UMR IRISA

Activity Report 2025

Team DYLISS

Dynamics, Logics and Inference for biological Systems and Sequences

Joint team with Centre Inria de l'Université de Rennes

D7 – Data and Knowledge Management

Université de Rennes | rennes | INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES | cnrs | Inria | Inserm | CentraleSupélec | IMT Atlantique Bretagne-Pays de la Loire École Mines-Télécom | Université Bretagne Sud ubs

Team DYLISS

Creation of the Team: 2025 July 01

Keywords

Computer sciences and digital sciences

A3.1.1. – Modeling, representation

A3.1.2. – Data management, quering and storage

A3.1.6. – Query optimization

A3.1.7. – Open data

A3.1.8. – Big data (production, storage, transfer)

A3.1.11. – Structured data

A3.2.1. – Knowledge bases

A3.2.2. – Knowledge extraction, cleaning

A3.2.4. – Semantic Web

A3.2.5. – Ontologies

A3.4. – Machine learning and statistics

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

A7.3.1. – Computational models and calculability

A9.1. – Knowledge

A9.2. – Machine learning

Other research topics and application domains

B1.1.2. – Molecular and cellular biology

B1.1.4. – Genetics and genomics

B1.1.7. – Bioinformatics

B1.1.10. – Systems and synthetic biology

B2.2.3. – Cancer

B2.2.5. – Immune system diseases

Contents

Team DYLISS **1**

1 Team members, visitors, external collaborators **4**

2 Overall objectives **5**

3 Research program **6**
3.1 Context: Computer science perspective on symbolic artificial intelligence 6
3.2 Scalable methods to query data heterogenity 6
3.2.1 Research topics 7
3.2.2 Associated software tools 7
3.3 Metabolism: from protein sequences to systems ecology 8
3.3.1 Research topics 8
3.3.2 Associated software tools 8
3.4 Regulation and signaling: detecting complex and discriminant signatures of phenotypes . . 9
3.4.1 Research topics 9
3.4.2 Associated software tools 9

4 Application domains **10**

5 Social and environmental responsibility **12**
5.1 Footprint of research activities 12
5.2 Impact of research results 12

6 Highlights of the year **12**
6.1 Members 12
6.2 Broadening International Visibility 13

7 Latest software developments, platforms, open data **13**
7.1 Latest software developments 13
7.1.1 AskOmics 13
7.1.2 Metage2Metabo 13
7.1.3 AuCoMe 14
7.1.4 prolipipe 14
7.1.5 EnzBert-GO 15
7.1.6 FUSE-PhyloTree 15

8 New results **16**
8.1 Scalable methods to query data heterogeneity 16
8.2 Metabolism: from protein sequences to systems ecology 16
8.3 Regulation and signaling: detecting complex and discriminant signatures of phenotypes . . 18

9 Bilateral contracts and grants with industry **19**
9.1 Bilateral Grants with Industry 19

10 Partnerships and cooperations **19**
10.1 International initiatives 19
10.1.1 Visits of international scientists 19
10.1.2 Visits to international teams 20
10.2 European initiatives 21
10.2.1 Other european programs/initiatives 21
10.3 National initiatives 21
10.3.1 Programs funded by Inria 23

11 Dissemination **24**
11.1 Promoting scientific activities . . . 24
11.1.1 Scientific events: organisation . . . 24
11.1.2 Scientific events: selection . . . 24
11.1.3 Journal . . . 24
11.1.4 Invited talks . . . 25
11.1.5 Leadership within the scientific community . . . 25
11.1.6 Scientific expertise . . . 25
11.1.7 Research administration . . . 26
11.2 Teaching - Supervision - Juries . . . 26
11.2.1 Teaching . . . 26
11.2.2 Supervision . . . 28
11.2.3 Doctoral advisory committees (CSID) . . . 29
11.2.4 Juries . . . 30
11.3 Popularization . . . 30
11.3.1 Participation in Live events . . . 30

12 Scientific production **30**
12.1 Major publications . . . 30
12.2 Publications of the year . . . 31
12.3 Cited publications . . . 33

1 Team members, visitors, external collaborators

Research Scientists

- Anne Siegel [Team leader, CNRS, Senior Researcher, until Jun 2025, HDR]
- Samuel Blanquart [INRIA, Researcher]
- François Coste [INRIA, Researcher]
- Anne Siegel [CNRS, Senior Researcher, from Jul 2025, HDR]
- Nathalie Theret [INSERM, Senior Researcher, HDR]

Faculty Members

- Emmanuelle Becker [Team leader, UNIV RENNES, Professor, from Jul 2025, HDR]
- Emmanuelle Becker [UNIV RENNES, Professor, until Jun 2025, HDR]
- Catherine Belleannée [UNIV RENNES, Associate Professor]
- Myriam Bontonou [UNIV RENNES, ATER, until Aug 2025]
- Olivier Dameron [UNIV RENNES, Professor, HDR]
- Yann Le Cunff [UNIV RENNES, Associate Professor, HDR]

PhD Students

- Moana Aulagner [INRIA]
- Cecile Beust [UNIV RENNES]
- Oceane Carpentier [UNIV RENNES]
- Elisa Chenel [UNIV RENNES]
- Pablo Espana Gutierrez [UNIV RENNES]
- Juliette Francis [UNIV RENNES]
- Pauline Giraud [UNIV RENNES]
- Ulysse Le Clanche [UNIV RENNES]
- Corentin Lucas [INRIA]
- Noe Robert [UNIV RENNES, from Nov 2025]
- Noryah Safla [INSERM, from Nov 2025]
- Yael Tirlet [UNIV RENNES]

Technical Staff

- Jeanne Got [CNRS, Engineer]
- Alice Mataigne [CNRS, Engineer, until Jul 2025]
- Noe Robert [CNRS, Engineer, until Jun 2025]

Interns and Apprentices

- Daniel Calvez [CNRS, Intern, from Apr 2025 until Jul 2025]
- Domenico Palladino [INRIA, Intern, from Nov 2025]
- Noryah Safla [INSERM, Intern, until Jul 2025]

Administrative Assistant

- Marie Le Roïc [INRIA]

2 Overall objectives

Bioinformatics context: from life data science to functional information about biological systems and unconventional species. Sequence analysis and systems biology both consist in the interpretation of biological information at the molecular level, that concern mainly intra-cellular compounds. Analyzing genome-level information is the main issue of **sequence analysis**. The ultimate goal here is to build a full catalogue of bio-products together with their functions, and to provide efficient methods to characterize such bio-products in genomic sequences. In regards, contextual physiological information includes all cell events that can be observed when a perturbation is performed over a living system. Analyzing contextual physiological information is the main issue of **systems biology**.

For a long time, computational methods developed within sequence analysis and dynamical modeling had few interplay. However, the emergence and the democratization of new sequencing technologies (NGS, metagenomics) provides information to link systems with genomic sequences. In this research area, the Dyliss team focuses on linking genomic sequence analysis and systems biology. **Our main applicative goal in biology is to characterize groups of genetic actors that control the phenotypic response of species when challenged by their environment**. **Our main computational goals are to develop methods for analyzing the dynamical response of a biological system, modeling and classifying families of gene products with sensitive and expressive languages, and identifying the main actors of a biological system within static interaction maps**. We first formalize and integrate in a set of logical or grammatical constraints both generic knowledge information (literature-based regulatory pathways, diversity of molecular functions, DNA patterns associated with molecular mechanisms) and species-specific information (physiological response to perturbations, sequencing...). We then rely on symbolic methods (Semantic Web technologies for data integration, querying as well as for reasoning with bio-ontologies, solving combinatorial optimization problems, formal classification) to compute the main features of the space of admissible models.

Computational challenges. The main challenges we face are **data incompleteness and heterogeneity, leading to non-identifiability**. Indeed, we have observed that the biological systems that we consider cannot be uniquely identifiable. Indeed, "omics" technologies have allowed the number of measured compounds in a system to increase tremendously. However, it appears that the theoretical number of different experimental measurements required to integrate these compounds in a single discriminative model has increased exponentially with respect to the number of measured compounds. Therefore, according to the current state of knowledge, there is no possibility to explain the data with a single model. Our rationale is that biological systems will still remain non-identifiable for a very long time. In this context, we favor **the construction and the study of a space of feasible models or hypotheses**, including known constraints and facts on a living system, rather than searching for a single discriminative optimized model. We develop methods allowing a precise and exhaustive investigation of this space of hypotheses. With this strategy, we are in the position of developing experimental strategies to progressively shrink the space of hypotheses and increase the understanding of the system.

Bioinformatics challenges. Our objectives in computer sciences are developed within the team in order to fit with three main bioinformatics challenges (1) data-science and knowledge-science for life sciences (see Section 3.2); (2) understanding metabolism (see Section 3.3); (3) characterizing regulatory and signaling phenotypes (see Section 3.4).

Implementing methods in software and platforms. Seven platforms have been developed in the team during the last five years: Askomics, AuReMe, FinGoc, Caspo, Cadbiom, Logol and Protomata. They aim at guiding the user to progressively reduce the space of models (families of sequences of genes or proteins, families of keys actors involved in a system response or dynamical models) which are compatible with both the knowledge and experimental observations. Most of our platforms are developed with the support of the GenOuest resource and data center hosted in the IRISA laboratory, including their computer facilities [More info]

3 Research program

3.1 Context: Computer science perspective on symbolic artificial intelligence

We develop methods that use an explicit representation of the relationships between heterogeneous data and knowledge in order to construct a space of hypotheses. Therefore, our objective in computer science is mainly to develop accurate representations (oriented graphs, Boolean networks, automata, or expressive grammars) to iteratively capture the complexity of a biological system.

Integrating data with querying languages: Semantic web for life sciences The first level of complexity in the data integration process consists in confronting heterogeneous datasets. Both the size and the heretogeneity of life science data make their integration and analysis by domain experts impractical and prone to the streetlight effect (they will pick up the models that best match what they know or what they would like to discover). Our first objective involves the formalization and management of symbolic knowledge, that is, the explicitation of relations occurring in structured data. In this setting, our main goal is to facilitate and optimize the integration of Semantic Web resources with local users data by relying on the implicit data scheme contained in biological data and Semantic Web resources.

Reasoning over structured data with constraint-based logical paradigms Another level of complexity in life science integration is that very few paradigms exist to model the behavior of a complex biological system. This leads biologists to perform and formulate hypotheses in order to interpret their data. Our strategy is to interpret such hypotheses as combinatorial optimization problems, allowing to reduce the family of models compatible with data. To that goal, we collaborate with Potsdam University in order to use and challenge the most recent developments of Answer Set Programming (ASP) [53], a logical paradigm for solving constraint satisfiability and combinatorial optimization issues.

Our goal is therefore to provide scalable and expressive formal models of queries on biological networks with the focus of integrating dynamical information as explicit logical constraints in the modeling process.

Characterizing biological sequences with formal syntactic models Our last goal is to identify and characterize the function of expressed genes such as transcripts, enzymes or isoforms in non-model species biological networks or specific functional features of metagenomic samples. These are insufficiently precise because of the divergence of biological sequences, the complexity of molecular structures and biological processes, and the weak signals characterizing these elements.

Our goal is therefore to develop accurate formal syntactic models (automata, grammars or abstract gene models) that would enable us to represent sequence conservation, sets of short and degenerated patterns, and crossing or distant dependencies. This requires both to determine the classes of formal syntactic models adequate for handling biological complexity, and to automatically characterize the functional potential embodied in biological sequences with these models.

3.2 Scalable methods to query data heterogenity

Confronted to large and complex data sets (raw data are associated with graphs depicting explicit or implicit links and correlations) almost all scientific fields have been impacted by the *big data issue*, especially genomics and astronomy [65]. In our opinion, life sciences cumulate several features that are very specific and prevent the direct application of big data strategies that proved successful in other domains such as experimental physics: the existence of **several scales of granularity** (from microscopic to macroscopic) and the associated issue of dependency propagation, datasets **incompleteness and uncertainty** (including highly **heterogeneous** responses to a perturbation from one sample to another), and highly fragmented sources of

information that **lacks interoperability** [51]. To explore this research field, we use techniques from symbolic data mining (Semantic Web technologies, symbolic clustering, constraint satisfaction, and grammatical modeling) to take into account those life science features in the analysis of biological data.

3.2.1 Research topics

Facilitating data integration and querying The quantity and inner complexity of life science data require semantically-rich analysis methods. A major challenge is then to combine data (from local project as well as from reference databases) and symbolic knowledge seamlessly. Semantic Web technologies (RDF for annotating data, OWL for representing symbolic knowledge, and SPARQL for querying) provide a relevant framework, as demonstrated by the success of Linked (Open) Data [33]. However, life science end users (1) find it difficult to learn the languages for representing and querying Semantic Web data, and consequently (2) miss the possibility they had to interact with their tabulated data (even when doing so was exceedingly slow and tedious). Our first objective in this axis is to develop accurate abstractions of datasets or knowledge repositories to facilitate their exploration with RDF-based technologies.

