finally, an increasingly popular method (RADseq) is used to develop SNP markers, but it was shown that using RADseq data to estimate diversity directly biases estimates [42]. Workflows to implement statistical methods that correct for diversity biases estimates now need an implementation for biologists.

5 Social and environmental responsibility

5.1 Footprint of research activities

Pleiade's policy is to rely on shared computing platforms for computations that consume significant energy, for two reasons. First, those platforms have greater leverage over total energy consumption, and have the technical means to implement green computing on a useful scale. Second, those platforms have staff with the requisite skills to develop and implement policies as part of their service offer. Our partner platforms are mesocenters in Bordeaux and Grenoble, and national centers including the Idris and the CEA. Some of them charge back the cost of CO_2 generation.

5.2 Promoting equality and diversity in science

Promoting the inclusion and diversity in science is essential at all levels: when planning science during project conception, when executing science and publishing its results, and also within the community of scientists itself. Members of Pleiade are involved in promoting the place of women in science through the participation in outreach activities, but also by committing to working groups and committees on the subject at the local and national level.

- Clémence Frioux – member the Inria national committee for equality and inclusion
- Clémence Frioux, Coralie Muller – members of the gender equality and diversity working group in the Inria Centre at the University of Bordeaux
- Clémence Frioux – participation to "MIMM, moi informaticienne, moi mathématicienne" 2023, a free internship at the University of Bordeaux for young girls in 9th and 10th grade in order to encourage them to choose mathematics and computer science, allowing them to discover training, research and jobs in these two disciplines.

6 Highlights of the year

Highlights for team members:

- Simon Labarthe successfully defended his HDR *Contributions to mathematical modelling in microbial ecology, from data to models* [20], and was promoted to research director effective January 2025.
- Guilhem Sommeria-Klein was recruted as an Inria research scientist and joined the team in October.
- Maxime Lecomte successfully defended his doctoral dissertation *Hybrid approach for explainable metabolic modelling of microbial ecosystems* [21].

A major study in collaboration with INRAE STLO (Rennes) was published in *Metabolic Engineering* [14], integrating heterogenous *omics* data to build a dynamic model of cheese production and reveal the roles of each microbe in the production of flavor compounds. Complete software is available in HAL and Software Heritage [39]. Complete data are available in ENA and Recherche Data Gouv.

Our review in the highly reputed journal *Microbial Biotechnology* [13] is a conceptual work covering the complete chain of numerical methods, from metagenomic data to community-scale models.

Several meetings of the Artemis consortium led by Simon Labarthe advanced towards a generic formalization of digital twins for microbial applications. clarifying the specificities of microbiology compared to a classic process-control definition of digital twins.

Through a combination of papers, we made significant progress towards true spatiotemporal modeling of microbial communities.

7 New software, platforms, open data

7.1 New software

7.1.1 Metage2Metabo

Keywords: Metabolic networks, Microbiota, Metagenomics, Workflow

Scientific Description: Flexible pipeline for the metabolic screening of large scale microbial communities described by reference genomes or metagenome-assembled genomes. The pipeline comprises several main steps. (1) Automatic and parallel reconstruction of metabolic networks. (2) Computation of individual metabolic potentials (3) Computation of collective metabolic potential (4) Calculation of the cooperation potential described as the set of metabolites producible by species only in a cooperative context (5) Computation of minimal-sized communities sastifying a metabolic objective (6) Extraction of key species (essential and alternative symbionts) associated to a metabolic function

Functional Description: Metabolic networks are graphs which nodes are compounds and edges are biochemical reactions. To study the metabolic capabilities of microbiota, Metage2Metabo uses multiprocessing to reconstruct metabolic networks at large-scale. The individual and collective metabolic capabilities (number of compounds producible) are computed and compared. From these comparisons, a set of compounds only producible by the community is created. These newly producible compounds are used to find minimal communities that can produce them. From these communities, the keytstone species in the production of these compounds are identified.

URL: https://github.com/AuReMe/metage2metabo

Publication: hal-02395024

Contact: Clemence Frioux

Participants: Clemence Frioux, Arnaud Belcour, Anne Siegel

7.1.2 MiSCoTo

Name: Microbiota Screening and COmmunity Selection with TOpology

Keywords: Metabolic networks, ASP - Answer Set Programming, Logic programming

Scientific Description: MiSCoTo solves combinatorial problems using Answer Set Programming. It aims at minimizing either the number of selected species or both the number of selected species and the cost of the interaction between them, characterized by the number of metabolic exchanges. In the first case, the level of modeling is called lumped or mixed-bag, in the latter, it is compartmentalized.

Functional Description: Metabolic networks are composed of biochemical reactions and gather the expected metabolic capabilities of species. For organisms that live in interaction altogether (microbiotas), complementarity between these networks can be exploited to predict cooperation events. This software takes as inputs metabolic networks for various species (host, symbionts of the microbiota), components of the growth medium and a metabolic objective (metabolites to be produced), and aims at selecting a minimal set of symbionts to ensure the metabolic objective can be achieved. The software can use two types of modelings: a simplified one and another that takes into account the cost of metabolic exchanges and aims at minimizing it.

Release Contributions: Memory usage optimization. Fix issues with input file formats.

URL: https://github.com/cfrioux/miscoto

Publication: hal-01871600

Contact: Clemence Frioux

Participants: Clemence Frioux, Anne Siegel, Enora Fremy, Camille Trottier, Arnaud Belcour

7.1.3 MeneTools

Name: Metabolic networks Topological tools

Keywords: Metabolic networks, Graph, Topology, Bioinformatics, Systems Biology, ASP - Answer Set Programming

Scientific Description: MeneTools are a set of tools for the exploration of the producibility potential in a metabolic network using the network expansion algorithm. The MeneTools can: - assess whether targets are producible starting from nutrients (Menecheck) - get all compounds that are producible starting from nutrients (Menescope) - get all reactions that are activable from nutrients (Meneacti) - get production paths of specific compounds (Menepath) - obtain compounds that if added to the nutrients, would ensure the producibility of targets (Menecof) - identify metabolic deadends, i.e. metabolites that act as reactants of reactions but never as products, or metabolites that act as products of reactions but never as reactants. This is a purely structural analysis. All MeneTools using modelling follow the producibility in metabolic networks as defined by the network expansion algorithm.

Functional Description: MeneTools consist in four topological tool to analyze metabolic models in a graph-based perspective. Menecheck verifies the producibility of target compounds from available substrates (growth medium) of the metabolic network. Menescope gives the whole range of accessible compounds in the metabolic network starting from substrates. Menepath give the production paths of given compounds in the model. Menecof proposes compounds that need to be produced or added as substrate for ensuring the producibility of targets.

URL: https://github.com/cfrioux/MeneTools

Publications: hal-01819150, hal-02395024

Contact: Clemence Frioux

Participants: Clemence Frioux, Anne Siegel, Arnaud Belcour

7.1.4 Emapper2GBK

Keywords: Bioinformatics, Metabolic networks, Functional annotation

Functional Description: Starting from FASTA and Eggnog-mapper annotation files, Emapper2GBK builds a GBK file that is suitable for metabolic network reconstruction with Pathway Tools, and adds the GO terms and EC numbers annotations in the GenBank file.

URL: https://github.com/AuReMe/emapper2gbk

Publication: hal-02395024

Contact: Clemence Frioux

Participants: Clemence Frioux, Arnaud Belcour, Anne Siegel

7.1.5 Biodiversiton

Name: Biodiversiton

Keywords: Biodiversity, Comparative metagenomics, Clustering, Dimensionality reduction, Masses of data

Functional Description: Biodiversiton is a suite of tools for biodiversity composed by Rsyst, pairwise_dis, diagno_syst, and yapotu. The global project provides tutorials, datasets, and a readme for the whole suite.

URL: https://gitlab.inria.fr/biodiversiton

Contact: Alain Franc

Participants: Alain Franc, Jean-Marc Frigerio, Franck Salin

7.1.6 Yapotu

Name: Yet Another Pipeline for OTU building

Keywords: Metagenomics, Biodiversity, Dimensionality reduction, Masses of data

Functional Description: The main functionalities are as follows: 1) building OTUs from a fasta file (swarm, vsearch, ..) or a distannce file (yapotu) for an environmenal sample 2) building a fasta file and a distance file per OTU 3) checking the consistency of the OTUs by displaying them as a graph (see OTU as a graph below) 4) displaying the shape of an OTU or of a set of OTUs by Multidimensional Scaling 5) implementing Hierachical Aggregative Clustering of an OTU or a set of OTUs with various aggregation methods

URL: https://gitlab.inria.fr/biodiversiton/yap

Contact: Alain Franc

Participants: Alain Franc, Jean-Marc Frigerio, Franck Salin

Partner: INRAE

7.1.7 pydiodon

Name: pydiodon

Keyword: Dimensionality reduction

Functional Description: Most of dimension reduction methods inherited from Multivariate Data Analysis, and currently implemented as element in statistical learning for handling very large datasets (the dimension of spaces is the number of features) rely on a chain of pretreatments, a core with a SVD for low rank approximation of a given matrix, and a post-treatment for interpreting results. The costly part in computations is the SVD, which is in cubic complexity. Diodon is a list of functions and drivers which implement (i) pre-treatments, SVD and post-treatments on a large diversity of methods, (ii) random projection methods for running the SVD which permits to bypass the time limit in computing the SVD, and (iii) an implementation in C++ of the SVD with random projection at prescribed rank or precision, connected to MDS.

Release Contributions: - completed documentation with sphynx - library now public through Inria git - availability of a readme - making a few "toy" datasets available - delivering a few jupyter notebooks as tutorials

URL: https://gitlab.inria.fr/diodon/pydiodon

Contact: Alain Franc

Participants: Alain Franc, Florent Pruvost, Romain Peressoni

7.1.8 TANGO

Keywords: Computational biology, Systems Biology, Metabolic networks, Bacterial strains

Functional Description: The organoleptic properties that provide the added value of fermented dairy products result from specific metabolites that are produced by metabolic processes performed in concert by consortia of microbial species. TANGO enable a deeper understanding of the molecular and cooperative mechanisms underlying the production of organoleptic compounds. Tango uses a combination of whole-genome metabolic modeling and dynamic numerical simulation to assemble a complete, precise model of cheese production using lactic acid and propionic acid

bacteria. The results of this modeling reveal interactions between the members of the bacterial community, follow dynamically organoleptic compounds and fit with experimental data.