Scalability of semantic web queries. A bottleneck in data querying is given by the performance of federated SPARQL queries, which must be improved by several orders of magnitude to allow current massive data to be analyzed. In this direction, our research program focuses on the combination of *linked data fragments* [71], query properties and dataset structure for decomposing federated SPARQL queries.

Building and compressing static maps of interacting compounds A final approach to handle heterogeneity is to gather multi-scale data knowledge into a functional static map of biological models that can be analyzed and/or compressed. This requires to link genomics, metabolomics, expression data and protein measurement of several phenotypes into unified frameworks. In this direction, our main goal is to develop families of constraints, inspired by symbolic dynamical systems, to link datasets together. We currently focus on health (personalized medicine) and environmental (role of non-coding regulations, graph compression) datasets.

3.2.2 Associated software tools

AskOmics platform AskOmics is an integration and interrogation software for linked biological data based on semantic web technologies[1]. AskOmics aims at bridging the gap between end user data and the Linked (Open) Data cloud (LOD cloud). It allows heterogeneous bioinformatics data (formatted as tabular files or directly in RDF) to be loaded into a Triple Store system using a user-friendly web interface. It helps end users (1) to take advantage of the information available in the LOD cloud for analyzing their own data, and (2) to contribute back to the linked data by representing their data and the associated metadata in the proper format, as well as by linking them to other resources. An originality is the graphical interface that allows any dataset to be integrated in a local RDF datawarehouse and SPARQL query to be built transparently and iteratively by a non-expert user.
Pax2graphml aims at easily manipulating BioPAX source files as regulated reaction graphs described in graph format. The goal is to be highly flexible and to integrate graphs of regulated reactions from a single BioPAX source or by combining and filtering BioPAX sources. The output graphs can then be analyzed with additional tools developed in the team, such as KeyRegulatorFinder.

FinGoc-tools The FinGoc tools allow filtering interaction networks with graph-based optimization criteria in order to elucidate the main regulators of an observed phenotype. The main added-value of these tools is the functionality allowing to make explicit the criteria used to highlight the role of the main regulators. (1) The KeyRegulatorFinder package searches key regulators of lists of molecules (like metabolites, enzymes or genes) by taking advantage of knowledge databases in cell metabolism and signaling[2]. (2) The PowerGrasp python package implements graph compression methods oriented toward visualization, and based on power graph analysis[3]. (3) The iggy package enables the repairing of an interaction graph with respect to expression data[4].

[1] askomics.org
[2] biowic.inria.fr/
[3] github.com/aluriak/powergrasp
[4] bioasp.github.io/iggy/

3.3 Metabolism: from protein sequences to systems ecology

Our research in bioinformatics in relation with metabolic processes is driven by the need to understand non-model (eukaryote) species. Their metabolism have acquired specific features that we wish to identify with computational methods. To that goal, we combine sequence analysis with metabolic network analysis, with the final goal to understand better the metabolism of communities of organisms.

3.3.1 Research topics

Genomic level: characterizing functions of protein sequences Precise characterization of functional proteins, such as enzymes or transporters, is a key to better understand and predict the actors involved in a metabolic process. In order to improve the precision of functional annotations, we develop machine learning approaches that take a sample of functional sequences as input and infer a model representing their key syntactical characteristics, including dependencies between residues.

System level: enriching and comparing metabolic networks for non-model organisms

Non-model organisms often lack both complete and reliable annotated sequences, which cause the draft networks of their metabolism to largely suffer from incompleteness. In former studies, the team has developed several methods to improve the quality of eukaryotic metabolic networks, by solving several variants of the so-called *Metabolic Network gap-filling problem* with logical programming approaches [10, 9]. The main drawback of these approaches is that they cannot scale to the reconstruction and comparison of families of metabolic networks. Our main objective is therefore to develop new tools for the comparison of species strains at the metabolic level.

Consortium level: exploring the diversity of community consortia The newly emerging field of system ecology aims at building predictive models of species interactions within an ecosystem, with the goal of deciphering cooperative and competitive relationships between species [50]. This field raises two new issues: (1) uncertainty on the species present in the ecosystem and (2) uncertainty about the global objective governing an ecosystem. To address these challenges, our first research focus is the inference of metabolic exchanges and relationships for transporter identification, based on our expertise in metabolic network gap-filling. The second challenging focus is the prediction of transporters families via refined characterization of transporters, which are quite unexplored apart from specific databases [63].

3.3.2 Associated software tools

Protomata[5] is a machine learning suite for the inference of automata characterizing (functional) families of proteins at the sequence level. It provides programs to build a new kind of sequence alignments (characterized as partial and local), learn automata, and search for new family members in sequence databases. By enabling to model local dependencies between positions, automata are more expressive than classical tools (PSSMs, Profile HMMs, or Prosite Patterns) and are well suited to predict new family members with a high specificity. This suite is for instance embedded in the cyanolase database [40] to automate its updade and was used for refining the classification of HAD enzymes [6] or identify shared conservations in the core proteome of extracellular vesicles produced by human and animal *S. aureus* strains [68].

PPSuite[6] is one of the first frameworks taking into account coevolutionary dependencies between residues for the comparison of protein sequences. It proposes a complete workflow enabling to infer direct couplings between the positions of a sequence of interest by a Potts model with the help of the sequence close homologs and to score the similarity of the sequences by alignment of the inferred Potts models, as well as tools to visualize the models and their alignments [67, 66].

AuReMe and AuCoMe workspaces is designed for tractable reconstruction of metabolic networks[7]. The toolbox allows for the Automatic Reconstruction of Metabolic networks based on the combination of multiple heterogeneous data and knowledge sources [1]. The main added values are the inclusion of graph-based tools relevant for the study of non-model organisms (Meneco and Menetools packages), the possibility to trace the reconstruction and curation procedures (Padmet package), and the exploration of reconstructed metabolic

[5] protomata-learner.genouest.org
[6] www-dyliss.irisa.fr/ppalign/
[7] aureme.genouest.org/

networks with wikis (wiki-export package, see: aureme.genouest.org/wiki.html) [32]. It also generates outputs to explore the resulting networks with Askomics. It has been used for reconstructing metabolic networks of micro and macro-algae [61], extremophile bacteria [43] and communities of organisms [4].

Mpwt, emmapper2gbk is a Python package for running Pathway Tools[8] on multiple genomes using multiprocessing. Pathway Tools is a comprehensive systems biology software system that is associated with the BioCyc database collection[9]. Pathway Tools is frequently used for reconstructing metabolic networks. In order to allow the output of the eggnoggmapper annotation tool to be used by Mpwt, we also developed emmaper2gbk to create relevant genome files.

Metage2metabo is a Python tool to perform graph-based metabolic analysis starting from annotated genomes (reference genomes or metagenome-assembled genomes) [30]. It uses Mpwt to reconstruct metabolic networks for a large number of genomes. The obtained metabolic networks are then analyzed individually and collectively in order to get the added value of metabolic cooperation in microbiota over individual metabolism and to identify and screen interesting organisms among all.

3.4 Regulation and signaling: detecting complex and discriminant signatures of phenotypes

On the contrary to metabolic networks, regulatory and signaling processes in biological systems involve agents interacting at different granularity levels (from genes, non-coding RNAs to protein complexes) and different time-scales. Our focus is on the reconstruction of large-scale networks involving multiple scales processes, from which controllers can be extracted with symbolic dynamical systems methods. Particular attention is paid to the characterization of products of genes (such as isoform) and of perturbations to identify discriminant signature of pathologies.

3.4.1 Research topics

Genomic level: characterizing gene structure with grammatical languages and conservation information The goal here is to accurately represent gene structure, including intron/exon structure, for predicting the products of genes, such as isoform transcripts, and comparing the expression potential of a eukaryotic gene according to its context (e.g. tissue) or according to the species. Our approach consists in designing grammatical and comparative-genomics based models for gene structures able to detect heterogeneous functional sites (splicing sites, regulatory binding sites...), functional regions (exons, promotors...) and global constraints (translation into proteins) [35]. Accurate gene models are defined by identifying general constraints shaping gene families and their structures conserved over evolution. Syntactic elements controlling gene expression (transcription factor binding sites controlling transcription; enhancers and silencers controlling splicing events...), i.e. short, degenerated and overlapping functional sequences, are modeled by relying on the high capability of SVG grammars to deal with structure and ambiguity [64].

System level: extracting causal signatures of complex phenotypes with systems biology frameworks Our main challenge is to set up a generic formalism to model inter-layer interactions in large-scale biological networks. To that goal, we have developed several types of abstractions: multi-experiments framework to learn and control signaling networks [11], multi-layer reactions in interaction graphs [36], and multi-layer information in large-scale Petri nets [29]. Our main issues are to scale these approaches to standardized large-scale repositories by relying on the interoperable Linked Open Data (LOD) resources and to enrich them with ad-hoc regulations extracted from sequence-based analysis. This will allow us to characterize changes in system attractors induced by mutations and how they may be included in pathology signatures.

3.4.2 Associated software tools

Logol software is designed for complex pattern modeling and matching[10]. It is a swiss-army-knife for pattern matching on DNA/RNA/Protein sequences, based on expressive patterns which consist in a complex combination of motifs (such as degenerated strings) and structures (such as imperfect stem-loop ou repeats) [2].

[8] bioinformatics.ai.sri.com/ptools/
[9] biocyc.org
[10] logol.genouest.org/

Logol key features are the possibilities (i) to divide a pattern description into several sub-patterns, (ii) to model long range dependencies, and (iii) to enable the use of ambiguous models or to permit the inclusion of negative conditions in a pattern definition. Therefore, Logol encompasses most of the features of specialized tools (Vmatch, Patmatch, Cutadapt, HMM) and enables interplays between several classes of patterns (motifs and structures), including stem-loop identification in CRISPR.

Caspo Cell ASP Optimizer (Caspo) software constitutes a pipeline for automated reasoning on logical signaling networks (learning, classifying, designing experimental perturbations, identifying controllers, take time-series into account)[11]. The software handles inherent experimental noise by enumerating all different logical networks which are compatible with a set of experimental observations [11]. The main advantage is that it enables a complete study of logical network without requiring any linear constraint programs.

Cadbiom package aims at building and analyzing the asynchronous dynamics of enriched logical networks[12]. It is based on Guarded transition semantic and allows synchronization events to be investigated in large-scale biological networks [29]. For example, it allowed to analyze controler of phenotypes in a large-scale knowledge database (PID) [5].

Recently, we have significantly refactored Cadbiom package towards a framework that allows the identification of causal regulators in large-scale models, formalized in the BioPAX language and automatically interpreted as guarded transitions. The Cadbiom framework was applied to the BioPAX version of two ressources (PID, KEGG) of the PathwayCommons database and to the Atlas of Cancer Signalling Network (ACSN). As a case-study, it was used to characterize the causal signatures of markers of the epithelial-mesenchymal transition.

4 Application domains

In terms of transfer and societal impact, we consider that our role is to develop fruitful collaborations with biology laboratories in order to consolidate their studies by a smart use of our tools and prototypes and to generate new biological hypotheses to be tested experimentally.

Marine Biology: seaweed enzymes and metabolism An important field of study is **marine biology**, as it is a transversal field covering challenges in integrative biology, dynamical systems and sequence analysis.

- **Protein functions in seaweed metabolism** Several years ago, our methods based on combinatorial optimization for the reconstruction of genome-scale metabolic networks and on classification of enzyme families based on local and partial alignments allowed the seaweed *E. siliculosus* metabolism to be deciphered [61, 44]. The study of the *HAD* superfamily of proteins thanks to partial local alignments produced by Protomata tools, allowed sub-families to be deciphered and classified. Additionally, the metabolic map reconstructed with Meneco enabled the reannotation of 56 genes within the *E. siliculosus* genome. These approaches also shed light on evolution of metabolic processes.

- **Elucidating algal metabolism thanks to large-scale metabolic network reconstructions** More recently, the tools developed by Dyliss (based on the AuReMe toolbox) allowed us to participate in the reconstruction of a metabolic network for the brown algae *Saccharina japonica* and *Cladosiphon okamuranus* in order to identify these species specificities on the synthesis of carotenoids biosynthesis [59]. We also participated in the study of the genome of *Ectocarpus subulatus*, a highly stress-tolerant algal strain [49]. Finally, AuReMe has been used to analyze the metabolic capacity of several strains of cyanobacteria, with results integrated in the Cyanorak database [52] and to characterize synergistic effects of the *synechococcus* strain WH7803 [55].

- **Metabolic pathway drift theory** Genome annotations can contribute to understanding algal metabolism. The tool PathModel was developed to add support for biochemical reactions and metabolite structures to the theory of metabolic pathway drift with an approach combining chemoinformatics knowledge reasoning and modeling. This approach was applied to the study of the red alga *Chondrus crispus*, which allowed to show that even for metabolic pathways supposed to be conserved between species (sterols, mycrosporins synthesis), we can see an important turnover in the order of reactions appearing

[11] bioasp.github.io/caspo/
[12] cadbiom.genouest.org

in a metabolic pathway. This work lays the foundations for the concept of "metabolic drift" analogous to the same concept in genomics. [31].

- **Algal-bacteria interactions** We reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* [48] and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. Similarily, our analyses suggested that the bacterium *Ca. P. ectocarpi* is able to provide both beta-alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway [62].

  These works paved the way to the study of host-microbial interactions, as shown in [41] where we evidenced the role of tools such as miscoto and metage2metabo to predict synthetic communities allowing to restore algal metabolic pathways. To validate these approaches experimentally, we worked with S. Dittami, researcher at the Roscoff biological station. We applied these methods on a set of about fifteen cultivable bacteria identified on the wall membrane of *Ectocarpus siliculosus*. Our approaches predicted that three bacteria were necessary to facilitate the growth of this alga in an axenic medium. The experiments were carried out, and indeed allowed the alga to grow in an axenic medium. This is therefore a proof of concept of the relevance of our approaches. More recently, the study of the freshwater strain of *Ectocarpus subulatus* evidenced the role of metabolism in adaptation, paving the way to biotechnological applications [57].

Microbiology: elucidating the functioning of extremophile consortiums of bacteria. Our main issue is the understanding of bacteria living in extreme environments. The context is mainly a collaboration with the group of bioinformatics at Universidad de Chile (co-funded by the Center of Mathematical Modeling, the Center of Regulation Genomics and Inria-Chile). In order to elucidate the main characteristics of these bacteria, our integrative methods were developed to identify the main groups of regulators for their specific response in their living environment. The integrative biology tools Meneco, Lombarde and Shogen have been designed in this context. In particular, genome-scale metabolic network been recently reconstructed and studied with the Meneco and Shogen approaches, especially on bacteria involved in biomining processes [37] and in Salmon pathogenicity [43]. We have also studied the specificities of two Microbacterium strains, CGR1 and CGR2, isolated in different soils of the Atacama Desert in Chile, showing significant differences on the connectivity of metabolite production in relation to pH tolerance and CO2 production [58].

Agriculture and environmental sciences: upstream controllers of cow, pork and pea-aphid metabolism and regulation. Our goal is to propose methods to identify regulators of complex phenotypes related to environmental issues. Our work on the identification of upstream regulators within large-scale knowledge databases (tool KeyRegulatorFinder) [36] and on semantic-based analysis of metabolic networks [34] was very valuable for interpreting the differences of gene expression in pork meat [56] and figure out the main gene-regulators of the response of porks to several diets [54]. Our expertise in microbiota analysis is also currently being applied to rumen microbial genomics [60].

Health: Dynamics of microenvironment in chronic liver diseases We develop methods and models to understand the dynamics of the microenvironment in order to propose evolutionary markers and effective therapeutic targets. The matrix microenvironment is the major regulator of events related to fibrosis-cirrhosis-cancer progression and Hepatic Stellate Cells (HSC) are the main actors of microenvironment remodeling. At molecular level, the transforming growth factor TGF-β plays a central role by promoting HSC activation, extracellular matrix remodeling and epithelial-mesenchymal transition. In that context we have developed three programs :

- *TGF-β signaling networks*. TGF-β is a multifunctional cytokine that binds to specific receptors and induce numerous signaling pathways depending on the context. Deciphering TGF-β signaling networks requires to take into account a system-wide view and develop predictive models for therapeutic benefit. For that purpose we developed Cadbiom and identified gene networks associated with innate immune response to viral infection that combine TGF-β and interleukin signaling pathways [29, 42]. More recently we have very significantly refactored Cadbiom package towards a framework that allows the identification of causal regulators in large-scale models, formalized in the BioPAX language and automatically interpreted as guarded transitions[13].The Cadbiom framework was applied to the BioPAX

[13] cadbiom.genouest.org

version of two resources (PID,KEGG) of the Pathway Commons database and to the Atlas of Cancer Signalling Network (ACSN). As a case-study, it was used to characterize the causal signatures of markers of the epithelial-mesenchymal transition.

- *Functional signature for ADAMTS*. Hepatic Stellate Cells produce a wide variety of molecules involved in ECM remodeling, such as adamalysins [69]. However, the limitations of discovering new functions of these proteins stem from the experimental approaches that are difficult to implement due to their structure and biochemical features. In that context we developed an original framework combining the identification of small modules in conserved regions independent of known domains and the concepts of phylogenomics (association of conservation and phenotype gained concurrently during evolution). The resulting evolutionary model of motif signatures and protein-protein interaction signatures of the ADAMTS family is validated by data from literature and provides biologists with many new potential functional motifs [45], [47], [46].

- *Dynamic model of hepatic stellate cells*. To characterize the dynamics of HSC activation upon TGFB1 stimulation, we developed a model using Kappa, a site graph rewriting language and its static analyzer Kasa [39]. We previously demonstrated the advantages of Kappa language for modeling TGF-β signaling and extracellular matrix [70]. Unlike previous model based on a population of interacting proteins, we now develop an original Kappa model based on a population of cells interacting with TGF-β [38]. The model recapitulates the dynamics of activation of HSC towards myofibroblast states and the reversion processes. Current work aims to identify the regulators of the repair likely to promote the resolution of fibrosis at the expense of its progression.

5 Social and environmental responsibility

5.1 Footprint of research activities

Our footprint for 2025 is 6.8T CO2 for the entire team. This is mainly driven by 3 transatlantic missions (2 long missions 2-3 weeks to see our collaborators in Chile, and 1 mission of 1 week in Chicago). For other missions in France and Switzerland (between 40 and 50), we favored train (0T CO2), except for one mission to Paris and another to Nantes by car (0.2T CO2, calculated on the ADEME website). For PhD juries ouside France (Canada), we opted for videoconferencing.

Outside missions, Dyliss research activities have low environmental footprints. Most of our software solution run on off-the-shelf computers and are not computationally intensive. Indirectly, the analyses and predictions we make intend to reduce the need for long, costly technically or ethically difficult biological experiments.

5.2 Impact of research results

Through our ongoing collaborations with INSERM and Rennes' Hospital, Dyliss research activities have a social impact on human health. Our collaborations with INRAe have a direct impact on vegetal and animal health, and an indirect impact in environment as these projects original motivation is to reduce fertilizers or pesticides.

6 Highlights of the year

In 2025, in accordance with the operating principle of INRIA project teams, which have a limited lifespan, the DYLISS team actively prepared for its future evolution. A request to create a future team was submitted and accepted (future BioGraphs team).

6.1 Members

- Yann Le Cunff successfully defended his habilitation, entitled "From Data to Phenotype: Integrating Data Structure and Prior Knowledge to Model Biological Systems" in May 2025.

- Samuel Blanquart was promoted to CRHC.
- Jeanne Got was promoted to IEHC.

6.2 Broadening International Visibility

Yaël Tirlet, PhD student in the DYLISS team, has been awarded by the doctoral school a fellowship for a mobility in Switzerland (3 months), to start a collaboration with the Swiss Institute of Bioinformatics. The mobility allowed to start a fruitful collaboration with Jerven Bolleman involving other members of the DYLISS team Olivier Dameron, Emmanuelle Becker (publications submitted).

Cécile Beust, PhD student in the DYLISS team, has been awarded by the doctoral school a fellowship for a mobility in the U.S. (UC San Diego, California) to start a collaboration with the Cytoscape team (2 months), but was unable to move to the U.S. at the proposed period because of the U.S. administration's visa blocking policy in June 2025. The visit was thus replaced by a remote collaboration involving other members of the DYLISS team Emmanuelle Becker, Olivier Dameron and Nathalie Theret.

7 Latest software developments, platforms, open data

7.1 Latest software developments

7.1.1 AskOmics

Name: Convert tabulated data into RDF and create SPARQL queries intuitively and "on the fly".

Keywords: RDF, SPARQL, Querying, Graph, LOD - Linked open data

Functional Description: AskOmics aims at bridging the gap between end user data and the Linked (Open) Data cloud. It allows heterogeneous bioinformatics data (formatted as tabular files) to be loaded in a RDF triplestore and then be transparently and interactively queried. AskOmics is made of three software blocks: (1) a web interface for data import, allowing the creation of a local triplestore from user's datasheets and standard data, (2) an interactive web interface allowing "à la carte" query-building, (3) a server performing interactions with local and distant triplestores (queries execution, management of users parameters).

URL: https://askomics.org/

Contact: Olivier Dameron

Partners: Université de Rennes 1, CNRS, INRA

7.1.2 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 keystone 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

7.1.3 AuCoMe

Name: Automatic Comparison of Metabolisms

Keywords: Bioinformatics, Workflow, Metabolic networks, Omic data, Data analysis

Functional Description: AuCoMe is a Python package that aims at reconstructing homogeneous metabolic networks and pan-metabolism starting from genomes with heterogeneous levels of annotations. Four steps are composing AuCoMe. 1) It automatically infers annotated genomes from draft metabolic networks thanks to Pathway Tools and MPWT. 2) The Gene-Protein-Reaction (GPR) associations previously obtained are propagated to protein orthogroups in using Orthofinder and, an additional robustness criteria. 3) AuCoMe checking the presence of supplementary GPR associations by finding missing annotation in all genomes. In this step, the tools BlastP, TblastN and, Exonerate are called. 4) It adding spontaneous reactions to metabolic pathways that were completed by the previous steps. AuCoMe generates several outputs to facilitate the analysis of results: tabuled files, SBML files, PADMET files, supervenn and a dendogram of reactions.

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

Publication: hal-03778267

Contact: Anne Siegel

Participants: Arnaud Belcour, Jeanne Got, Meziane Aite, Ludovic Delage, Jonas Collen, Clemence Frioux, Catherine Leblanc, Simon M. Dittami, Samuel Blanquart, Gabriel V. Markov, Anne Siegel

7.1.4 prolipipe

Keywords: Metabolic networks, Workflow, Bacterial strains

Scientific Description: This pipeline evaluates in silico the ability of several thousand bacteria to produce specific metabolites. (1) Reconstruction of large-scale metabolic networks using three annotation software programs, thanks to the AuFAMe tool (https://github.com/AuReMe/AuFAMe). (2) Analysis of the synthesis pathways producing specific metabolites in the bacterial metabolic networks created. (3) Generation of a heatmap for each synthesis pathway studied. (4) Production of a SPARQL-queryable file to easily exploit the results produced.

Functional Description: This pipeline evaluates in silico the ability of several thousand bacteria to produce specific compounds. Prolipipe generates bacterial metabolic networks from their genomes. By focusing on certain synthesis pathways chosen by the user, it will create as many heatmaps as there are synthesis pathways studied. The metabolic specifications of each bacterium will be visualized on these heatmaps. Prolipipe also generates an easily queryable file.

News of the Year: With this software, we obtained a PCI recommendation in 2025 and are currently in the process of publishing it.

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

Publication: hal-05045657

Contact: Noe Robert

Participants: Noe Robert, Jeanne Got, Pauline Giraud, Hélène Falentin, Anne Siegel

Partner: INRAE

7.1.5 EnzBert-GO

Keywords: Proteins, Biological sequences, Functional annotation, Deep learning, Ontologies

Scientific Description: Code for learning and using BERT deep neural architectures for the prediction of multi-level and multi-class functional enzymatic GO (Gene Ontology) annotations of protein sequences.

Functional Description: Prediction of the functional enzymatic GO annotations of protein sequences

News of the Year: EnzBert-GO has been generalized to support hierarchical labels beyond those from Gene Ontology. This includes expert refinements of Gene Ontology, Enzyme Commission numbers, and others.

URL: https://gitlab.inria.fr/fcoste/enzbert-go

Contact: François Coste

Participant: François Coste

7.1.6 FUSE-PhyloTree

Name: FUnctions and SEquence conservations on a Phylogenetic Tree

Keywords: Bioinformatics, Biological sequences, Sequence alignment, Phylogenomics, Proteins

Scientific Description: FUSE-PhyloTree is dedicated to estimate the sequence regions which are potentially associated to functions of interest in a multi-functional protein families, such as paralogous and multi-domain protein families. The method uses state-of -the-art programs to estimate a mapping of both the ancestral functions and the ancestral sequence content at each node in the phylogenetic family tree. It enables the association of functions with local sequence conservations through the inference of their co-appearance along the evolutionary gene tree, and it generates interactive Itol representations allowing to explore the annotated tree.