Contact: Simon Labarthe

Participants: Julie Aubert, Hélène Falentin, Clemence Frioux, Simon Labarthe, Maxime Lecomte, David Sherman

7.1.9 Mapler

Name: Metagenome Assembly and Evaluation Pipeline for Long Reads

Keywords: Metagenomics, Genome assembly, Benchmarking, Bioinformatics

Functional Description: Mapler is a pipeline to compare the performances of long-read metagenomic assemblers. The pipeline is focused on assemblers for high fidelity long read sequencing data (e.g. pacBio HiFi), but it supports also assemblers for low-fidelity long reads (ONT, PacBio CLR) and hybrid assemblers. It currently compares metaMDBG, metaflye, Hifiasm-meta, opera-ms and miniasm as assembly tools, and uses reference-based, reference-free and binning-based evalutation metrics. It is implemented in Snakemake.

URL: https://gitlab.inria.fr/mistic/mapler

Publication: hal-04142837

Contact: Nicolas Maurice

Participants: Nicolas Maurice, Claire Lemaitre, Riccardo Vicedomini, Clemence Frioux

7.1.10 seed2lp

Keywords: ASP - Answer Set Programming, Metabolic networks, Logic programming, Linear programming

Functional Description: Seed2lp is a tool for seed searching from metabolic networks using logic and/or linear programming (reasoning, FBA or hybrid). The solution, developed in python, uses Answer Set programming with clingo and clingo-lpx, and allows solution verification with Cobrapy.

Contact: Clemence Frioux

Participants: Chabname Ghassemi Nedjad, Clemence Frioux

7.1.11 GeMeNet

Name: Genomes to Metabolic Networks

Keywords: Bioinformatics, HPC, Metabolic networks, Genomics

Scientific Description: GeMeNet is a pipeline for generating multiple metabolic networks essentially from their genomes but from other data (gbk, ect...).

Functional Description: GeMeNet is a pipeline for generating multiple metabolic networks essentially from their genomes but from other data (gbk, ect...).

URL: https://gitlab.inria.fr/slimmest/gemenet

Contact: Coralie Muller

Participants: Clemence Frioux, Coralie Muller

7.1.12 CoCoMiCo

Name: Cooperation and competition potentials in large microbial communities

Keywords: Automated Reasoning, Metabolic networks, Answer Set Programming, Microbiota, Systems Biology

Scientific Description: By discretely modelling metabolic cross-feeding and dependency on limiting metabolites between organisms, CoCoMiCo defines novel optimisation criteria that can be used at scale for screening microbial communities using combinatorial methods. The criteria can be used for evaluation of large sets of naturally-occurring communities, or of large sets of generated candidate communities screened to identify species of interest for health or ecology applications.

Functional Description: Metabolic cross-feeding and dependency on limiting metabolites between organisms are logically modeled using an ad hoc knowledge base derived from whole genome metabolic models in SBML format, and analyzed by logical inference rules defined using the answer set programming paradigm.

URL: https://gitlab.inria.fr/CCMC/CoCoMiCo

Contact: David James Sherman

Participants: Maxime Lecomte, David Sherman, Coralie Muller, Chabname Ghassemi Nedjad, Clemence Frioux

7.1.13 MetagenoPIC

Name: Metagenomic pipeline creation

Keywords: Metagenomics, Genome assembly

Scientific Description: MetagenoPIC pipeline contain differents steps in order to reconstruct Metagenome Assembled Genomes (MAGs).

The first step of pipeline is read trimming using kneaddata, with read filtering against organism databases. The next step is assembly step, using either Megahit or metaSpades. Reads are aligned to these contigs using BWA-MEM2. The following step is binning: the pipeline groups contigs into bins using one of two ways. The first way is to use a single one of the MetaBAT2, MaxBin2, or CONCOCT tools. The second way is to run a combination of these tools and then concatenate the results using a bin refiner, either DASTool or MetaWrap. The next step is to dereplicate the resulting bins, filtering reads of poor quality (high contamination and low completeness) and dereplicating bins that can be represent the same metagenome. In all cases, a threshold for the contamination and completeness can be specified.

Once the pipeline has constructed MAGs, it can run further analyses: CheckM to evaluate the quality of bins, and GTDB-TK to run a taxonomic assignment.

Functional Description: Metagenome reconstruction comprises four steps: assembly, binning, dereplication, annotation.

Release Contributions: Internal availability

News of the Year: The first internal release of MetagenoPIC based on work by Ariane Badoual provides a full pipeline that can be executed by any system that respects the Common Workflow Language. Validation was performed using Calrissian on Kubernetes and Toil on slurm. Work by Leonard Brindel this year included integration of long reads, improvements to functional and structural annotation using Prodigal and Eggnog-mapper, and an overall increase in reliability.

URL: https://gitlab.inria.fr/metagenopic

Contact: Clemence Frioux

Participants: Ariane Badoual, Clemence Frioux, David Sherman, Leonard Brindel

7.1.14 pherosensor-toolbox

Keywords: Data assimilation, Computational biology

Scientific Description: Insect pests are a major threat to agricultural systems, leading to intensive use of pesticides for crop protection with unsustainable drawbacks on the environment, biodiversity, and human health. Most insects produce pheromones for conspecific communication, making pheromone sensors an effective tool for early specific detection of pests, in order to reduce pesticide use within the context of precision agriculture.

'Pherosensor-toolbox' is a Python package containing numerical tools for pheromone sensor data assimilation to infer the position of emitting pest insects. It contains specific tools to model pheromone propagation and solve the corresponding inverse problem to determine emitters' position taking into account the environmental context (wind, landscape, vegetation...). A specific focus is put on the integration of biological knowledge of pest behavior during inference.

Functional Description: This toolbox brings together numerical methods for solving a data assimilation problem for a reaction-convection-diffusion PDE describing pheromone propagation in an agricultural landscape using variational methods penalised by biology-informed regularisation terms (population dynamics of the emitting insect, preferred habitat, exclusion zones).

News of the Year: publication dans JOSS

URL: https://forgemia.inra.fr/pherosensor/pherosensor-toolbox

Publications: hal-04669546, hal-04572831

Contact: Simon Labarthe

Participant: Simon Labarthe

7.1.15 AsebaHub

Name: Turn-key bridging of Aseba mobile robots to wifi networks

Keywords: Robotics, Education

Scientific Description: The Thymio-II educational robot teaches robotics programming to 8-18 year old children. Since 2014 Inria has contributed to the Thymio ecosystem. Thymio-II robots only communicate through USB connections from a host computer, so until now it has been necessary to install Aseba software on that computer. This is not always possible in schools, which increasingly use tablets.

This software provides an OpenWRT-based firmware image for a wifi Access Point that takes responsibility for communication with each robot, advertising it on the local network to learner programming environments running on tablets. It provides device manager and web-based configuration services, out of the box.

Functional Description: AsebaHub is a small wifi access point for Thymio educational robots, connected by USB or LR-WPAN. Every robot is made available as a network target on the wifi and wired local-area networks, where they can be discovered using mDNS-sd (Zeroconf/Bonjour). No network or user configuration is necessary. The device works out of the box with Thymio programming environments, including VPL3, Scratch, Python and Aseba Studio.

Release Contributions: Release v1.0.0 provides a complete firmware image based on OpenWRT, for compatible mini-routers including GL.iNET and Raspberry Pi.

News of the Year: AsebaHub compiled for ramips-mt76x8 is deployed as the firmware for Mobsya's Thymio 2+ mini-router (https://www.thymio.org/products/thymio-2plus) and is currently installed in schools.

Contact: David Sherman

Participant: David Sherman

7.2 New platforms

Participants: David Sherman, , Ahmed Kallel (DSI), , Jean-François Scariot (DSI) .

As a founding principle, Pleiade supports reproducible scientific analyses and promotes a declarative approach using reusable software modules, rigorous documentation of data provenance, and systematic recording of workflows. The latter is a challenge when interactive interfaces are used, but can be addressed, to cite two examples, in Galaxy by extracting workflows, and in other systems by using Jupyter notebooks. Part of Pleiade's mission is to automate the deployment of environments that support these goals, for non-technical end users.

Pleiade has built a Kubernetes platform *Pleiadès* hosted in the Inria research center at the Univ. Bordeaux, which in 2023 was been integrated into Inria's IT Management (DSI-SP).

Use cases were identified by the project-team and from the MISTIC data management plan:

- Fast deployment of **containerized user environments**, combining biological data and databases, software modules specified by version, a CWL executor, and interactive tools including web front ends, notebooks, or Galaxy. A user environment will provide at least one specific HTTPS endpoint, created dynamically. A single researcher may deploy several different environments in the course of one day.
- Support for **development and testing of workflows**, as above but configured for team members who are developing software modules or interfaces, and who must often deploy several different environments simultaneously.
- Dynamically allocated **containerized compute tasks**, including both individual analysis steps in workflows and GitLab runner containers used for continuous integration. These tasks arrive in bursts that often cannot be planned in advance.
- Long-running **stream preprocessing**, a low-priority background task that watches external databases for changes, chooses pertinent data, precomputes representations and ingests them into local data bases.

The following requirements were derived from these use cases:

- Tasks must run in OCI containers. A typical environment will be constructed from ten to one hundred containers, grouped in Kubernetes Pods of co-localized containers that share a private network.
- Containers run unprivileged and must rely on role-based access control (RBAC), secrets, and service accounts.
- Different storage classes must be available for dynamic volume allocation: ReadWriteSingle, ReadWriteMany, Object (S3) Bucket.
- An application must be able to allocate a route with wildcard DNS in order to offer an endpoint, internally to the Inria network.
- A collection of Kubernetes custom resource definitions and RBAC definitions, specific to Pleiade's applications, is needed.
- A collection of OpenShift Operators for deployment of applications, is needed. These include database services, workflow execution, and container building using source-to-image (S2I).
- A management interface through the OKD console that allows inspection and management of app topologies, pods, volumes, and Kubernetes objects.

We support community best practices for reproducible computing in bioinformatics, using biocontainers generated by bioconda, in CWL or Galaxy workflows. For internal use we provide model serving endpoints and host JupyterHub environments.

The Pleiadès platform is built on OKD 4, the community distribution of Kubernetes developed alongside of RedHat Openshift. OKD4 in particular uses the CRI-O runtime, not Docker, and containers run unprivileged. Software-defined storage and S3 endpoints are provided by Ceph. Pleiadès follows the *gitops* pattern and all management and implementation use Git repositories as the single source of truth.