Functional Description: FUSE-PhyloTree takes as input: 1) protein sequences of the target family (including both paralogs and orthologs), 2) a gene tree corresponding to these sequences, and 3) functional annotations of interest of proteins, for instance their identified protein-protein interactions (PPI). As a result, FUSE-PhyloTree provides a gene tree annotated with both predicted conserved sequence modules and functions of ancestral genes, enabling the association of functions with specific sequence regions based on their co-emergence during gene evolution.

News of the Year: The primary improvement in the software is the introduction of an estimate for the robustness of predictions regarding the appearance of modules in ancestral genes. This allows users to prioritize stronger predictions while still retaining weaker ones for consideration. In addition, the update includes general improvements to the user experience, enhanced documentation, and has been published as an Application Note in Bioinformatics.

URL: https://github.com/OcMalde/fuse-phylotree

Publications: hal-05238604, hal-04248728

Contact: François Coste

Participants: Olivier Dennler, Elisa Chenel, François Coste, Samuel Blanquart, Catherine Belleannée, Nathalie Theret

8 New results

8.1 Scalable methods to query data heterogeneity

Participants: Emmanuelle Becker, Cécile Beust, Océane Carpentier, Olivier Dameron, Ulysse Le Clanche, Yann Le Cunff, Alice Mataigne, Anne Siegel, Nathalie Théret, Yael Tirlet.

ACUITEE: A Comprehensive Tool for Visualization, Editing and Curating textual Annotations in Clinical Data [Olivier Dameron] [18] Annotation and management of clinical data remains a critical but challenging task due to the complexity and diversity of medical records. We developed ACUITEE (Annotation and Curation User Interface for Terms Extraction Engines), a web application that offers a simple way to improve clinical data annotation workflows by integrating automatic analysis, manual processing, and real-time visualization of medical notes. Using advanced natural language processing (NLP) techniques for phenotypes extraction such as PhenoBERT and efficient string-matching algorithms, ACUITEE maps free-text medical notes to ontology terms and enables clinicians to validate or refine these annotations through a user-friendly interface. The system supports fully automated, semi-automated and manual annotation modes, providing flexibility for different use cases. A key feature of ACUITEE is its interactive annotation interface, which enables clinicians to validate, edit, and curate ontology terms with precision, thereby speeding up the annotation process while maintaining high accuracy.

Biological Knowledge Extraction from BioPAX Graphs [Emmanuelle Becker, Cécile Beust, Olivier Dameron, Nathalie Théret] [25] In systems biology, the study of biological pathways is a key to understand the complexity of biological systems. The recent massification of biological pathway data available online through various databases raised an important need of standardization of these data. The BioPAX format (Biological Pathway Exchange), created in 2010, is a semantic web format designed for the standardization and exchange of pathway data. BioPAX is highly expressive but intrinsically complex, limiting its wider adoption. We reported on the use of the BioPAX format in 2024 and present abstraction methods to simplify knowledge extraction from BioPAX graphs.

Assessing bioinformatics software annotations: bio.tools case-study [Olivier Dameron, Ulysse Le Clanche, Yann Le Cunff] [20] Reproducibility and reuse of digital bioinformatics resources are essential for the development of open and cumulative science, in line with FAIR principles. To search and reuse bioinformatics tools, scientists need to be confident enough with the reliability of their annotations. Our study focuses on the quantitative and qualitative evaluation of semantic annotations in the bio.tools registry, which serves more than 30,000 bioinformatics tool descriptions, annotated with the EDAM ontology. In this work we propose to study how the EDAM ontology is used to categorize software based on scientific disciplines and the kind of data processing they allow. We also evaluate how qualitative are the annotations based on Shannon entropy. We emphasize that a particular attention should be given to the whole set of inherited annotations, from the used ontology. Our results underline the need for automatic tools to support annotation curation, reducing the annotation cost for domain experts. This study is a preliminary work aimed at designing novel annotation approaches based on the combination of knowledge graphs and large language models towards more findable and reusable bioinformatics tools.

MLOps best practices for bioinformatics [Yann Le Cunff] [27] Machine learning is increasingly used in bioinformatics for various applications. Developing and maintaining machine learning models requires methods to ensure reproducibility and facilitate the deployment. Unfortunately, these methods are until now rarely used in bioinformatics, and there is a critical need for the adoption of good practices in this field, just as was done in the last years for FAIR management of data, tools or workflows. Machine Learning Operations (MLOps) is a set of practices and tools that offer a very good framework for optimizing machine learning lifecycle management.

8.2 Metabolism: from protein sequences to systems ecology

Participants: Moana Aulagner, Emmanuelle Becker, Catherine Belleannée, Samuel Blanquart, Myriam Bontonou, Elisa Chenel, François Coste, Pablo Espana Gutierrez, Pauline Giraud, Jeanne Got, Yann Le Cunff, Alice Mataigne, Noé Robert, Anne Siegel, Nathalie Théret.

Modeling the emergent metabolic potential of soil microbiomes in Atacama landscapes [Pauline Giraud, Yann Le Cunff, Anne Siegel] [12] The Atacama Desert's extreme Talabre Lejía transect serves as a natural lab to study how microbial communities adapt through metabolic interactions. A new computational framework—combining taxonomic/functional profiling, metabolic modeling, and regression—identifies key species and metabolites across six soil samples. Results reveal functional redundancy in metagenomes and site-specific adaptations, linking environmental stressors to microbial survival strategies. The approach is scalable for any (meta)genomic dataset with robust environmental data, offering insights into metabolism-driven resilience in extreme ecosystems.

Evolutionary history and association with seaweeds shape the genomes and metabolisms of marine bacteria [Pauline Giraud, Anne Siegel] [15] Seaweeds support a rich diversity of bacteria, offering metabolic resources and surfaces for biofilm development. To determine whether seaweed-associated bacteria possess unique genetic and metabolic traits compared to their free-living counterparts in seawater, we analyzed genomes from 72 bacterial genera across 16 different seaweed hosts. The study revealed that taxonomic classification plays a major role in shaping genomic features like GC content, gene number, and genome size. Their genomes reveal metabolic adaptations: enriched pathways for B vitamin synthesis, complex carbohydrate breakdown, and amino acid production—especially in Flavobacteriia. No evidence of host-metabolism complementarity was found in *Ectocarpus subulatus* and its bacteria. These adaptations may impact coastal carbon, nitrogen, and sulfur cycling.

A duo of fungi and complex and dynamic bacterial community networks contribute to shape the Ascophyllum nodosum holobiont [Samuel Blanquart] [16] The brown alga *Ascophyllum nodosum* and its microbiota form a dynamic functional entity named holobiont. Some microbial partners may play a role in seaweed health through bioactive compounds crucial for normal morphology, development, and physiological acclimation. However, the full spectrum of the microbial diversity and its variations according to algal life stage, season, and location have not been comprehensively studied. This study uses 208 short-read metabarcoding samples to characterize the bacterial, archaeal, and microeukaryotic communities of *A. nodosum* across three nearby sites, four thallus parts, and a monthly survey, aiming to explore the dynamics of ecological interactions within the holobiont. Our results revealed that *A. nodosum* harbors a predominantly bacterial microbiota, varying significantly across all covariables, while archaea were virtually absent. An innovative normalization using the co-amplified host reads provided an estimation of bacterial abundance, revealing a drastic decline in May, potentially linked to epidermal shedding. In contrast, fungal communities were stable, dominated by *Mycophycias ascophylli* and *Moheitospora sp.*, which remained closely associated with the host year-round. We identified a core microbiome of 22 ASVs, consistently found in all samples, including Granulosicoccus, a genus consistently abundant in other brown algal microbiota. Sequence clustering revealed multiple species which vary according to seasons, even in the overall stable Granulosicoccus genus. Co-occurrence network analysis revealed putative interactions between microbial groups in response to ecological niches. Overall, these findings highlight the dynamic of bacterial interactions and stable fungal associations within the *A. nodosum* holobiont, providing new insights into the ecology of its microbiota.

Methods for a species-specific genome-scale metabolic model designed for eukaryotes and applied to the Ascophyllum nodosum macroalga [Pauline Giraud, Jeanne Got, Anne Siegel] [19] The Prolipipe pipeline addresses the challenge of assessing functional variability in food industry-relevant bacteria by enabling large-scale metabolic potential evaluation from genomic data. Leveraging public genome repositories, it automates the construction of metabolic networks, with enzyme identification as a key focus. For hundreds to thousands of bacterial genomes, Prolipipe integrates triple-tool annotation to predict gene functions, builds genome-scale metabolic networks, and maps the presence/absence of pathway-specific reactions. Applied to 1,494 lactic acid bacteria genomes, it evaluated 761 pathways, revealing 137 pathways operational in at least one strain, while four Metacyc functional classes remained unrepresented. The pipeline also uncovered infraspecific variability, highlighting strain-dependent phenotypic differences within species, which underscores the functional diversity critical for industrial applications.

Studying metabolic cross-feedings insides phycospheres during cyanobacterial harmful blooms (HCBs) [Jeanne Got] [22] Using metagenomic, metabolomic and metabolic modelling, we characterised 12 Microcystis cyanobacteria and 97 MAGs from the phycosphere cultured after isolation from a pond near Paris. Metabolic modelling, identification of biosynthetic gene clusters, and secondary metabolites highlighted differences between the metabolic capacities of the phycosphere and the importance of manual curation of secondary metabolism in metabolic networks. These results deepen our understanding of Microcystis' phycosphere functioning, demonstrate the relevance of multi-omics systems biology approaches, and lay the fundation for further characterisation of freshwater HCB's microbial interactions and inter-species complementarity.

Carbon substrates utilization determine antagonistic fungal-fungal interactions among root-associated fungi [Alice Mataigne] [14] This study explores how fungal metabolism shapes fungal-fungal interactions in the plant microbiome, an area far less understood than bacterial competition. By profiling carbon substrate utilization in 91 root-associated fungal isolates, the authors reveal that fungal carbon usage strategies vary widely—independent of host plant species, root compartment, or geography. Notably, fungi with antifungal-mediated antagonism exhibit broader, faster carbon utilization, while those relying on direct competition use fewer substrates at slower rates. Combined with taxonomy-based enzyme predictions, these findings suggest that carbon utilization profiles and enzymatic reactions could serve as markers of fungal antagonistic potential. Ecologically, this highlights how metabolic diversity among root fungi drives their competitive dynamics, offering new insights into microbiome assembly and fungal interaction networks.

Metagenomic taxonomic assignment using Nanopore reads, reconstruction of metabolic networks and prediction of metabolite production [Jeanne Got, Anne Siegel] [21] The intestinal microbiota shapes the early-life gut barrier through metabolite production. Using INRAE's Holopig program, colonic samples from control and colistin-treated piglets were analyzed. The AuCoMe and MeneTools pipelines reconstructed bacterial metabolic networks, revealing strain-level metabolic diversity—shared and unique pathways—within species. Gram-negative bacteria emerged as key producers of metabolites critical for intestinal immunity, permeability, inflammation, and gut-brain signaling. Future work aims to scale this approach for broader microbiota metabolite predictions.

FUSE-PhyloTree: Linking functions and sequence conservation modules of a protein family through phylogenomic analysis [Catherine Belleannée, Samuel Blanquart, Elisa Chenel, François Coste, Nathalie Theret] [13] FUSE-PhyloTree is a phylogenomic analysis software for identifying local sequence conservation associated with the different functions of a multi-functional (e.g., paralogous or multi-domain) protein family. FUSE-PhyloTree introduces an original approach that combines advanced sequence analysis with phylogenetic methods. First, local sequence conservation modules within the family are identified using partial local multiple sequence alignment. Next, the evolution of the detected modules and known protein functions is inferred within the family's phylogenetic tree using three-level phylogenetic reconciliation and ancestral state reconstruction. As a result, FUSE-PhyloTree provides a gene tree annotated with both predicted sequence modules and ancestral gene functions, enabling the association of functions with specific sequence regions based on their co-emergence. FUSE-PhyloTree is provided as Docker and Singularity images including all the required software tools.

8.3 Regulation and signaling: detecting complex and discriminant signatures of phenotypes

Participants: Emmanuelle Becker, Catherine Belleannée, Samuel Blanquart, Myriam Bontonou, Olivier Dameron, Juliette Francis, Yann Le Cunff, Corentin Lucas, Noryah Safla, Anne Siegel, Nathalie Théret.

Pervasive formation of double-stranded RNAs by overlapping sense/antisense transcripts in budding yeast mitosis and meiosis [Emmanuelle Becker] [17] Previous RNA profiling studies revealed co-expression of overlapping sense/antisense (s/a) transcripts in pro- and eukaryotic organisms. Functional analyses in yeast have shown that certain s/a mRNA/mRNA and mRNA/lncRNA pairs form stable double-stranded RNAs (dsRNAs) that affect transcript stability. Little is known, however, about the genome-wide prevalence of dsRNA formation and its potential functional implications during growth and development in diploid budding yeast. To address this question, we monitored dsRNAs in a Saccharomyces cerevisiae strain expressing the

ribonuclease DCR1 and the RNA binding protein AGO1 from Naumovozyma castellii. We identify dsRNAs at 347 s/a loci that express partially or completely overlapping transcripts during mitosis, meiosis or both stages of the diploid life cycle. The data are interesting from an evolutionary perspective, since natural antisense transcripts that form stable dsRNAs have been detected in many species from bacteria to humans. This work was driven by Michaël Primig, collaborator at IRSET (Rennes).

Identifying coevolving residues by factoring out the evolutionary distance covariance matrix [François Coste, Pablo Espana Gutierrez] [26] The identification of coevolving residues in protein families underlies recent breakthroughs in predicting protein structure from sequence, from early Direct Coupling Analysis (DCA) methods to modern tools like AlphaFold. However, as shown by Qin and Colwell (2018), residue covariations in multiple sequence alignments are heavily influenced by phylogenetic relationships. A key challenge remains: distinguishing covariations due to shared evolutionary history from those driven by structural or functional constraints. Some deep learning architectures, such as MSA Transformers, have been introduced to handle these two sources of signal separately. Here, we investigate a more direct approach: explicitly analyzing this separation within the classical DCA covariance framework. To handle this two-source signal, we introduce a novel method based on a matrix normal law that explicitly separates sequence-level and residue-level dependencies via two covariance matrices instead of one: one for coevolution among residue positions and the other for evolutionary distances between sequences. From this theoretical framework, we derive an estimator of the residue-residue covariance matrix by factoring out the contribution of evolutionary relationships, encoded in the sequence distance covariance matrix, from the observed dependencies. We perform a spectral analysis of this estimator, revealing the need for more refined strategies to estimate the covariance of sequence evolutionary distances. We then present two alternative approaches that better incorporate the phylogenetic history of sequences for improved practical estimation of evolutionary distances and more accurate identification of coevolving residues through their removal in covariance-based Direct Coupling Analysis.

9 Bilateral contracts and grants with industry

9.1 Bilateral Grants with Industry

BeCycle

Participants: Jeanne Got, Noé Robert, Anne Siegel.

In the context of the Grand Défi "Ferment du futur", this private-public project aims at scanning thousands of bacterial genomes to identify the best consortium of strains capable of producing metabolites of interest. Duration: 2024-2026, total of the grant 400k€.

10 Partnerships and cooperations

10.1 International initiatives

10.1.1 Visits of international scientists

Other international visits to the team

Cathy Pfiser

Status (researcher, PhD, post-Doc, intern (master/eng))

Institution of origin: Univ. Chicago

Country: USA

Dates: December 2025

Context of the visit: Meeting algometabionte, Rennes, December 2025, 30 participants [Anne Siegel]

Mobility program/type of mobility: research stay

Domenico Palladino

Status PhD student

Institution of origin: Univ. Salerno

Country: Italy

Dates: November 2025 - January 2026

Context of the visit: Erasmus+ [Olivier Dameron]

Mobility program/type of mobility: internship

10.1.2 Visits to international teams

Research stays abroad

Anne Siegel

Visited institution: University of Chile

Country: Chile

Dates: January 2025

Context of the visit: Cloture of the associated team biointegrative-chile. Organisation of the workshop metabolic'lub.

Mobility program/type of mobility: associated team Inria.

Anne Siegel

Visited institution: University of Chicago

Country: United States

Dates: March 2025

Context of the visit: Visit of the department of environment

Mobility program/type of mobility: Invitation from the AI Schmidt program.

Anne Siegel

Visited institution: University of Chile

Country: Chile

Dates: December 2025

Context of the visit: Collaboration with CMM and CRG.

Mobility program/type of mobility: Local invitation

Yael Tirlet

Visited institution: Swiss Institute for Bioinformatics

Country: Switzerland

Dates: May 2025 – July 2025

Context of the visit: Starting collaboration

Mobility program/type of mobility: Doctoral school fellowship

10.2 European initiatives

10.2.1 Other european programs/initiatives

ERC HoloE2Plant, Exploring the Holobiont concept through a Plant Evolutionary Experiment study

Participants: Moana Aulagner, Samuel Blanquart, Anne Siegel.

Exploring the Holobiont concept through a Plant Experimental Evolution study. In her ERC project, Claudia Bartoli aims at validating the holobiont concept, highlighting how the interactions with its microbiota influence a species evolution. The study will apply to a host/pathogen system, Brassica rapa / Rhizoctonia solani, associated with bacterial and fungal synthetic communities. Examining nine plant generations in an experimental-evolution apparatus should reveal the molecular outcomes of the applied selective pressures. 2022-2027, total of the grant 1500k€.

10.3 National initiatives

SEABIOZ : Potential microbial origins of the biostimulant properties of extracts from a brown algae holobinte

Participants: Samuel Blanquart, Olivier Dameron, Jeanne Got, Anne Siegel.

For sustainable agriculture, new bio-based solutions include biocontrol and the use of plant biostimulants such as aqueous seaweed extracts. The most widely exploited biomass for biostimulant production is the brown seaweed *Ascophyllum nodosum* and its commercial extracts, including products from the Roullier Group, have demonstrated their ability to improve plant growth and mitigate certain abiotic and biotic stresses. A unique feature of the alga is its mutualistic association with the fungal endophyte *Mycophycias ascophylli* and other microbes constituting an holobiont. Many questions remain as to the nature and origin of the active compounds in algal extracts. Are these bioactive metabolites produced by the host or by its microbiota? The main objective of SEABIOZ is to answer these questions by combining a multi-omics approach and systems biology. 2021–2025. Dyliss grant: 120k€.

DeepImpact : Deciphering plant-microbiome interactions to enhance crop defense to bioagressors

Participants: Samuel Blanquart, Olivier Dameron, Jeanne Got, Alice Mataigne, Pauline Giraud, Anne Siegel.

DEEP IMPACT is a multidisciplinary consortium-based project that aims at combining ecology, biology, plant genetics and mathematics to identify, characterize and validate the microbial communities, plant communities and abiotic factors (including agricultural managements) explaining variation in *Brassica napus* and *Triticum aestivum* resistance to several pests. For this, we will start from an *in situ* approach by

characterizing 100 fields (50 for each crop species) for both habitat (climatic and edaphic variables) and biotic (microbiota, virome, weed communities, pest attacks and pathobiota prevalence) features. Information from this broad characterization will be integrated into sparse and correlative statistical models to describe the relative part of the variance explained by both habitat and biotic features and correlated with a reduction of pest's attacks. This analysis will allow us to identify a combination of microbial species and soils, correlated with an increase of crop's resistance to pests. These microbial consortia will be isolated by taking advantages of newly developed culturomics methods and characterized by both whole genome sequencing and biochemical assays. Synthetic Consortia (SynComs) will be reconstructed to test their efficacy on a broad range of pests attacking both crops. 2021–2026. Dyliss grant: 176k€.

ENDOVIRE (ANR)

Participants: Emmanuelle Becker, Olivier Dameron, Yael Tirlet.

The whole ANR project gathers 4 partners : the BIPAA platform (INRAe), the DGIMI laboratory, the BF2I laboratory and the Dyliss team of IRISA. The project is focused about the understanding of how genes of a endogeneized viral genome in a parasitoid wasp are the activated and regulated. The available data produced by the consortium will cover genomics, epigenomics, pathways, regulation and orthology. We will contribute to identify the key actors involved in the activation of parasitoids genes, to propose a data and knowledge integration framework for the data of the global project, and to develop integrative data analysis methods for elucidating the mechanism involving the key actors identified in the first point. It will consist in proposing a library of queries (which contains a reasoning part), and further to propose regulation mechanisms based on heterogeneous -omics data across interacting organisms. To tackle the different challenges, our appoach will be based on (1) adequate statistical analysis workflows or methods, (2) Semantic Web technologies and AskOmics developed within the team, (3) knowledge-guided traversal strategies across multiplex graphs. 2023–2026. Dyliss grant: 176k€.

PEPR Digital health : ShareFAIR

Participants: Olivier Dameron, Ulysse Le Clanche, Yann Le Cunff.

The increasing availability of life science data offers unprecedented opportunities for healthcare research, it has the potential to revolutionize the way we understand and treat diseases, as it allows researchers to identify trends and patterns that may not have been apparent with smaller data sets. However, exploiting this potential requires innovative solutions for the annotation of biomedical and clinical datasets and extraction of provenance. Challenges thus include standardization and annotation for datasets and protocols, extracting protocols from text and datasets, and synthesizing them into interoperable, yet shareable protocols. ShareFAIR will provide (i) standards to uniformly annotate datasets and protocols with ontologies/common vocabularies and provenance to trace their origin, (ii) an interoperable framework to index, design and annotate reliable and shareable analysis protocols, (iii) approaches to extract new protocols, based on the literature, learned from biomedical and clinical datasets, and from international data challenges in neuroimaging. Dyliss contribution consists in designing a semi-automated dataset FAIRification method that will extend low-level metadata by higher level descriptions inferred from the workflow specification and execution. These descriptions will provide a summary focusing on the "what" rather than the "how", that will be instrumental to workflow recommendation as well as improved reusability of data analysis results. To this end, we will leverage domain-specific knowledge associated to biomedical datasets, as well as fine-grained workflow execution provenance traces so that data analysis results can be more easily understood, explained and shared, in line with critical open and reproducible sciences initiatives. The PhD of Ulysse Le Clanche is co-supervized by Olivier Dameron at Dyliss and Alban Gaignard at Institut du Thorax, INSERM and Univ. Nantes. 2023–2027. Dyliss grant: 185k€.

PEPR Digital agro-ecology : HOLOBIONT

Participants: Juliette Francis, Yann Le Cunff.

Animals and their microbiota form a composite organism, called a holobiont, which can be considered the ultimate unit on which evolution and selection act. Host genes and the environment influence the colonization, development, and function of the various microbiota, which in turn help shape the host's phenotypes. The phenotypes of the holobiont thus result from the combined action of the host genes and those of its microbiota, and their determinism can be explored by implementing hologenetic approaches capable of considering host genomes and metagenomes jointly. The overall objective of this PEPR is to develop integrative hologenetic approaches for animal breeding, using state-of-the-art technologies to generate, process and analyze genetic and genomic datasets of the host and its microbiota as well as the phenotypes and environmental parameters in which the holobionts evolve. To this end, the project aims to develop methods for the analysis of new-generation phenotyping data of the holobiont (mainly high-throughput and continuous), for their modeling and for the analysis of their interrelationships with the microbiota data. Juliette Francis's Ph.D, co-supervised by Yann Le Cunff (Dyliss) and Mahendra Mariadassou (INRAe, MaIAGe), focuses on co-analyzing genomic data, microbiota data and metabolomic data to efficiently predict a phenotype of interest (food intake efficiency in this case). 2024–2028. Dyliss grant : 178k€.

PEPR Digital health : M4DI

Participants: Emmanuelle Becker, Océane Carpentier, Yann Le Cunff.

The main objective of the Methods and Models for Multimodal and Multiscale Data Integration (M4DI) project is to develop innovative methodological frameworks for the integration of biomedical datasets. In particular, the team is involved in designing robust machine learning approaches enhanced by prior knowledge. In particular, Océane Carpentier's Ph.D, co-supervised by Emmanuelle Becker, Yann Le Cunff (Dyliss), Nicolas Jay and Aurélie Bannay (LORIA, Nancy) is dedicated to exploiting the ontology structure of the Gene Ontology database in machine learning algorithms. One key application will be carried out on a local cohort of Crohn's patients with the CHU of Rennes. 2024–2028. DYLISS grant: 169k€.

CRLnet (ANR)

Participants: Emmanuelle Becker, Olivier Dameron.