Continuous integration for software development is supported for Inria's Gitlab instance. Thirteen project-specific CI runners are currently hosted on Pleiadès.

To support our scientific users, Pleiadès hosts an instance of *Open Data Hub* (ODH), an AI platform for the hybrid cloud. Each project in ODH can host Jupyter workbenches, shared cluster storage and data connections to S3 buckets, pipelines, and AI model serving runtimes including Kserve and OpenVINO.

7.3 Open data

TANGO

Contributors: Hélène Falentin, Maxime Lecomte, Wenfan Cao, Julie Aubert, David Sherman, Clémence Frioux, Simon Labarthe

Description: Genomes, metatranscriptomics, targeted metabolomics and modeling of L. lactis CIRM-BIA1206, L. plantarum CIRM-BIA465 and P. freudenreichii CIRM-BIA122 throughout making and ripening of semihard model cheeses

Dataset PID: https://doi.org/10.57745/X4C5VR

Project link: https://forgemia.inra.fr/tango/tango_models

Publications: [14, 46, 61, 63]

Contact: Simon Labarthe

Release contributions: Initial release.

CoCoMiCo

Contributors: Maxime Lecomte, Coralie Muller, Ariane Badoual, Helene Falentin, David Sherman, Clemence Frioux

Description: Curated metabolic networks for root microbiota (https://doi.org/10.1038/nature16192) leaf microbiota (https://doi.org/10.1038/nature16192) soil microbiota (https://doi.org/10.1038/ismej.2016.168) gut microbiota (https://doi.org/10.1038/s41587-018-0008-8)

Dataset PID: https://doi.org/10.5281/zenodo.7551375

Project link: https://gitlab.inria.fr/CCMC/CoCoMiCo.git

Publications: [62, 63]

Contact: Clémence Frioux

Release contributions: Initial release.

8 New results

8.1 Inferring seeds from metabolic networks

Participants: Chabname Ghassemi Nedjad, Clémence Frioux.

As a part of her PhD project, Chabname Ghassemi Nedjad develops an approach for the identification of seed metabolites in metabolic networks. The underlying biological challenge is to facilitate the culture of microbial dark matter, micro-organisms that cannot be cultured in controlled media experimentally. Usually, these organisms exhibit auxotrophies that prevent their growth on most nutritional conditions.

We take advantage of the possibility to reconstruct metabolic networks from genomic information, thereby obtaining a blueprint of the metabolic potential of species. We apply to this metabolic network a set of rules and constraints under the reasoning paradigm of Answer Set Programming (ASP) to provide sets of metabolites that would enable the activation of functions of interest in the corresponding species. The core of the model is network expansion (NE), a Boolean abstraction of metabolic activity. Yet, because most models of metabolism are numerical, relying on the widespread paradigm of Flux Balance Analysis (FBA), this works developped hybrid NE-FBA formalisms to guarantee both a positive flux in the objective reaction, and the Boolean reachability of its reactants. A preprint of this work was released in 2024 [25], and a tool – Seed2LP – implementing the approach is available (7.1.10).

Chabname is co-supervised by Loïc Paulevé (CNRS, LaBRI) and Clémence Frioux.

8.2 Metagenomic assembly of complex microbial ecosystems

Participants: Nicolas Maurice, Franck Salin, Clémence Frioux.

The interest of Pleiade for the treatment of DNA sequences has renewed over the past few years with the MISTIC project aiming at developing models for complex microbial communities in an Agroecology context.

The PhD project of Nicolas Maurice targets the difficult task of assembly genomes in high-diversity microbiomes such as soil. He developed MAPLER (7.1.9) for the assessment of metagenomic assembly quality. In 2024, Nicolas presented his work at the 24th Genome Informatics Conference in Hinxton, United Kingdom [19]. He is co-supervised by Claire Lemaitre, Ricardo Vicedomini (Inria centre at the University of Rennes, Genscale team) and Clémence Frioux, and is hosted in the Genscale Inria team.

Application of Nicolas' work to soil microbiome samples was presented in EAGS 2024 - The International Environmental and Agronomical Genomics symposium (Toulouse, France) [30].

8.3 Modeling the emergent metabolic potential of soil microbiomes in Atacama landscapes

Participants: Coralie Muller, Clémence Frioux.

Pleiade (Clémence Frioux) and Dyliss (Anne Siegel) Inria teams co-lead an *Associated team* with the group of Alejandro Maass at the Center for Mathematical Modelling in Santiago de Chile. This collaboration led to a work published as a preprint in 2024 [23] characterising the metabolic potential in soil microbiomes along the Talabre-Lejía transect (TLT) in the Atacama desert.

Metagenomic sequences and environmental metadata were available for six sites along the altitudinal gradient of the TLT. We designed and applied a systems biology framework to characterize the metabolism of the soil bacterial communities associated to each site. We built metabolic models both at the level of the metagenome and genome-resolved, the latter related to metagenome-assembled genomes reconstructed

at each site. This two-pronged approach enabled us to compare the global metabolic potential at a site, with the predicted metabolisms of the organisms that were abundant enough for their genomes to be assembled. Among our results, we observed that the metabolic potential of all sites was overall stable at the scale of the metagenome, but differences emerge at the genomic level, permitting adaptation to the diverse environmental conditions along the transect.

8.4 Numerical models of microbial community metabolism

Participants: Maxime Lecomte, Amandine Paulay, Coralie Muller, David Sherman, Clémence Frioux, Simon Labarthe.

Building up on metabolic models of microbial strains to derive dynamical models of microbial communities is still a major scientific challenge that Pleiade is addressing. At the interface between bioinformatics and applied mathematics, the objective is to solve numerical issues in order to couple large-scale accurate metabolic models of microorganisms with ordinary or partial differential equations describing the population dynamics, in order to integrate multi-omic data.

A first achievement in 2024 was a conceptual work with the publication of a review covering the whole chain of numerical method, from the metagenomic data to the community-scale model [13]. This paper has been published in a highly reputed journal in microbial ecology community and positioned the team in the community of modelers of microbial ecosystems.

Next, a dynamic model of protein degradation by a gut symbiont (*B.caccae*) has been developed during Amandine Paulay's PhD. Amandine Paulay defended her PhD in june 2024 at the Abies doctoral school of Paris Saclay university. Her PhD project was a collaborative work between Micalis and MaIAGE units (INRAE), the startup Biomathematica, and the Inria teams Pléiade and Musca. The model, that coupled an accurate metabolic model of a *B.caccae* strain, curated to account with proteolysis, with a dynamic model describing microbial growth in a media containing high or low levels of proteins. The model was then fitted and validated on independant datasets of metabolomics data. This work was published in 2024 in *Msystems* [16].

A part of Maxime Lecomte's PhD project consisted in integrating heterogenous *omics* data to build a dynamic model of cheese production. The associated manucript was published in *Metabolic Engineering* in 2024 [14] and the corresponding codes were released [39]. It highlights a methodological approach for the construction and validation of dynamic metabolic models using dFBA. Using metabolic models for three bacterial species involved in cheese making, a careful curation and validation of metabolic functions was performed. Individual dynamics were calibrated using pure culture experimental data. The dynamics was then validated at the community level, illustrating the roles of each microbe in the production of flavour compounds.

Coupling metabolic models and dynamic systems is also the topic of Sahak YEGHIAZARYAN's PhD, started in dec. 2023. This project is a collaborative work with the LBE INRAE Laboratory (Narbonne), and is co-supervised by Nicolas Bernet, Elie Le Quemener and Simon Labarthe. The aim of the project is to study a syntrophy articulated around hydrogen metabolism in anaerobic digestion in a bioprocess context. Sahak YEGHIAZARYAN will couple metabolic models to thermodynamics-based kinetic models of microbial growth.

8.5 Digital twins of microbial communities

The topic of digital twins of microbial communities has been a hot topic during 2024, through the Artemis project (INRAE metaprogram Digitbio), and the WP2 of Mistic (PEPR Agroecology and IT).

8.5.1 Artemis project

Participants: Simon Labarthe, Clémence Frioux, David Sherman, Coralie Muller, Chabname Ghassemi Nedjad, Franck Salin, Sthyve Tatho.

The Artemis project is a research network founded by the INRAE metaprogram Digitbio. It gathers about 40 researchers from INRAE and Inria, either modelers developping models of microbial systems or experimentalists with strong interest in mathematical models and digital twins. The different microbial systems represented in the network covers the whole set of applications of microbiology at INRAE : environmental, bioprocesses, plant, animal, food and human microbiology. The network gathers different Inria teams, in particular Macbes, Musca and Pleiade. The final goal of the network is to produce an opinion paper about the concept of digital twins applied to microbial systems.

In 2024, we organized the kickoff meeting (webinar), a workshop about digital twins (hybrid format: at Inria Bordeaux and on line) and a training (at Inria Bordeaux) about the whole modeling chain, from sequence to dynamical model of microbial community.

8.5.2 Community-scale Metabolic Flux Analysis

Participants: Sthyve Tatho, Simon Labarthe.

In the Mistic project, the task WP2.2 is devoted to the development of a digital twin of a microbial experiment conducted in WP2.1 on a simplified community (SimCom) controling powdery mildew. The first bottleneck to overcome towards a digital twin is the capacity to integrate multi-omics time series to rapidely build a dynamic system.

It is the subject of Sthyve Tathos's PhD, started in january 2024. During this year, he extended a method previously used for single strains, Metabolic Flux Analysis (MFA) at the community scale (c-MFA). The method consists in extracting growth, consumption and production rates in multi-omics time series, and to analyse them using a metabolic model to identify the co-variation structures between metabolic fluxes. During his first year, Sthyve Tatho presented his work at the "Congrés des jeunes chercheur.e. en mathématiques appliquées" [35].

8.6 Pherosensor

Participants: Thibault Malou, Simon Labarthe.

Pherosensor is a project founded by the ANR-PPR "Cultiver et protéger Autrement" program dedicated to the developement of pheromone sensor to track insect pests. The final goal is to use pheromone sensors in a precision agriculture framework as a early detection tool, in order to reduce pesticide use.

Pleiade team is involved in the WP3 of the project, in collaboration with MaIAGE and ISEE units (INRAE). The objective is to develop mathematical models to solve the inverse problem of pheromone propagation: from time-series obtained from a spatial network of sensors, the goal is to track back pheromone plumes towards the emission source, i.e. the targeted pest insect. To this end, we specifically focus on integrating biological prior knowledge about pest in the inverse problem, by developing a biology-informed data assimilation method (BI-DA).

In 2024, a python toolbox has been released 7.1.14 and published in *Journal of Open Source Software* [15]. The work has been also presented in conferences [17], and a preprint benchmarking diverse BI-DA methods [26].

Pesticide-free agriculture is also the topic of the VITAE project that resulted in a review work about pestice-free viticulture presented to a conference [18].

8.7 Short Time-series Modelling

Participants: Guilhem Sommeria-Klein, Chandler Ross.

Ecosystems tend to fluctuate around stable equilibria in response to internal dynamics and environmental factors. Occasionally, they enter an unstable tipping region and collapse into an alternative stable state. This is, for instance, the case of lake plankton communities, or bacterial communities in the human gut microbiota. The latter case is relevant to human health, since many diseases are associated with altered states of the gut microbiota.

Our understanding of how ecological communities vary over time and respond to perturbations depends on our ability to quantify and predict these dynamics. Mechanistic models of microbial community dynamics often fail to characterise observed fluctuations in naturally occurring microbiomes and inform us about key dynamical properties such as stability and resilience. An alternative approach is to characterise the dynamical landscape using non-parametric models. However, the scarcity of long, dense time series data creates a severe bottleneck for characterising community dynamics using existing methods.

With Chandler Ross, a student at the University of Turku (Finland) co-supervised with Leo Lahti, we have worked on overcoming this limitation by combining information across multiple short time series using Bayesian inference. By decomposing dynamics into deterministic and stochastic components using Gaussian process priors, we predict stable and tipping regions along the community landscape and quantify resilience, while addressing uncertainty. In particular, we estimate a recently proposed probabilistic metric for resilience in multistable systems: the expected time before exiting the current stable state under stochastic fluctuations (or "exit time").

We validated our approach using simulated data and showed in particular that our model is able to distinguish bistability from bimodality, which are often conflated in classical potential analyses. We further demonstrated our approach by re-analysing ecological time series data of lake cyanobacteria abundance, for which we recovered similar results as a previous study (Arani et al. 2021, Science) using three orders of magnitude fewer data points. Finally, we used our model to re-evaluate the stability of previously proposed “tipping elements" in the human gut microbiota. We confirmed the apparent bistability of Prevotella bacterial genus but invalidated that of Dialister, a distinction impossible to make with approaches not based on a probabilistic model.

The corresponding manuscript was published on arXiv [28] in January 2025 and is currently under review in Methods in Ecology and Evolution. We are currently working on extensions of this approach to model the full community composition.

8.8 Microbial Metacommunity Modeling

Participants: Guilhem Sommeria-Klein.

Microbial communities are increasingly being sampled in spatial-temporal contexts that require taking into account the influence of dispersal fluxes between sampled communities. Few probabilistic approaches currently exist for explictly modelling communities connected by dispersal, forming what is called a “metacommunity” in ecology. With collaborators in the UK (Aura Raulo, Oxford University, and Christopher Quince, Earlham Institute), we are working on modelling the colonisation-extinction dynamics of microbial taxa in the case of a metacommunity where most constituent communities are being sampled though time.

We are applying this approach to model bacterial gut microbiota data from a population of around 100 wild mice with known interaction network and individual home ranges, sampled through time over one season in Wytham Woods, Oxford (UK). Most individuals in the interaction network can be considered to have been tracked. The modelling approach has already allowed us to pinpoint marked differences in colonisation and extinction dynamics depending on bacterial phenotype (anaerobes vs. aerobes). Guilhem Sommeria-Klein visited Oxford for 2 weeks in December 2024 as part of this collaboration.

This type of approach could be applied to other data types, including human microbiota. A promising case study is the study of antibiotic resistance gene transmission. A collaboration is under way on this topic with Finnish collaborators, which led to a preprint published in August 2024 on medRxiv [27], currently under review in Nature Communications.

8.9 Human Gut Microbiota Phylosymbiosis

Participants: Guilhem Sommeria-Klein.

The human gut microbiota exhibits large variations in microbiota taxonomic composition between different populations. The role of industrialization in generating these variations has been recently emphasized, however less attention has been paid to pre-existing variations across non-industrial populations. Yet, the rapid adaptation of our species to nearly all terrestrial habitats over the last 150,000 years is suspected to have been partly mediated by the gut microbiota, and it remains unclear how much this history has shaped the gut microbiota variations observed across modern populations.

With collaborators at ENS in Paris, Benoît Perez-Lamarque and Hélène Morlon, we reconstructed a tree of divergence times summarising the demographic history of 24 representative non-industrial populations from across Africa, Eurasia and the Americas whose gut microbiota has been sequenced, and we built a Bayesian model to quantify the relative contribution of demographic history, mode of subsistence and natural environment in shaping their microbiota. We found that, at the level of bacterial genera and families, similarity in microbiota composition tends to increase with more recent divergent time between populations. We previously reported this pattern, known as phylosymbiosis, over longer evolutionary timescale across different mammal species in a 2023 paper [67]. We inferred systematic covariations between bacterial taxa and reconstructed the most likely changes in ancestral microbiota composition along human demographic history. The manuscript is in preparation.

8.10 Diversity Estimation

Participants: Guilhem Sommeria-Klein.

It has been observed that, while the number of beetle species in a given plot of Amazonian forest is as high as 100,000, two distant enough Amazonian sites of similar habitat have virtually no species in common. If one knows the shape taken by the decay of the similarity in species composition with distance, one could in principle use this information to extrapolate local diversity to the regional scale, and thus estimate the number of beetle species in Amazonia.

More generally, with colleagues at ENS in Paris, Odile Maliet and Hélène Morlon, as well as with a master student supervised during spring 2024, we have worked on estimating the number of species in a large area of homogeneous habitat, such as Amazonia, for a species group with a high spatial turnover in species composition, given at least one well-sampled estimate of local diversity and the shape of the distance-decay of community similarity. We have derived a mathematical relation between these quantities based on geometrical arguments and developped a Bayesian inference model based on it. We tested it on simulated data, as well as on global mammal and bird data of known diversity. The manuscript is in preparation.

8.11 Metabarcoding and Taxonomic Diversity

Participants: Jean-Marc Frigerio, Alain Franc.

Metabarcoding is a series of technical procedures to build molecular based inventories from large datasets of amplicons. The underlying information needs to be compacted without losing its information content before it can be further processed with domain-specific tools. This links metabarcoding tools to dimension reduction techniques, which is an important topic in Pleiade.

Alain Franc and Jean-Marc Frigerio have continued their efforts to complete the *yapotu* software (§7.1.6), which makes the main stages of barcoding and metabarcoding data analysis accessible. The main

progress in 2023 has been, beyond the development of some methods, the development of a Domain Specific Language, in Python, called *yapsh*, that makes the execution of instructions more automatic and comfortable for the user. The project, still in development, is accessible publicly on the Inria Gitlab.

The work on the evaluation of the intrinsic quality of OTUs produced in metabarcoding (Marie-Josée Cros, Jean-Marc Frigerio, Nathalie Peyrard and Alain Franc,Simple approaches for evaluation of OTU quality based on dissimilarity arrays) , has been achieved by submitting a paper to the Journal MBMG, currently in revision. The preprint of the manuscript is available at [49]. As a companion to this work is available in a Git project on the public "Forge MIA" of the Inrae. It contains the documentation for the programs implementing the methods, as well as a toy data set with a tutorial on how to use it.

8.12 Dimension Reduction

Participants: Alain Franc.

Alain Franc's comprehensive report on Linear Dimensionality Reduction, Research Report 9488, Inria Bordeaux Sud-Ouest. 2023, 99 pages, is available at [52] and arXiv 2209.13597. In 2024 editorial work was performed to adapt the manuscript to monograph form for book publication by a major academic press.

9 Bilateral contracts and grants with industry

In 2024 Pleiade's impact on industry was channeled through agroecological project MISTIC (§10.3.1) and did not require bilateral projects.

10 Partnerships and cooperations

10.1 International initiatives

10.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

Clémence Frioux co-leads the Inria associated team SymBioDiversity with the University of Santiago de Chile (Center for Mathematical Modeling).

10.2 International research visitors

10.2.1 Visits of international scientists

Other international visits to the team

Sebastián Mendoza

Status researcher

Institution of origin: Centro de Modelamiento Matematico, Universidad de Chile

Country: Chile

Dates: October 7th to October 14th, 2024

Context of the visit: Inria Associated Team SymBioDiversity

Mobility program/type of mobility: Research stay

10.2.2 Visits to international teams

Research stays abroad

Clémence Frioux

Visited institution: Centro de Modelamiento Matematico, Universidad de Chile

Country: Chile

Dates: January 11th to January 19th, 2024

Context of the visit: Inria Associated Team SymBioDiversity

Mobility program/type of mobility: Research stay

Guilhem Sommeria-Klein

Visited institution: University of Oxford

Country: UK

Dates: December 2nd to 13th, 2024

Context of the visit: On-going collaboration with Aura Raulo, Christopher Quince and Sarah Knowles' lab

Mobility program/type of mobility: Research stay

Guilhem Sommeria-Klein

Visited institution: University of Turku, University of Helsinki

Country: Finland

Dates: January 7th to 10th, 2025

Context of the visit: Visiting supervised student Chandler Ross (Turku), collaborators Matti Ruuskanen and Katariina Pärnänen (Helsinki)

Mobility program/type of mobility: Research stay

10.3 National initiatives

10.3.1 MISTIC (PEPR Agroecology and ICT)

Participants: David Sherman, Clémence Frioux, Simon Labarthe, Franck Salin, Alain Franc, Jean-Marc Frigerio, Nicolas Maurice, Coralie Muller, Sthyve Tatho.

MISTIC, *Microbial communities and ICT*, has been selected as a five-year flagship project in the PEPR Agroecology and ICT program of the French Government. MISTIC will develop methodological tools for defining spatio-temporal models of microbial community dynamics in the phyllosphere and rhizosphere of crop plants, with the goal of creating new understanding of the role of these communities in plant adaptation to environmental stresses, including climate change. MISTIC is a partnership between seven Inria and INRAE teams in Bordeaux, Rennes, and Sophia Antipolis. The project formally began in November 2022.