The whole ANR aims at better understanding the ubiquitin system, a vital regulatory network that controls many different proteins in our cells. It works by attaching a small protein called ubiquitin to other proteins, which either adjusts their activity or marks them for degradation. This system plays a key role in various diseases, including cancer, neurodegenerative disorders, and infections. Cullin RING ligases (CRLs) are crucial components of the ubiquitin system, found in organisms ranging from yeast to humans. They are made up of several interchangeable subunits. Despite two decades of research, the different ways they operate and the cellular proteins they target is still poorly understood. The whole ANR will use budding yeast as a model organism and employ a cutting-edge technique called NanoBiT for identifying CRL interaction partners and create a comprehensive catalog of CRL interaction partners. Our contribution to the ANR aims at priorizing the potential targets identified with the NanoBit experiments, by leveraging knowledgebases about biological interactions (protein-protein interaction, metabolic networks, genetic interactions...). 2025–2029. Dyliss grant: 144k€.

10.3.1 Programs funded by Inria

Exploratory Action ECxit: Exiting the EC Classification for Better Enzyme Annotation by Deep Learning

Participants: François Coste.

Scientific leader: François Coste

Duration: 2025–2029

Description: Deep language models, such as those behind ChatGPT, have revolutionized natural language processing. By treating protein sequences as a language, the ECxit project aims to transfer these advances to the field of biology. Its goal is to develop a novel method and a redesigned classification of enzymes, enabling their identification and the precise prediction of their functions from amino acid sequences, ultimately improving genome annotation.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- Meeting algometabionte, Rennes, December 2025, 30 participants [Anne Siegel]

11.1.2 Scientific events: selection

Member of the conference program committees

- ISMB-ECCB 2025 (International Symposium on Molecular Biology) [Anne Siegel]
- Jobim 2025 (Journées Ouvertes Biologie Informatique Mathématiques), France [Olivier Dameron]
- Journée Santé et IA 2025 (plateforme IA, Dijon) [Olivier Dameron]
- Journées de Biologie In Silico Rennes, France [François Coste]

Reviewer

- ISMB/ECCB 2025 [Yann Le Cunff, Emmanuelle Becker]
- Jobim 2025 [Olivier Dameron, Yann Le Cunff]
- Santé et IA 2025 [Olivier Dameron]

11.1.3 Journal

Reviewer - reviewing activities

- Matrix Biology [Nathalie Théret]

11.1.4 Invited talks

- AI Schmidt seminar series, Chicago university, *Discovering Functions in Taxonomic Characterization of Environmental Samples: Combining Symbolic Data Science and Machine Learning*, Chicago, US, March 2025 [Anne Siegel]
- Metaboliclub, *What can we say from taxonomic affiliation : from metabolic functions to classification*, Chile, January 2025 [Anne Siegel]
- Dynamics in Patagonia, *Symbiosis in Environmental Biology, Discretization of Dynamical Systems*, Puerto Natales, Chile, Decembre 2025 [Anne Siegel]
- "Data and knowledge integration, analysis, life cycle and reproducibility in life sciences" INRAe PEPI IBIS, Rennes October 2025 [Olivier Dameron]
- Algometabiont days, Rennes December 2025. *Beyond EC annotation of enzymes* [François Coste]

11.1.5 Leadership within the scientific community

National responsibilities

- Deputy Scientific Directory (CNRS Informatics), in charge of interdisciplinarity between numerical sciences and other disciplines, gender equality in computer sciences, groupements de recherches (GDR), until 01/10/2025 [Anne Siegel]
- Scientific officer (CNRS Informatics, gender equality in computer sciences, since 01/10/2025 [Anne Siegel]
- Mediator and Member of the steering committee of the programme LORIER: The Organization for Ethical and Responsible Research at Inserm [Nathalie Théret]

Local responsibilities

- Scientific Director of the GenOuest platform [Yann Le Cunff]
- Responsability of the cross-cutting axis "Health-biology" at IRISA [Yann Le Cunff]
- Full member of the Sciences Faculty Counsil at University of Rennes [Emmanuelle Becker]
- Head of the Master degree "Bioinformatics" [Emmanuelle Becker]
- Head of the double-diploma Licence degree "Life Science, Maths and Artificial Intelligence" [Yann Le Cunff]
- Responsability of the 2nd and 3rd years of the "Yes-if" ISTN licence program, Univ. Rennes, France [Catherine Belleannée]
- In charge of the "Open Day and student fair" for Istic, Univ. Rennes, France [Catherine Belleannée]
- Referent teacher, 15h, L1 informatique, Univ. Rennes, France [Catherine Belleannée]
- Member of the Parcoursup jury for the Life Science Licence, Univ. Rennes, France [Yann Le Cunff, Emmanuelle Becker]

11.1.6 Scientific expertise

Evaluation of national projects

- Full member of the Evaluation Committee for ANR section "Interfaces: mathematics, numerical sciences for health and biology" [Emmanuelle Becker]
- ANRT CIFRE [Olivier Dameron]

11.1.7 Research administration

Institutional boards for the recruitment and evaluation of researchers

- Professor selection committee University of Rennes (internal promotion, CNU 64) [Emmanuelle Becker]
- Junior professor selection committee University of Poitiers (CNU 27) [Emmanuelle Becker]
- Junior professor in Biology and computer sciences selection committee, CNRS [Anne Siegel]
- Associate professor selection committee University of Marseille (CNU 26/27)[Emmanuelle Becker]
- Associate professor selection committee, Paris City University [Yann Le Cunff]
- Non-permanent Associate professor selection committee University of Rennes [Emmanuelle Becker]
- Research Engineer selection committee INRAe IGEPP [Emmanuelle Becker]

Scientific councils

- Scientific referent (for CNRS) of the PEPR exploratoire Molecularxiv [Anne Siegel]
- Comité de pilotage of the Mission for Interdisciplinarity (MITI) at CNRS [Anne Siegel]
- Scientific advisory Board of the LPHI lab [Anne Siegel]
- Scientific Advisory Board of the BioGenOuest network (37 platforms) [Emmanuelle Becker]
- Scientific Advisory Board of the GenOuest platform [Olivier Dameron]

Local responsibilities

- Member of the social committee of Univ. Rennes [Catherine Belleannée]
- Member of the emergency aid commission of Univ. Rennes and Rennes 2 [Catherine Belleannée]
- Member of CUMI (Commission des utilisateurs des moyens informatiques) of Inria Rennes [François Coste]
- Member of the thesis committee of the Matisse doctoral school [Olivier Dameron]
- Member of the Inria Rennes center council [Jeanne Got]

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Master : Emmanuelle Becker, "R, Data, and Visualisation (SIR + PAR + DVI)", 50h, Master 1 in Bioinformatics, Master 1 in Ecology and Environment, Univ. Rennes, France
- Master : Emmanuelle Becker, "Object oriented programming (OOP)", 60h, Master in Bioinformatics, Univ. Rennes, France
- Licence : Emmanuelle Becker, "Manipulate and Visualize Data (MVD)", 40h, double-diploma Licence degree "Life Science, Maths and Artificial Intelligence" , Univ. Rennes, France
- Master : Emmanuelle Becker, "Method (METH)", 15h, Master 2 in Computer Sciences, Univ. Rennes, France
- Master : Emmanuelle Becker, "Manipulate Data with R (MDR)", 30h, Bioinformatics Minor for Master Students, Univ. Rennes, France
- Licence : Emmanuelle Becker, "Biostatistics with R", 12h, L3 Life Science Licence, Univ. Rennes, France

- Licence: Catherine Belleannée, "Formal Languages", 20h, L3 informatique, Univ. Rennes, France
- Licence: Catherine Belleannée, "Projet professionnel et communication", 16h, L1 informatique, Univ. Rennes, France
- Licence: Catherine Belleannée, "Projet professionnel et communication", 12h, L2 informatique, Univ. Rennes, France
- Licence: Catherine Belleannée, Spécialité informatique, "Functional and immutable programming", 44h, L1 mathématiques, Univ. Rennes, France
- Master: Catherine Belleannée, "Answer Set Programming", 15h, M1 informatique, Univ. Rennes, France
- Master: Catherine Belleannée, "Programmation logique et contraintes", 32h, M1 informatique, Univ. Rennes, France
- Licence: Catherine Belleannée, "Outils formels pour l'informatique", 46h, L2 informatique, Univ. Rennes, France
- Licence: Catherine Belleannée, "Fondements mathématiques", 49h, L1 informatique, Univ. Rennes, France
- Licence : Myriam Bontonou, "Data: Sciences des Données", 36h, L2 Informatique, ISTIC, Univ. Rennes, France
- Licence : Myriam Bontonou, "Programmation Linéaire", 26h, L3 MIAGE, ISTIC, Univ. Rennes, France
- Licence : Myriam Bontonou, "GInitiation aux sciences informatiques", 6h, Licence 3 SVT-ME, Faculté des Sciences, Univ. Rennes, France
- Licence: Olivier Dameron, "Programmation 1", 98h, Licence 1 informatique, Univ. Rennes, France
- Licence: Olivier Dameron, "Introduction à l'IA", 6h, Licence 1 sciences de la vie et de l'environnement, Univ. Rennes, France
- Licence: Olivier Dameron, "Algorithmes de parcours de données", 25h, Licence 2 sciences de la vie, Univ. Rennes, France
- Licence: Olivier Dameron, "Graph Modeling and Algorithms", 21h, Licence 2 informatique, Univ. Rennes, France
- Licence: Olivier Dameron, "Programmation avancée", 36h, Licence 3 miage, Univ. Rennes, France
- Master: Olivier Dameron, "Data Engineering in Life Science", 36h, Master 2 in bioinformatics, Univ. Rennes, France
- Master: Olivier Dameron, "Internship", 10h, Master 2 in bioinformatics, Univ. Rennes, France
- Licence: Pablo Espana Gutierrez, "Langages Formels et Calculabilité", 20h, L3SIF, ENS Rennes, France
- Licence: Pablo Espana Gutierrez, "Remise à niveau MPI", 10h, L3SIF, ENS Rennes, France
- Licence: Pablo Espana Gutierrez, "Préparation à l'agrégation", 5h, ENS Rennes, France
- Master : Juliette Francis, "Apprentissage Statistique", 30h, Master 1 in Bioinfortmatics, Univ. Rennes, France
- Licence : Yann Le Cunff "Modélisation des phénomènes du vivant", 30h, L2 Biologie, Univ. Rennes, France

- Master: Yann Le Cunff, "Apprentissage statistique", 110h, Master 1 in Bioinfortmatics Univ. Rennes, France
- Master: Yann Le Cunff, "Biologie aux interfaces", 25h, Master 1 in Biology, Univ. Rennes, France
- Master: Yann Le Cunff,"Simulating dynamic systems in biology", 20h, Master 2 in bioinformatics, Univ. Rennes, France
- Master: Yann Le Cunff, "Applied Interdisciplinarity", 20h, Master 2 in biology, Univ. Rennes, France
- Master: Yann Le Cunff, "ESG Challenges of Artificial Intelligence", 20h, Master 2 ATN & RSE, Univ. Rennes, France
- Licence : Cécile Beust, "Informatique", 16h, Licence 1 PCSTM, Univ. Rennes, France
- Licence : Cécile Beust, "Data : Sciences des données", 24h, Licence 2 ISTN, ISTIC, France

11.2.2 Supervision

HDR

- HDR Yann Le Cunff "From Data to Phenotype: Integrating Data Structure and Prior Knowledge to Model Biological Systems" (defended in May 2025)

PhD thesis

- PhD in progress: Moussa Baddour, Extraction de phénotypes à partir de comptes-rendus médicaux textuels et mise en relation avec le génotype, started in May 2023, supervized by Olivier Dameron, M. De Tayrac (Rennes Hospital), S. Paquelet (b<>com) and T. Labbé (Orange)
- PhD in progress: Yael Tirlet, Integrative method for multi-omics data analysis with application to the activation and regulation of an endogeneized viral genome in a parasitoid wasp, started in Oct 2023, supervized by Emmanuelle Becker, Olivier Dameron and F. Legeai (INRAe)
- PhD in progress: Pablo Espana Gutierrez, Learning models with explicit dependencies between residues to predict protein functions, started in September 2023, supervized by François Coste and Olivier Dameron
- PhD in progress: Cécile Beust, Knowledge-guided rules for generating context-specific views on a knowledge graph: application to biological networks, started in Oct 2023, supervized by Emmanuelle Becker, Olivier Dameron and Nathalie Théret
- PhD in progress: Corentin Lucas, Integration of multi-modal data for longitudinal follow-up of Crohn's disease patients, started in Oct 2023, supervized by Emmanuelle Becker, Yann Le Cunff
- PhD in progress: Moana Aulagner, Modeling microbiota interactions in plants to build synthetic microbial communities for enhanced biocontrol and biostimulation, started in Oct 2023, supervized by Samuel Blanquart, Anne Siegel and C. Bartoli-Kautski (INRAe)
- PhD in progress: Océane Carpentier, Integrating prior knowledge for a better patient representation, started September 2024, supervized by Emmanuelle Becker, Yann Le Cunff, A. Bannay and N. Jay (LORIA)
- PhD in progress: Elisa Chenel, Study of protein co-evolution to identify interaction regions involved in TGFbeta growth factor activation, started in Oct 2024, supervized by Samuel Blanquart, François Coste and N. Nathalie Théret
- PhD in progress: Juliette Francis, Integration of heterogeneous data for phenotype prediction, started in October 2024, supervized by Yann Le Cunff and M. Mariadasssou (INRAe)