10.3.2 CULTISSIMO (PEPR Food Systems, Microbiome, and Health)

Participants: Clémence Frioux, Simon Labarthe.

The Cultissimo project, fonded by the PEPR Food Systems, Microbime and Health (SAMS), is dedicated to the development of culturomics approaches to study the human gut. The Pleiade team is involved, together with MaIAGE (INRAE) and Musca team, in the developement of modeling approaches to predic culture media to cultivate microbial defined communities.

10.3.3 VITAe, Pherosensor (PPR Cultiver et proteger autrement)

Participants: Simon Labarthe.

Pleiade participates to two projects of the PPR CPA, dedicated to research towards an agriculture without pesticide. Pléiade co-leads a work package of the VITAe project, taking in charge modeling tasks to analyse culturomics data in order to identify antagonist micro-organisms against powdery mildew in grapewine. Pléiade leads a work package of the Pherosensor project, dedicated to the design of new sensors of pheromone. The main task of the team is to solve an inverse problem on a PDE model of pheromone propagation to track back the pheromone emitters.

10.3.4 Holovini (Holoflux INRAE metaprogram)

Participants: Simon Labarthe, Clémence Frioux.

Pleiade is involved in the Holovini flagship project of the Holoflux metaprogram. Holovini studies the berry microbiome of grapewine, focusing on the microbial flux involved in the assembly of the berry microbiome. Pléiade takes in charge the analysis of metagenomics data and the co-lead of a modeling workpackage.

10.3.5 REBON (ANR)

Participants: Clémence Frioux.

REBON, piloted by Joachim Niehren, will abstract reaction networks to boolean networks with the goal of improving inference and control in systems biology.

10.3.6 Artemis (Digit-bio INRAE metaprogram)

Participants: Simon Labarthe, Clémence Frioux, David Sherman.

Pleiade pilots the Artemis pre-project funded by the Digit-bio metaprogram, aimed at developing methodologies for defining digital twins in microbial ecology.

10.3.7 Microsentry (SPE and MAthnum INRAE departments)

Participants: Simon Labarthe, Clémence Frioux.

In collaboration with SAVE unit (INRAE), this project aims to explore the use of q-PCR screening of sentry microorganisms to get early detection of powdery mildew infection in environmental samples. Pleiade team is involved in methodological developments to select a small set of sentry microorganisms.

10.3.8 COMIC (Holoflux INRAE metaprogram)

Participants: Clémence Frioux, Juliette Audemard, Coralie Muller.

The Metaprogramme HOLOFLUX at INRAE funded in 2025 the COMIC project, led by Binta Diémé and Nicolas Creusot on the application of a systems biology approach to the understanding of cyanobacterial blooms in freshwater. The main objective is to better understand the metabolic relationship between the cyanotbacteria and their heterotrophic symbionts. Juliette Audemard's internship in the team was dedicated to this project.

10.3.9 Agence Française pour la Biodiversité

Participants: Alain Franc, Jean-Marc Frigerio.

The AFB is a public law agency of the French Ministry of Ecology that supports public policy in the domains of knowledge, preservation, management, and restoration of biodiversity in terrestrial, aquatic, and marine environments. Pleiade is a partner in two AFB projects developed with the former ONEMA: one funded by ONEMA, the second by labex COTE, where BioGeCo/Pleiade is responsible for data analysis, with implementaton of the tools recently developed for scaling MDS. Calculations have been made on CURTA at MCIA and PlaFRIM at INRIA.

10.4 Regional initiatives

10.4.1 Nouvelle-Aquitaine research project MicroMod

Participants: Simon Labarthe, Clémence Frioux, Franck Salin, Coralie Muller, Alain Franc, Jean-Marc Frigerio, Guilhem Sommeria-Klein.

MicroMod is a research project that unifies deferent researches previously conducted in the team in an holistic framework. MicroMod aims to 1) identify protective simplified communities (SimCom), 2) develop more accurate models of SimCom, 3) develop simple biomarkers of health of microbial community at field.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- Clémence Frioux, Simon Labarthe - Co-head of the 2025th edition of the French Bioinformatics Conference, JOBIM

11.1.2 Scientific events: selection

Member of the conference program committees

- Clémence Frioux – Proceedings Program Committee of International Conference on Intelligent Systems for Molecular Biology (ISMB 2024)

Reviewer

- Clémence Frioux – Proceedings Program Committee of International Conference on Intelligent Systems for Molecular Biology (ISMB 2024)

11.1.3 Journal

Reviewer - reviewing activities

- Cell Reports – Clémence Frioux
- Microbiome – Clémence Frioux
- Biotechnology and Bioengineering Journal – Clémence Frioux
- Peer Community In (PCI) Math Comp Biol – Clémence Frioux
- Journal of Theoretical Biology – Simon Labarthe
- Peer Community in (PCI) microbiology – Simon Labarthe
- Communications Earth & Environment – Guilhem Sommeria-Klein

11.1.4 Invited talks

- Clémence Frioux – Invited Keynote at the French Bioinformatics Conference, JOBIM
- Simon Labarthe – Invited Keynote at the Workshop Modélisation du Métabolisme, Toulouse
- Simon Labarthe – Invited to the Workshop *Mathematical modeling in life and health sciences* at Wolfgang Pauli Institute, Wien
- David Sherman – Invited talk at the LABIP Expert Workshop on *AI tools for fermentation modeling*, Lesaffre Campus, Lille

11.1.5 Leadership within the scientific community

- David Sherman is on the steering committee of Biosena, a regional research network of the New Aquitaine region dedicated to Biodiversity and Ecosystemic Services. Biosena associates actors from the academic and socio-economic sectors, with the goal of contributing to the understanding and preservation of biodiversity and to the improvement of ecosystemic services. Biosena contributes to this goal through research, knowledge dissemination, outreach, and skill transfer in the form of Research Action, in keeping with the recommendations of Ecobiose.
- Simon Labarthe is a on the steering committee of the Holoflux metaprogram of the INRAE. Holoflux works to achieve a clearer understanding of the functioning of holobionts and their interactions with the environment in various agronomic contexts. The interactions between microbiota and hosts (plant, animal, human) and the flux of microorganisms between holobionts and more broadly in the agri-food system as a whole, are levers that can be acted upon to improve the performance and sustainability of agricultural sectors.

11.1.6 Scientific expertise

- Clémence Frioux – Reviewer for ANR AAPG 2024.
- Simon Labarthe – Reviewer for Abies (U.Paris Saclay) doctoral school PhD grant call.
- Simon Labarthe – Reviewer for KFG (Austrian national funding agency for research)

11.1.7 Research administration

National responsabilities

- Project coordinator, MISTIC (§10.3.1) – David Sherman
- Participation to the Inria national committee for equality and inclusion – Clémence Frioux
- Participation to the Holoflux INRAE metaprogram steering committee.
- Participation to the CSS (commission scientifique spécialisée) MISTI of INRAE

Local responsabilities

- Co-creation of and participation in a gender equality and diversity working group in the Inria Centre at the University of Bordeaux – Clémence Frioux
- Resource person for the 3D printing and electronics workshop – David Sherman
- Workplace first-aid (SST) – David Sherman
- Workplace psychosocial risk first-aid – Clémence Frioux

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Master – ENSTBB Bordeaux INP - Bioinformatics (20 hours) – Clémence Frioux
- Master – ENSEIRB Bordeaux INP - Research algorithms (10 hours) – Clémence Frioux
- Doctorate – Computer science and mathematics doctoral school (7 hours) – Clémence Frioux

11.2.2 Supervision

Ongoing PhD projects

- PhD of Chabname Ghassemi Nedjad (2022-2025) - *Combinatorial optimisation problems for reverse ecology* - Clémence Frioux (co-director), Loïc Paulevé (CNRS, LaBRI, co-director).
- PhD of Nicolas Maurice (2023-2026) (Genscale, Inria centre at the university of Rennes) - *Sequence algorithmics for genome reconstruction from complex metagenomic data* - Claire Lemaitre (Inria, director), Ricardo Vicedomini (CNRS, co-director), Clémence Frioux (co-advisor).
- PhD of Sahak Yeghiazaryan (2023-2026) (LBE, INRAE, Narbonne), *Coupling genome based and energy based approaches of syntrophic microbial interactions for the modelling of high rate anaerobic digestion* - Nicolas Bernet (INRAE, director), Simon Labarthe (co-director), Elie Le Quemener (INRAE) (co-advisor).
- PhD of Sthyve Tatho (2024-2027) - *Integration of multi-omics data for the analysis of microbial community dynamics in plant health* - Simon Labarthe (co-director) and Valentina Baldazzi (co-director).

- PhD of Mathilde Sola (2024-2027) - *Characterization of large-scale "gut microbiota-diet-health" links in humans using combined approaches of digital microbial ecology, metabolic modeling, and artificial intelligence* - Patrick Veiga (INRAE, director), Magali Berland (INRAE, co-advisor) Clémence Frioux (co-advisor).
- PhD of Coralie Muller (2024-2027) - *Generation of metabolomic-informed models of metabolism in complex microbial communities* - Clémence Frioux (co-director), Sylvain Prigent (INRAE, co-director).
- PhD of Chandler Ross (2022-2026) - *Reconstructing community dynamics from limited observations* - Guilhem Sommeria-Klein (co-supervisor), Leo Lahti (University of Turku, Finland, director).
- PhD of Moein Khalighi (2021-2025) - *Impact of memory on complex dynamics* - Guilhem Sommeria-Klein (co-supervisor), Leo Lahti (University of Turku, Finland, director).

PhD defended this year

- PhD of Maxime Lecomte (2020-2024) - *Hybrid approach for explainable metabolic modelling of microbial ecosystems* - David Sherman (Director), Hélène Falentin (director, INRAE STLO), Clémence Frioux (advisor). Defended on April 29th, 2024.
- PhD of Amandine Paulay (2020-2024) (Micalis, INRAE Jouy en Josas) - *Modeling the degradation of dietary proteins by the human gut microbiota* - Emmanuelle Maguin (Director, INRAE, Micalis), Beatrice Laroche (co-Director, INRAE, MaIAGE), Simon Labarthe (advisor, INRAE-Inria, Biogeco-Pléiade), Ghjuvan Grimaud (supervisor, Biomathematica)

Master's students and apprentices

- Juliette Audemard (Univ. Paris Cité, France) - Apprentice (Master 2) supervised by Clémence Frioux
- Juliette Audemard (Univ. Paris Cité, France) - Intern 5 months (Master 1) supervised by Clémence Frioux

Bachelor student

- Juliette Gronier (ENS Lyon, France) - Intern 7 weeks supervised by Clémence Frioux

11.2.3 Juries

PhD defense juries

- Mathieu Bolteau (Univ. Nantes, France) - Clémence Frioux as *examiner*

PhD follow-up committees

- Morgane Roger-Margueritat (TIMC, Univ. Grenoble, France) – Clémence Frioux
- Coralie Rousseau (Sorbonne Univ., Roscoff, France) – Clémence Frioux
- Emma Crisci (Inria-LBBE, Univ. Lyon, France) – Clémence Frioux
- Charles Goedefroit (Inria, Univ. Bordeaux, France) – Clémence Frioux

CRCN and IR selection committees

- INRAE IR selection committee "Multi-omics data analysis" - Simon Labarthe (president)
- INRAE CRCN selection for INRAE GA department - Simon Labarthe.

11.3 Popularization

11.3.1 Specific official responsibilities in science outreach structures

- David Sherman is member of the board (membre du Conseil d'administration) and secretary of the Mobsya Association, Lausanne. Mobsya develops and commercializes the Thymio educational robot, geared towards K-12.
- David Sherman is member of the board (membre du Conseil d'Administration) and lead advisor for software of the Poppy Station Association. Poppy Station develops open-hardware open-source humanoid robots for research and education.

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

UCIA, *Usage et connaissances des intelligences artificielles*, designed by the Ligue de l'Enseignement de la Gironde, Inria, and Poppy Station, is presented to young learners as a serious game combining playing cards, debates and negotiation, and maniulation of a two-wheeled robot capable of image recognition.

11.3.3 Participation in Live events

- Clémence Frioux taught 4 workshops during the "MIMM, moi informaticienne, moi mathématicienne" 2024 week, a free internship at the University of Bordeaux for young girls in 9th and 10th grade in order to encourage them to choose mathematics and computer science, allows them to discover training, research and jobs in these two disciplines.
- Chiche ! Un ou une scientifique, une classe — Clémence Frioux (2 classes), Simon Labarthe (9 classes), David Sherman (3 classes).
- Mathematics Career Days (University of Bordeaux) — Simon Labarthe
- Poppy Rosa UCIA Formation des formateurs (2 sessions, Bazas et Artigues-près-Bordeaux) – David Sherman

11.3.4 Other science outreach relevant activities

- Hosting of *Classe de troisième* student observers (4 students) – David Sherman, Clémence Frioux, Simon Labarthe, Coralie Muller, Leonard Brindel, Sthyve Tatho
- Hosting of *Classe de seconde* student observers (17 students) – David Sherman

12 Scientific production

12.1 Major publications

[1] P. Almeida, C. Gonçalves, S. Teixeira, D. Libkind, M. Bontrager, I. Masneu-Pomarède, W. Albertin, P. Durrens, D. J. Sherman, P. Marullo, C. Todd Hittinger, P. Gonçalves and J. P. Sampaio. 'A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum.' In: *Nature Communications* 5 (2014), p. 4044. DOI: 10.1038/ncomms5044. URL: https://hal.inria.fr/hal-01002466 (cit. on p. 8).

[2] R. Assar, M. A. Montecino, A. Maass and D. J. Sherman. 'Modeling acclimatization by hybrid systems: Condition changes alter biological system behavior models'. In: *BioSystems* 121 (June 2014), pp. 43–53. DOI: 10.1016/j.biosystems.2014.05.007. URL: https://hal.inria.fr/hal-01002987.

[3] M. Bahram, T. Netherway, C. Frioux, P. Ferretti, L. P. Coelho, S. Geisen, P. Bork and F. Hildebrand. 'Metagenomic assessment of the global distribution of bacteria and fungi'. In: *Environmental Microbiology* (13th Nov. 2020). DOI: 10.1111/1462-2920.15314. URL: https://hal.inria.fr/hal-03033570.

[4] A. Belcour, C. Frioux, M. Aite, A. Bretaudeau, F. Hildebrand and A. Siegel. 'Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species'. In: *eLife* 9 (29th Dec. 2020). DOI: 10.1101/803056. URL: https://hal.inria.fr/hal-02395024 (cit. on p. 5).

[5] B. Burgunter-Delamare, H. Kleinjan, C. Frioux, E. Fremy, M. Wagner, E. Corre, A. Le Salver, C. Leroux, C. Leblanc, C. Boyen, A. Siegel and S. Dittami. 'Metabolic Complementarity Between a Brown Alga and Associated Cultivable Bacteria Provide Indications of Beneficial Interactions'. In: *Frontiers in Marine Science* 7 (21st Feb. 2020), pp. 1–11. DOI: 10.3389/fmars.2020.00085. URL: https://hal.inria.fr/hal-02866101.

[6] S. M. Dittami, E. Corre, L. Brillet-Guéguen, A. Lipinska, N. Pontoizeau, M. Aite, K. Avia, C. Caron, C. H. Cho, J. Collen, A. Cormier, L. Delage, S. Doubleau, C. Frioux, A. Gobet, I. González-Navarrete, A. Groisillier, C. Herve, D. Jollivet, H. Kleinjan, C. Leblanc, X. Liu, D. Marie, G. V. Markov, A. E. Minoche, M. Monsoor, P. Péricard, M.-M. Perrineau, A. F. Peters, A. Siegel, A. Siméon, C. Trottier, H. S. Yoon, H. Himmelbauer, C. Boyen and T. Tonon. 'The genome of Ectocarpus subulatus – A highly stress-tolerant brown alga'. In: *Marine Genomics* 52 (Jan. 2020), p. 100740. DOI: 10.1016/j.margen.2020.100740. URL: https://hal.inria.fr/hal-02866117.

[7] C. Frioux, R. Ansorge, E. Özkurt, C. Ghassemi Nedjad, J. Fritscher, C. Quince, S. M. Waszak and F. Hildebrand. 'Enterosignatures define common bacterial guilds in the human gut microbiome'. In: *Cell Host & Microbe* (June 2023). DOI: 10.1016/j.chom.2023.05.024. URL: https://inria.hal.science/hal-04141300.

[8] C. Frioux, S. Dittami and A. Siegel. 'Using automated reasoning to explore the metabolism of unconventional organisms: a first step to explore host–microbial interactions'. In: *Biochemical Society Transactions* 48.3 (7th May 2020), pp. 901–913. DOI: 10.1042/BST20190667. URL: https://hal.archives-ouvertes.fr/hal-02569935 (cit. on p. 5).

[9] C. Frioux, D. Singh, T. Korcsmaros and F. Hildebrand. 'From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes'. In: *Computational and Structural Biotechnology Journal* (June 2020). DOI: 10.1016/j.csbj.2020.06.028. URL: https://hal.inria.fr/hal-02883309.

[10] F. Leese, A. Bouchez, K. Abarenkov, F. Altermatt, A. Borja, K. Bruce, T. Ekrem, F. Ċiampor, Z. Ċiampor, F. Costa, S. Duarte, V. Elbrecht, D. Fontaneto, A. A. Franc, M. Geiger, D. Hering, M. Kahlert, B. Kalamujić Stroil, M. Kelly, E. Keskin, I. Liska, P. Mergen, K. Meissner, J. Pawlowski, L. Penev, Y. Reyjol, A. Rotter, D. Steinke, B. van der Wal, S. S. Vitecek, J. Zimmermann and A. Weigand. 'Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action'. In: *Next Generation Biomonitoring: Part 1*. Vol. 58. Elsevier, 2018, pp. 63–99. URL: https://hal.inria.fr/hal-01984996.

[11] N. D. P. Peyrard, M.-J. Cros, S. De Givry, A. A. Franc, S. S. Robin, R. R. Sabbadin, T. Schiex and M. Vignes. 'Exact or approximate inference in graphical models: why the choice is dictated by the treewidth, and how variable elimination can be exploited'. In: *Australian and New Zealand Journal of Statistics* 61.2 (June 2019). to appear, pp. 89–133. DOI: 10.1111/anzs.12257. URL: https://hal.inria.fr/hal-02433018.

[12] D. J. Sherman, T. Martin, M. Nikolski, C. Cayla, J.-L. Souciet and P. Durrens. 'Genolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes.' In: *Nucleic Acids Research* 37 (2009), pp. D550–D554. DOI: 10.1093/nar/gkn859. URL: https://hal.inria.fr/inria-00341578 (cit. on p. 8).

12.2 Publications of the year

International journals

[13] K. Cerk, P. Ugalde-Salas, C. Ghassemi Nedjad, M. Lecomte, C. Muller, D. J. Sherman, F. Hildebrand, S. Labarthe and C. Frioux. 'Community-scale models of microbiomes: articulating metabolic modelling and metagenome sequencing'. In: *Microbial Biotechnology* (20th Jan. 2024). DOI: 10.1111/1751-7915.14396. URL: https://inria.hal.science/hal-04409251 (cit. on pp. 9, 19).

[14] M. Lecomte, W. Cao, J. Aubert, D. J. Sherman, H. Falentin, C. Frioux and S. Labarthe. 'Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling'. In: *Metabolic Engineering* 83 (May 2024), pp. 24–38. DOI: 10.1016/j.ymben.2024.02.014. URL: https://hal.inrae.fr/hal-04509395 (cit. on pp. 9, 17, 19).

[15] T. Malou and S. Labarthe. 'Pherosensor-toolbox: a Python package for Biology-Informed Data Assimilation'. In: *Journal of Open Source Software* (26th Sept. 2024). DOI: 10.21105/joss.06863. URL: https://hal.inrae.fr/hal-04669546 (cit. on p. 20).

[16] A. Paulay, G. M. Grimaud, R. Caballero, B. Laroche, M. Leclerc, S. Labarthe and E. Maguin. 'Design of a proteolytic module for improved metabolic modeling of Bacteroides caccae'. In: *mSystems* 9.4 (22nd Mar. 2024). DOI: 10.1128/msystems.00153-24. URL: https://hal.inrae.fr/hal-04540116 (cit. on p. 19).

International peer-reviewed conferences

[17] T. Malou, S. Labarthe, B. Laroche, K. Adamczyk and N. Parisey. 'Pest detection from a biology-informed inverse problem and pheromone sensors'. In: DSABNS 2024 - 15th Conference on Dynamical Systems Applied to Biology and Natural Sciences. Lisbonne, Portugal, 6th Feb. 2024. URL: https://hal.science/hal-04481765 (cit. on p. 20).