- PhD in progress: Ulysse Le Clanche, Knowledge-driven dataset FAIRification: from workflow runs to domain-specific annotations, started in October 2024, supervized by Olivier Dameron and A. Gaignard (CNRS, Institut du Thorax INSERM Nantes)
- PhD in progress: Pauline Giraud, Hybrid methods for ab initio inference of metabolic pathways in marine eukaryotes, started in November 2024, supervized by Anne Siegel and G. Markov (CNRS, Station biologique de Roscoff)
- PhD in progress: Noé Robert, Data mining for high-throughput genome screening: predicting microbial synthesis capabilities of targeted metabolites, started in November 2025, supervized by Anne Siegel and Hélène Falentin (INRAE).
- PhD in progress: Noryah Safla, Integration of a priori knowledge into spatial transcriptomics models: application to the characterization of immune response and prediction of treatment resistance in cholangiocarcinoma, started in November 2025, supervized by Yann Le Cunff, Myriam Bontonou and Joachim Lupberger (INSERM)

Internship

- M2 internship: Noryah Safla, Bioinformatic analysis of mechanisms of resistance to immunotherapy in liver cancer. Jan-Jul 2025 supervized by Yann Le Cunff and Myriam Bontonou.
- M1 internship Daniel Calvez Interprétation des fonctions enzymatiques inférées à partir de données de microbiotes. April - July 2025, supervized by Anne Siegel and Myriam Bontonou.
- M1 internship Samuel Fosse Raisonnement sur des métadonnées issues de génomes. October 2025 - May 2026, supervized by Anne Siegel

11.2.3 Doctoral advisory committees (CSID)

- Rim Ait Ben Aoumar, Univ. Rennes [Yann Le Cunff]
- Maria-Mafalda Almeida, Univ. Rennes [Emmanuelle Becker]
- Alexandre Asset, AgroParisTech [Yann Le Cunff]
- Juan Andrés Cisneros–Jacome, Univ. Rennes [Emmanuelle Becker]
- Maëlys Auffret, Univ. Rennes 2 [Emmanuelle Becker]
- Dorian Chenet, Univ. Rennes [Samuel Blanquart]
- Guénolé Dande, Univ. de Rennes [Olivier Dameron]
- Guillaume Doré, Univ. Rennes [Emmanuelle Becker]
- Jin-Mei Gao, Université Paris-Saclay [Emmanuelle Becker]
- Zainab Ghrayeb, Univ. de Rennes [Olivier Dameron]
- Silvia Grosso, INSA Lyon [Yann Le Cunff]
- Jedrej Kubica, Univ. Grenoble-Alpes [Yann Le Cunff]
- Mats Kohler–Dijkstra, Univ. Rennes [Emmanuelle Becker]
- Adam Lakdhari, Univ Rennes [Anne Siegel]
- Gabriel Mastrilli, Univ. de Rennes [François Coste]
- Meije Mathé, Univ. Toulouse [Olivier Dameron]
- Thiviya Parthipan, Univ. Rennes [Samuel Blanquart]

- Quentin Rouger, Univ. Rennes [Emmanuelle Becker]
- Quentin Vacher, Univ. Rennes [Emmanuelle Becker]
- Maelle Zonnequin, Sorbonne Université [Anne Siegel]

11.2.4 Juries

Referee of PhD thesis

- Sofiane Bouirdene, Univ. Laval Canada [Emmanuelle Becker]
- Samuel Dussault, Univ. Sherbrooke Canada [Olivier Dameron]
- Danilo Dursoniah, Univ. Lille [Anne Siegel]
- Ludivine Vasseur, Univ. Lille [Nathalie Théret]
- Catalina Gomez-Gonzalez, Univ. Lyon [Emmanuelle Becker]
- Yanis Asloudj, Univ. Bordeaux [Emmanuelle Becker]
- Rola Shaaban, Univ. Nantes [Emmanuelle Becker]

Member of PhD thesis juries

- Maelle Zonnequin, Univ. Paris Sorbonne [Anne Siegel]
- Matheo Lode, Univ. Rennes [Nathalie Théret, president]
- Fabien Foucher, Univ. Rennes [Nathalie Théret, president]
- Dzenis Koca, Univ. Grenoble Alpes [Emmanuelle Becker, president]

Member of habilitation thesis juries

- Clémence Frioux, Univ. Bordeaux [Emmanuelle Becker, referee]
- Yann Le Cunff, Univ. Rennes [Emmanuelle Becker]

11.3 Popularization

11.3.1 Participation in Live events

- Intervention and supervision of research workshops at "Réunions des Jeunes Mathématiciennes et Informaticiennes" (RJMI) organized by Animath and Femmes & mathématiques at ENS Rennes [Pablo Espana Gutierrez].
- Scientific outreach intervention for middle-school students as part of the “Parcours Avenir” programme, meeting with women scientists, Collège François Truffaut, Betton [Elisa Chenel]

12 Scientific production

12.1 Major publications

[1] 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 (May 2018). e1006146. DOI: 10.1371/journal.pcbi.1006146. URL: https://hal-univ-rennes1.archives-ouvertes.fr/hal-01807842 (cit. on p. 8).

[2] C. Belleannée, O. Sallou and J. Nicolas. 'Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling'. In: *PRIB2014 - Pattern Recognition in Bioinformatics, 9th IAPR International Conference*. Ed. by M. Comin, L. Kall, E. Marchiori, A. Ngom and J. Rajapakse. Vol. 8626. Lukas KALL. Stockholm, Sweden: Springer International Publishing, Aug. 2014, pp. 34–47. DOI: 10.1007/978-3-319-09192-1_4. URL: https://hal.inria.fr/hal-01059506 (cit. on p. 9).

[3] C. Bettembourg, C. Diot and O. Dameron. 'Optimal Threshold Determination for Interpreting Semantic Similarity and Particularity: Application to the Comparison of Gene Sets and Metabolic Pathways Using GO and ChEBI'. In: *PLoS ONE* (2015), p. 30. DOI: 10.1371/journal.pone.0133579. URL: https://hal.inria.fr/hal-01184934.

[4] P. Bordron, M. Latorre, M.-P. Cortés, M. Gonzales, S. Thiele, A. Siegel, A. Maass and D. Eveillard. 'Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach'. In: *MicrobiologyOpen* 5.1 (2015), pp. 106–117. DOI: 10.1002/mbo3.315. URL: https://hal.inria.fr/hal-01246173 (cit. on p. 9).

[5] J. Coquet, N. Théret, V. Legagneux and O. Dameron. 'Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering: Application to TGF-β Signaling'. In: *CMSB 2017 - 15th International Conference on Computational Methods in Systems Biology*. Lecture Notes in Computer Sciences. Darmstadt, Sept. 2017, p. 17. URL: https://hal.archives-ouvertes.fr/hal-01559249 (cit. on p. 10).

[6] F. Coste, G. Garet, A. Groisillier, J. Nicolas and T. Tonon. 'Automated Enzyme classification by Formal Concept Analysis'. In: *ICFCA - 12th International Conference on Formal Concept Analysis*. Cluj-Napoca, Romania: Springer, June 2014. URL: https://hal.inria.fr/hal-01063727 (cit. on p. 8).

[7] F. Coste and J. Nicolas. 'Learning local substitutable context-free languages from positive examples in polynomial time and data by reduction'. In: *ICGI 2018 - 14th International Conference on Grammatical Inference*. Vol. 93. Wrocław, Poland, Sept. 2018, pp. 155–168. URL: https://hal.inria.fr/hal-01872266.

[8] 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 (Sept. 2018), pp. i934–i943. DOI: 10.1093/bioinformatics/bty588. URL: https://hal.inria.fr/hal-01871600.

[9] C. Frioux, T. Schaub, S. Schellhorn, A. Siegel and P. Wanko. 'Hybrid Metitebolic Network Completion'. In: *Theory and Practice of Logic Programming* (Nov. 2018), pp. 1–23. URL: https://hal.inria.fr/hal-01936778 (cit. on p. 8).

[10] S. Prigent, C. Frioux, S. M. Dittami, S. Thiele, A. Larhlimi, G. Collet, G. Fabien, 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 (Jan. 2017), p. 32. DOI: 10.1371/journal.pcbi.1005276. URL: https://hal.inria.fr/hal-01449100 (cit. on p. 8).

[11] S. Videla, J. Saez-Rodriguez, C. Guziolowski and A. Siegel. 'caspo: a toolbox for automated reasoning on the response of logical signaling networks families'. In: *Bioinformatics* (2017). DOI: 10.1093/bioinformatics/btw738. URL: https://hal.inria.fr/hal-01426880 (cit. on pp. 9, 10).

12.2 Publications of the year

International journals

[12] 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'. In: *Environmental Microbiome* 20.142 (16th Nov. 2025), pp. 1–19. DOI: 10.1186/s40793-025-00749-8. URL: https://inria.hal.science/hal-05368332 (cit. on p. 17).

[13] O. Dennler, E. Chenel, F. Coste, S. Blanquart, C. Belleannée and N. Théret. 'FUSE-PhyloTree: Linking functions and sequence conservation modules of a protein family through phylogenomic analysis'. In: *Bioinformatics* (2025), btaf479. DOI: 10.1093/bioinformatics/btaf479. URL: https://hal.science/hal-05238604 (cit. on p. 18).

[14] R. Kemmerling, L.-E. Dintilhac, A. Zancarini, A. Mataigne, C. Mougel and N. Vannier. 'Carbon substrates utilization determine antagonistic fungal-fungal interactions among root-associated fungi'. In: *Frontiers in Microbiology* 16 (2025), p. 1645107. DOI: 10.3389/fmicb.2025.1645107. URL: https://hal.science/hal-05227488 (cit. on p. 18).

[15] C. A. Pfister, J. Berlinghof, M. Bogan, U. Cardini, A. Gobet, P. Hamon-Giraud, J. Hart, N. Jimenez, A. Siegel, E. Stanfield, M. Vallet, C. Leblanc, C. Rousseau, F. Thomas, W. Stock and S. M. Dittami. 'Evolutionary history and association with seaweeds shape the genomes and metabolisms of marine bacteria'. In: *MSphere* 10.6 (2025), e00996–24. DOI: 10.1128/msphere.00996-24. URL: https://hal.umontpellier.fr/hal-05195842 (cit. on p. 17).

[16] C. Rousseau, G. Tanguy, E. Legeay, S. Blanquart, A. Belcour, S. Rousvoal, P. Potin, C. Leblanc and S. Dittami. 'A duo of fungi and complex and dynamic bacterial community networks contribute to shape the Ascophyllum nodosum holobiont'. In: *Environmental Microbiome* (11th Dec. 2025), pp. 1–53. DOI: 10.1186/s40793-025-00825-z. URL: https://inria.hal.science/hal-05413201 (cit. on p. 17).

[17] U. Szachnowski, E. Becker, I. Stuparevic, M. Wery, O. Sallou, M. Boudet, A. Bretaudeau, A. Morillon and M. Primig. 'Pervasive formation of double-stranded RNAs by overlapping sense/antisense transcripts in budding yeast mitosis and meiosis'. In: *RNA* 31.4 (23rd Jan. 2025), pp. 497–513. DOI: 10.1261/rna.080290.124. URL: https://hal.science/hal-04935733 (cit. on p. 18).

International peer-reviewed conferences

[18] M. Baddour, O. Dameron, M. de Tayrac, S. Paquelet, P. Rollier, T. Labbé and M. Saleh. 'ACUITEE: A Comprehensive Tool for Visualization, Editing and Curating textual Annotations in Clinical Data'. In: *Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)*. EECSS 2025 - 11th World Congress on Electrical Engineering and Computer Systems and Sciences. Paris, France, 2025, pp. 1–9. URL: https://inria.hal.science/hal-04974765 (cit. on p. 16).