Conferences without proceedings

[18] F. Delmotte, H. Hannin, C. Vacher, S. Labarthe, P. Mestre, A. Rusch, A. Mérot, L. Latruffe, E. Giraud-Héraud, L. Deliere, P. Darriet, C. Pretet-Lataste and M. Adrian. 'Pesticide-Free viticulture: towards agroecological wine-producing socioecosystems'. In: 45ème congrès mondial de la Vigne et du Vin. Dijon, France, 14th Oct. 2024. URL: https://hal.science/hal-04732866 (cit. on p. 20).

[19] N. Maurice, C. Frioux, C. Lemaitre and R. Vicedomini. 'Assessing assembly quality in metagenomes of increasing complexity sequenced with HiFi long reads'. In: 2024 - 24th Genome Informatics meeting. Hinxton, United Kingdom, Nov. 2024, pp. 1–21. URL: https://hal.science/hal-04822870 (cit. on p. 18).

Doctoral dissertations and habilitation theses

[20] S. Labarthe. 'Contributions to mathematical modelling in microbial ecology, from data to models'. Université de bordeaux, 10th July 2024. URL: https://hal.inrae.fr/tel-04654712 (cit. on p. 9).

[21] M. Lecomte. 'Hybrid approach for explainable metabolic modelling of microbial ecosystems'. Université de Bordeaux, 29th Apr. 2024. URL: https://theses.hal.science/tel-04583312 (cit. on p. 9).

Reports & preprints

[22] M. A. Abouabdallah, O. Coulaud, N. Peyrard and A. Franc. *Computing WSBM marginals with Tensor-Train decomposition*. 15th Jan. 2024. URL: https://hal.inrae.fr/hal-04394024.

[23] C. M. Andreani-Gerard, N. E. Jiménez, R. Palma, C. Muller, P. Hamon-Giraud, Y. Le Cunff, V. Cambiazo, M. González, A. Siegel, C. Frioux and A. Maass. *Modeling the emergent metabolic potential of soil microbiomes in Atacama landscapes*. 2024. DOI: 10.1101/2024.12.23.630026. URL: https://hal.science/hal-04854948 (cit. on p. 18).

[24] A. Belcour, P. Hamon-Giraud, A. Mataigne, B. Ruiz, Y. L. Cunff, J. Got, L. Awhangbo, M. Lebreton, C. Frioux, S. Dittami, P. Dabert, A. Siegel and S. Blanquart. *Estimating consensus proteomes and metabolic functions from taxonomic affiliations*. 2025. DOI: 10.1101/2022.03.16.484574. URL: https://hal.science/hal-03697249.

[25] C. Ghassemi Nedjad, M. Bolteau, L. Bourneuf, L. Pauleve and C. Frioux. *Seed2LP: seed inference in metabolic networks for reverse ecology applications*. 27th Sept. 2024. DOI: 10.1101/2024.09.26.615309. URL: https://hal.science/hal-04713829 (cit. on p. 18).

[26] T. Malou, N. Parisey, K. Adamczyk-Chauvat, E. Vergu, B. Laroche, P.-A. Calatayud, P. Lucas and S. Labarthe. *Biology-Informed inverse problems for insect pests detection using pheromone sensors*. 28th Jan. 2025. DOI: 10.24072/pci.mcb.100313. URL: https://hal.inrae.fr/hal-04572831 (cit. on p. 20).

[27] K. Pärnänen, M. Ruuskanen, G. Sommeria-Klein, V. Laitinen, P. Kantanen, G. Méric, C. Gazolla Volpiano, M. Inouye, R. Knight, V. Salomaa, A. S. Havulinna, T. Niiranen and L. Lahti. *Population variation and prognostic potential of gut antibiotic resistome*. 8th Aug. 2024. DOI: 10.1101/2024.08.08.24311663. URL: https://hal.science/hal-04908728 (cit. on p. 21).

[28] C. Ross, V. Laitinen, M. Khalighi, J. Salojärvi, W. de Vos, G. Sommeria-Klein and L. Lahti. *Reconstructing community dynamics from limited observations*. 2025. DOI: 10.48550/arXiv.2501.03820. URL: https://hal.science/hal-04908644 (cit. on p. 21).

Other scientific publications

[29] J. Audemard, S. Halary, G. Markov, J. Got, A. Siegel, M. Lefebvre, J. Leloup, B. Marie, N. Creusot, B. Dieme and C. Frioux. 'Metagenome-scale metabolic modelling for the characterization of cross-feeding interactions in Microcystis-associated microbial communities, in the context of freshwater cyanobacterial blooms'. In: JOBIM 2024 - Journées Ouvertes en Biologie, Informatique et Mathématiques. Toulouse, France, 2024, pp. 1–1. URL: https://inria.hal.science/hal-04654522.

[30] C. Belliardo, N. Maurice, C. Frioux, C. Lemaitre, R. Vicedomini, S. Mondy, A. Péré, M. Bailly-Bechet and E. Danchin. 'Unlocking the Soil Microbiome: Unraveling Soil Microbial Complexity Using Long-Read Metagenomics'. In: EAGS 2024 - The International Environmental and Agronomical Genomics symposium. Toulouse, France, 2024, pp. 1–1. URL: https://hal.science/hal-04509213 (cit. on p. 18).

[31] C. Belliardo, S. Mondy, A. Pere, C. Lemaitre, R. Vicedomini, C. Frioux, D. J. Sherman, P. Abad, M. Bailly-Bechet and E. Danchin. 'Exploring and quantifying the soil genetic diversity captured by long and short-read shotgun metagenomic sequencing'. In: Journées 2024 du programme Agroécologie et Numérique. Rennes, France, 2024, pp. 1–1. URL: https://hal.science/hal-04423917.

[32] C. Belliardo, A. Péré, A. Franc, J.-M. Frigerio, N. Maurice, C. Frioux, C. Lemaitre, S. Mondy, M. Bailly-Bechet, P. Abad, D. J. Sherman and E. G. Danchin. 'Enhancing Microbial Genome Reconstruction in Complex Environments by combining Short-and Long-read Sequencing'. In: 2025 - Journées scientifiques Agroécologie et Numérique. Dijon, France, 29th Jan. 2025, pp. 1–1. URL: https://hal.science/hal-04920790.

[33] N. Maurice, C. Frioux, C. Lemaitre and R. Vicedomini. 'Mapler: Assessing assembly quality in taxonomically-rich metagenomes sequenced with HiFi reads'. In: 2025 - Journées scientifiques Agroécologie et Numérique. Dijon, France, 2025, pp. 1–1. URL: https://hal.science/hal-04941137.

[34] N. Maurice, C. Lemaitre, C. Frioux and R. Vicedomini. 'Metagenomic assembly of complex ecosystems with highly accurate long-reads'. In: Journées 2024 du PEPR Agroécologie et Numérique. Rennes, France, 2024, pp. 1–1. URL: https://inria.hal.science/hal-04425626.

[35] S. J. Tatho Djeanou, S. Labarthe and V. Baldazzi. 'cMFA for multi-omics data integration in microbial community models'. In: CJC MA2024 - Congrès des Jeunes Chercheur.e.s en Mathématiques Appliquées. Lyon, France, 28th Oct. 2024. URL: https://hal.science/hal-04874275 (cit. on p. 20).

[36] S. J. Tatho Djeanou, S. Labarthe and V. Baldazzi. 'cMFA for multi-omics data integration in microbial community models'. In: Journées scientifiques 2025 Agroécologie et Numérique. Dijon, France, 28th Jan. 2025. URL: https://inria.hal.science/hal-04939529.

Scientific popularization

[37] C. Muller, S. Prigent and C. Frioux. 'Generation of metabolomic-informed models of metabolism in complex microbial communities'. In: Journées scientifiques Agroécologie et Numérique 2025. Dijon, France, 28th Jan. 2025. URL: https://inria.hal.science/hal-04919231.

[38] S. J. Tatho Djeanou, S. Labarthe and V. Baldazzi. 'cMFA for multi-omics data integration in microbial community models'. In: 16th Dynamical Systems applied on Biology and Natural Sciences | DSABNS 2025. Naples, Italy, 20th Jan. 2025. URL: https://inria.hal.science/hal-04939527.

Software

[39] [SW] J. Aubert, H. Falentin, C. Frioux, S. Labarthe, M. Lecomte and D. J. Sherman, *TANGO models: numerical multi-omics reconciliation of bacterial fermentation in cheese production* version 0.9.0, 15th Mar. 2024. Inria; INRAE. LIC: GNU General Public License v3.0. HAL: ⟨hal-04562266⟩, URL: https://inria.hal.science/hal-04562266, VCS: https://forgemia.inra.fr/tango/tango_models, SWHID: ⟨swh:1:dir:0a1da28c523c2f32e1214b5d1c1a60ac17ec29df;origin=https://forgemia.inra.fr/tango/tango_models.git;visit=swh:1:snp:722365d9bf1004d05edfe63717763283d4adf26f;anchor=swh:1:rev:bc0bd9895942ed6d82faa9553c1f0a32f60cbbef⟩ (cit. on pp. 9, 19).

12.3 Cited publications

[40] G. Ravel, M. Bergmann, A. Trubuil, J. Deschamps, R. Briandet and S. Labarthe. 'Inferring characteristics of bacterial swimming in biofilm matrix from time-lapse confocal laser scanning microscopy'. In: *eLife* 11 (14th June 2022). DOI: 10.7554/eLife.76513. URL: https://hal.inrae.fr/hal-03695580 (cit. on p. 6).

[41] M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron, N. Guillaudeux, M. Latorre, N. Loira, G. V. Markov, A. Maass and A. Siegel. 'Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models'. In: *PLOS Computational Biology* 14.5 (2018), e1006146. DOI: 10.1371/journal.pcbi.1006146 (cit. on p. 5).

[42] B. Arnold, R. Corbett-Detig, D. Hartl and K. Bomblies. 'RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling'. In: *Mol. Ecol.* 22.11 (2013), pp. 3179–90 (cit. on p. 8).

[43] M. Bakonyi and C. R. Johnson. 'The Euclidean Distance Matrix Completion Problem'. In: *SIAM J. Matrix Anal. App.* 16.2 (1995), pp. 646–654 (cit. on p. 3).

[44] M. Bourgin, S. Labarthe, A. Kriaa, M. Lhomme, P. Gérard, P. Lesnik, B. Laroche, E. Maguin and M. M. Rhimi. 'Exploring the Bacterial Impact on Cholesterol Cycle: A Numerical Study'. In: *Frontiers in Microbiology* 11 (2020), p. 1121. DOI: 10.3389/fmicb.2020.01121. URL: https://hal.science/hal-02863236 (cit. on p. 5).

[45] E. J. Candès and B. Recht. 'Exact Matrix Completion via Convex Optimization'. In: *Found. Comput. Math.* 9 (2009), pp. 717–772 (cit. on p. 3).

[46] W. Cao, M. Lecomte, S. Le Fur, J. J. Aubert, M.-B. Maillard, A. Nicolas, S.-M. Deutsch, S. Parayre, F. Boissel, A. Leduc, A. Kerjouh, M. Harel-Oger, G. Garric, C. Frioux, D. J. Sherman, S. Labarthe, A. Thierry and H. Falentin. 'Metatranscriptomics and metabolic modeling to identify bacterial metabolic interactions during the manufacture of a model pressed cheese.' In: *FOODMICRO 2022*. Athènes, Greece, Aug. 2022. URL: https://hal.inrae.fr/hal-03781287 (cit. on p. 17).

[47] C. Combes. *Parasitism: The Ecology and Evolution of Intimate Interactions*. University of Chicago Press, 2001 (cit. on p. 8).

[48] 'Comparative genomics of protoploid Saccharomycetaceae'. In: *Genome Research* 19 (2009), pp. 1696–1709. DOI: 10.1101/gr.091546.109. URL: http://hal.inria.fr/inria-00407511/en/ (cit. on p. 7).

[49] M.-J. Cros, J.-M. Frigerio, N. Peyrard and A. Franc. 'Simple approaches for evaluation of OTU quality based on dissimilarity arrays'. working paper or preprint. Oct. 2022. URL: https://hal.science/hal-03824588 (cit. on p. 23).

[50] B. Dujon, D. Sherman, G. Fischer, P. Durrens et al. 'Genome evolution in yeasts'. In: *Nature* 430 (2004), pp. 35–44. DOI: 10.1038/nature02579. URL: http://www.nature.com/nature/journal/v430/n6995/abs/nature02579.html (cit. on p. 7).

[51] O. Ebenhöh, T. Handorf and R. Heinrich. 'Structural analysis of expanding metabolic networks.' In: *Genome informatics. International Conference on Genome Informatics* 15.1 (2004), pp. 35–45 (cit. on p. 5).

[52] A. Franc. *Linear Dimensionality Reduction*. Research Report 9488. Inria Bordeaux Sud-Ouest, May 2023, p. 99. URL: https://inria.hal.science/hal-03784623 (cit. on p. 23).

[53] A. A. Franc, P. Blanchard and O. Coulaud. 'Nonlinear mapping and distance geometry'. In: *Optimization Letters* 14.2 (May 2019), pp. 453–467. DOI: 10.1007/s11590-019-01431-y. URL: https://hal.inria.fr/hal-02124882 (cit. on p. 4).

[54] C. Frioux, E. Fremy, C. Trottier and A. Siegel. 'Scalable and exhaustive screening of metabolic functions carried out by microbial consortia'. In: *Bioinformatics* 34.17 (2018), pp. i934–i943. DOI: 10.1093/bioinformatics/bty588 (cit. on p. 5).

[55] C. Frioux, T. Schaub, S. Schellhorn, A. Siegel and P. Wanko. 'Hybrid metabolic network completion'. In: *Theory and Practice of Logic Programming* 19.1 (2019), pp. 83–108. DOI: 10.1017/s1471068418000455 (cit. on p. 5).

[56] P. Gayral, J. Melo-Ferreira and S. Glemin. 'Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap'. In: *PLoS Genetic* 9.4 (2013). e1003457 (cit. on p. 8).

[57] M. Gebser, R. Kaminski, B. Kaufmann and T. Schaub. 'Clingo = ASP + Control: Preliminary Report'. In: *CoRR* abs/1405.3694 (2014) (cit. on p. 5).

[58] M. Gebser, R. Kaminski, A. Konig and T. Schaub. 'Advances in gringo Series 3'. In: *LPNMR*. Vol. 6645. Lecture Notes in Computer Science. Springer, 2011, pp. 345–351 (cit. on p. 5).

[59] 'Génolevures – a novel approach to 'evolutionary genomics''. In: *FEBS Letters (complete issue)* 487.1 (2000). URL: https://febs.onlinelibrary.wiley.com/toc/18733468/2000/487/1 (cit. on p. 7).

[60] S. Labarthe, B. Polizzi, T. Phan, T. Goudon, M. Ribot and B. Laroche. 'A mathematical model to investigate the key drivers of the biogeography of the colon microbiota.' In: *Journal of Theoretical Biology* 462.7 (2019), pp. 552–581. DOI: 10.1016/j.jtbi.2018.12.009. URL: https://hal.science/hal-01761191 (cit. on p. 5).

[61] M. Lecomte, W. Cao, J. J. Aubert, D. J. Sherman, H. Falentin, C. Frioux and S. Labarthe. 'Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling'. working paper or preprint. May 2023. URL: https://hal.inrae.fr/hal-04088301 (cit. on p. 17).

[62] M. Lecomte, D. J. Sherman, H. Falentin and C. Frioux. *Characterising competition and cooperation potentials in microbial communities using discrete models of metabolism. An exploratory analysis*. JOBIM 2022 - JOURNÉES OUVERTES EN BIOLOGIE, INFORMATIQUE ET MATHÉMATIQUES. Poster. July 2022. URL: https://hal.inrae.fr/hal-03839337 (cit. on p. 17).

[63] M. Lecomte, D. J. Sherman, H. Falentin and C. Frioux. 'Reasoning approaches for the characterization of cooperation and competition in large-scale microbial communities'. In: *Journée GT Bioss 2022*. Nantes, France, Nov. 2022. URL: https://hal.inrae.fr/hal-03857781 (cit. on p. 17).

[64] L. Liberti, C. Lavor, N. Maculan and A. Mucherino. 'Euclidean Distance Geometry and Applications'. In: *SIAM review* 56(1) (2014), pp. 3–69 (cit. on p. 3).

[65] M. Lynch. 'Estimation of Nucleotide Diversity, Disequilibrium Coefficients, and Mutation Rates from High-Coverage Genome-Sequencing Projects'. In: *Mol. Biol. Evol.* 25.11 (2008), pp. 2409–19 (cit. on p. 8).

[66] J. D. Orth, I. Thiele and B. Ø. Palsson. 'What is flux balance analysis?' In: *Nature Biotechnology* 28.3 (2010), pp. 245–248. DOI: 10.1038/nbt.1614 (cit. on p. 5).

[67] B. Perez-Lamarque, G. Sommeria-Klein, L. Duret and H. Morlon. 'Phylogenetic Comparative Approach Reveals Evolutionary Conservatism, Ancestral Composition, and Integration of Vertebrate Gut Microbiota'. In: *Molecular Biology and Evolution* 40.7 (July 2023). DOI: 10.1093/molbev/msad144. URL: https://hal.science/hal-04294906 (cit. on p. 22).

[68] S. Prigent, C. Frioux, S. M. Dittami, S. Thiele, A. Larhlimi, G. Collet, F. Gutknecht, J. Got, D. Eveillard, J. Bourdon, F. Plewniak, T. Tonon and A. Siegel. 'Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks'. In: *PLOS Computational Biology* 13.1 (2017), e1005276. DOI: 10.1371/journal.pcbi.1005276 (cit. on p. 5).

[69] R. E. Ricklefs. 'A comprehensive framework for global patterns in biodiversity'. In: *Ecology Letters* 7.1 (2004), pp. 1–15. DOI: 10.1046/j.1461-0248.2003.00554.x. URL: http://dx.doi.org/10.1046/j.1461-0248.2003.00554.x (cit. on p. 8).

[70] D. J. Sherman, T. Martin, M. Nikolski, C. Cayla, J.-L. Souciet and P. Durrens. 'Génolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes'. In: *Nucleic Acids Research* 37.suppl 1 (2009), pp. D550–D554. DOI: 10.1093/nar/gkn859. eprint: http://nar.oxfordjournals.org/content/37/suppl_1/D550.full.pdf+html. URL: http://nar.oxfordjournals.org/content/37/suppl_1/D550.abstract (cit. on p. 7).

[71] D. J. Sherman, P. Durrens, F. Iragne, E. Beyne, M. Nikolski and J.-L. Souciet. 'Genolevures complete genomes provide data and tools for comparative genomics of hemiascomycetous yeasts.' In: *Nucleic Acids Research* 34.Database issue (Jan. 2006), pp. D432–5. DOI: 10.1093/nar/gkj160. URL: https://hal.archives-ouvertes.fr/hal-00118142 (cit. on p. 7).

[72] B. de Thoisy, O. Duron, L. Epelboin, L. Musset, P. Quénel, B. Roche, F. Binetruy, S. Briolant, L. Carvalho, A. Chavy, P. Couppié, M. P. Demar, M. Douine, I. Dusfour, S. Gourbière, Y. Epelboin, C. Flamand, A. Franc, M. Ginouvès, E. S. Houël, A. Kocher, A. Lavergne, P. Le Turnier, L. Mathieu, J. Murienne, M. Nacher, R. Schaub, S. Pelleau, G. Prévot, D. Rousset, E. Roux, S. Talaga, P. Thill, S. Tirera and J.-F. Guégan. 'Ecology, evolution, and epidemiology of zoonotic and vector-borne infectious diseases in French Guiana: Transdisciplinarity does matter to tackle new emerging threats'. In: *Infection, Genetics and Evolution* 93 (Sept. 2021), p. 104916. DOI: 10.1016/j.meegid.2021.104916. URL: https://hal-pasteur.archives-ouvertes.fr/pasteur-03261181 (cit. on p. 5).

[73] S. Tirera, B. de Thoisy, D. Donato, C. Bouchier, V. Lacoste, A. Franc and A. Lavergne. 'The influence of habitat on viral diversity in neotropical rodent hosts'. In: *Viruses* 13.9 (Aug. 2021), pp. 1–29. DOI: 10.3390/v13091690. URL: https://hal.inrae.fr/hal-03370479 (cit. on p. 4).