National peer-reviewed Conferences

[19] P. Hamon-Giraud, B. B. Pinto, J. Got, C. Rousseau, E. Corre, S. Dittami, G. Markov and A. Siegel. 'Methods for a species-specific genome-scale metabolic model designed for eukaryotes and applied to the Ascophyllum nodosum macroalga'. In: *Proceedings of JOBIM 2025*. JOBIM 2025 - Journées Ouvertes en Biologie, Informatique et Mathématiques. Bordeaux, France, 2025, pp. 109–117. URL: https://hal.science/hal-05374724 (cit. on p. 17).

[20] U. Le Clanche, S. Cohen-Boulakia, Y. L. Cunff, O. Dameron and A. Gaignard. 'Assessing bioinformatics software annotations: bio.tools case-study'. In: *Proceedings JOBIM*. JOBIM 2025 - Journées Ouvertes en Biologie, Informatique et Mathématiques. Bordeaux & Online, France, 2025, pp. 1–7. URL: https://inria.hal.science/hal-05096849 (cit. on p. 16).

Conferences without proceedings

[21] L. Angevin, P. Peterlongo, R. Vicedomini, A. Siegel, J. Got and E. Roux. 'Metagenomic taxonomic assignment using Nanopore reads, reconstruction of metabolic networks and prediction of metabolite production'. In: 2025 - Journées Nutrition et Ecosystèmes Microbiens. Rennes, France, 2025. URL: https://hal.science/hal-05384891 (cit. on p. 18).

[22] J. Audemard, N. Creusot, J. Leloup, C. Duval, S. Halary, J. Got, B. Marie, G. V. Markov, B. Diémé and C. Frioux. 'Metagenome-scale metabolic modelling for the characterization of cross-feeding interactions in freshwater cyanobacteria-associated microbial communities'. In: JOBIM 2025 - Journées Ouvertes en Biologie, Informatique et Mathématiques. Bordeaux, France, 10th July 2025. URL: https://inria.hal.science/hal-05281552 (cit. on p. 18).

[23] N. Robert, J. Got, P. Hamon-Giraud, H. Falentin and A. Siegel. 'Evidencing strain-dependency of metabolic pathways within 1,494 lactic bacteria genomes with the in silico screening Prolipipe pipeline'. In: *Proceedings of JOBIM 2025*. JOBIM 2025 - Journées Ouvertes en Biologie, Informatique et Mathématiques. Bordeaux, France, 2025. URL: https://hal.science/hal-05045657.

Reports & preprints

[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.

Other scientific publications

[25] C. Beust, E. Becker, N. Théret and O. Dameron. 'Biological Knowledge Extraction from BioPAX Graphs'. In: IRISA - DKM (Data and Knowledge Management) days 2025. Rennes (Campus de Beaulieu), France, 4th Mar. 2025. URL: https://inria.hal.science/hal-05433535 (cit. on p. 16).

[26] P. Espana Gutierrez and F. Coste. 'Identifying coevolving residues by factoring out the evolutionary distance covariance matrix'. In: LEGEND 2025 - Machine Learning for Evolutionary Genomics Data. Aussois, France, 2025. URL: https://inria.hal.science/hal-05424193 (cit. on p. 19).

[27] P. Le Corre, A. Bretaudeau and Y. Le Cunff. 'MLOps best practices for bioinformatics'. In: JOBIM - Journées Ouvertes en Biologie, Informatique et Mathématiques. Bordeaux (France), France, 2025, pp. 1–1. DOI: 10.1093/bioinformatics/btad6[2]. URL: https://hal.science/hal-05168225 (cit. on p. 16).

Software

[28] [SW] J. Pécréaux, H. Bouvrais and Y. Le Cunff, *DiLiPop — Code supporting Bouvrais et al. EMBO Reports (2021)* 1st Dec. 2025. LIC: CeCILL Free Software License Agreement v2.1. DOI: 10.5281/zenodo.4552484, HAL: ⟨hal-05236907⟩, URL: https://hal.science/hal-05236907, VCS: https://github.com/JacquesPecreaux/DiLiPop/tree/Bouvrais_EmboR_21.

12.3 Cited publications

[29] G. Andrieux, M. Le Borgne and N. Théret. 'An integrative modeling framework reveals plasticity of TGF-Beta signaling'. In: *BMC Systems Biology* 8.1 (2014), p. 30. DOI: 10.1186/1752-0509-8-30. URL: http://www.hal.inserm.fr/inserm-00978313 (cit. on pp. 9–11).

[30] 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 (Dec. 2020). DOI: 10.7554/eLife.61968. URL: https://inria.hal.science/hal-02395024 (cit. on p. 9).

[31] A. Belcour, J. Girard, M. Aite, L. Delage, C. Trottier, C. Marteau, C. J.-J. Leroux, S. M. Dittami, P. Sauleau, E. Corre, J. Nicolas, C. Boyen, C. Leblanc, J. Collén, A. Siegel and G. V. Markov. 'Inferring Biochemical Reactions and Metabolite Structures to Understand Metabolic Pathway Drift'. In: *iScience* 23.2 (Feb. 2020), p. 100849. DOI: 10.1016/j.isci.2020.100849. URL: https://hal.inria.fr/hal-01943880 (cit. on p. 11).

[32] A. Belcour, J. Got, M. Aite, L. Delage, J. Collén, C. Frioux, C. Leblanc, S. M. Dittami, S. Blanquart, G. V. Markov and A. Siegel. 'Inferring and comparing metabolism across heterogeneous sets of annotated genomes using AuCoMe'. In: *Genome Research* 33 (June 2023), pp. 972–987. DOI: 10.1101/gr.277056.122. URL: https://hal.science/hal-04192851 (cit. on p. 9).

[33] T. Berners Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt and D. J. Weitzner. 'A Framework for Web Science'. In: *Foundations and Trends in Web Science* 1.1 (2007), pp. 1–130 (cit. on p. 7).

[34] C. Bettembourg, C. Diot and O. Dameron. 'Semantic particularity measure for functional characterization of gene sets using gene ontology'. In: *PLoS ONE* 9.1 (2014). e86525. DOI: 10.1371/journal.pone.0086525. URL: https://hal.inria.fr/hal-00941850 (cit. on p. 11).

[35] S. Blanquart, J.-S. Varré, P. Guertin, A. Perrin, A. Bergeron and K. M. Swenson. 'Assisted transcriptome reconstruction and splicing orthology'. In: *BMC Genomics* 17.10 (Nov. 2016), p. 786. DOI: 10.1186/s12864-016-3103-6. URL: https://doi.org/10.1186/s12864-016-3103-6 (cit. on p. 9).

[36] P. Blavy, F. Gondret, S. Lagarrigue, J. Van Milgen and A. Siegel. 'Using a large-scale knowledge database on reactions and regulations to propose key upstream regulators of various sets of molecules participating in cell metabolism'. In: *BMC Systems Biology* 8.1 (2014), p. 32. DOI: 10.1186/1752-0509-8-32. URL: https://hal.inria.fr/hal-00980499 (cit. on pp. 9, 11).

[37] P. Bordron, M. Latorre, M.-P. Cortés, M. Gonzales, S. Thiele, A. Siegel, A. Maass and D. Eveillard. 'Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach'. In: *MicrobiologyOpen* 5.1 (2015), pp. 106–117. DOI: 10.1002/mbo3.315. URL: https://hal.inria.fr/hal-01246173 (cit. on p. 11).

[38] M. Bouguéon, P. Boutillier, J. Feret, O. Hazard and N. Théret. 'The rule-based model approach. A Kappa model for hepatic stellate cells activation by TGFB1'. In: *Systems Biology Modelling and Analysis: Formal Bioinformatics Methods and Tools*. Ed. by E. D. Maria. Wiley, Nov. 2022, pp. 1–76. URL: https://inria.hal.science/hal-03388100 (cit. on p. 12).

[39] P. Boutillier, F. Camporesi, J. Coquet, J. Feret, K. Q. Lý, N. Théret and P. Vignet. 'KaSa: A Static Analyzer for Kappa'. In: *CMSB 2018 - 16th International Conference on Computational Methods in Systems Biology*. Ed. by M. Češka and D. Šafránek. Vol. 11095. LNCS. Brno, Czech Republic: Springer Verlag, Sept. 2018, pp. 285–291. DOI: 10.1007/978-3-319-99429-1_17. URL: https://hal-univ-rennes1.archives-ouvertes.fr/hal-01888951 (cit. on p. 12).

[40] A. Bretaudeau, F. Coste, F. Humily, L. Garczarek, G. Le Corguillé, C. Six, M. Ratin, O. Collin, W. M. Schluchter and F. Partensky. 'CyanoLyase: a database of phycobilin lyase sequences, motifs and functions'. In: *Nucleic Acids Research* (Nov. 2012), p. 6. DOI: 10.1093/nar/gks1091. URL: https://hal.inria.fr/hal-01094087 (cit. on p. 8).

[41] 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 (Feb. 2020), pp. 1–11. DOI: 10.3389/fmars.2020.00085. URL: https://hal.inria.fr/hal-02866101 (cit. on p. 11).

[42] J. Coquet, N. Théret, V. Legagneux and O. Dameron. 'Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering: Application to TGF-β Signaling'. In: *CMSB 2017 - 15th International Conference on Computational Methods in Systems Biology*. Lecture Notes in Computer Sciences. Darmstadt, France, Sept. 2017, p. 17. URL: https://hal.archives-ouvertes.fr/hal-01559249 (cit. on p. 11).

[43] M.-P. Cortés, S. N. Mendoza, D. Travisany, A. Gaete, A. Siegel, V. Cambiazo and A. Maass. 'Analysis of Piscirickettsia salmonis Metabolism Using Genome-Scale Reconstruction, Modeling, and Testing'. In: *Frontiers in Microbiology* 8 (Dec. 2017), p. 15. DOI: 10.3389/fmicb.2017.02462. URL: https://hal.inria.fr/hal-01661270 (cit. on pp. 9, 11).

[44] F. Coste, G. Garet, A. Groisillier, J. Nicolas and T. Tonon. 'Automated Enzyme classification by Formal Concept Analysis'. In: *ICFCA - 12th International Conference on Formal Concept Analysis*. Cluj-Napoca, Romania: Springer, June 2014. URL: https://hal.inria.fr/hal-01063727 (cit. on p. 10).

[45] O. Dennler. ‘Caractérisation en modules fonctionnels de la famille de protéines ADAMTS / ADAMTSL’. MA thesis. Univ Rennes, June 2019. URL: https://hal.inria.fr/hal-02403084 (cit. on p. 12).

[46] O. Dennler. ‘Caractérisation en modules fonctionnels des protéines ADAMTS-TSL, par approches de phylogénies’. Theses. Université Rennes 1, Dec. 2022. URL: https://hal.science/tel-03927428 (cit. on p. 12).

[47] O. Dennler, S. Blanquart, F. Coste, C. Belleannée and N. Theret. *Phylogenetic Functional Module Characterization of the ADAMTS / ADAMTS like Protein Family*. WABI 2021 - Workshop on Algorithms in Bioinformatics. Poster. Aug. 2021. URL: https://hal.archives-ouvertes.fr/hal-03543214 (cit. on p. 12).

[48] S. M. Dittami, T. Barbeyron, C. Boyen, J. Cambefort, G. Collet, L. Delage, A. Gobet, A. Groisillier, C. Leblanc, G. Michel, D. Scornet, A. Siegel, J. E. Tapia and T. Tonon. ‘Genome and metabolic network of "Candidatus Phaeomarinobacter ectocarpi" Ec32, a new candidate genus of Alphaproteobacteria frequently associated with brown algae’. In: *Frontiers in Genetics* 5 (2014), p. 241. DOI: 10.3389/fgene.2014.00241. URL: https://hal.inria.fr/hal-01079739 (cit. on p. 11).

[49] 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 (cit. on p. 10).

[50] K. Faust and J. Raes. ‘Microbial interactions: from networks to models’. In: *Nat. Rev. Microbiol.* 10.8 (July 2012), pp. 538–550 (cit. on p. 8).

[51] M. Y. Galperin, D. J. Rigden and X. M. Fernández-Suárez. ‘The 2015 Nucleic Acids Research Database Issue and molecular biology database collection’. In: *Nucleic acids research* 43.Database issue (2015), pp. D1–D5 (cit. on p. 7).

[52] L. Garczarek, U. Guyet, H. Doré, G. Farrant, M. Hoebeke, L. Brillet-Guéguen, A. Bisch, M. Ferrieux, J. Siltanen, E. Corre, G. Le Corguillé, M. Ratin, F. Pitt, M. Ostrowski, M. Conan, A. Siegel, K. Labadie, J.-M. Aury, P. Wincker, D. Scanlan and F. Partensky. ‘Cyanorak v2.1: a scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes’. In: *Nucleic Acids Research* 49.D1 (Oct. 2020), pp. D667–D676. DOI: 10.1093/nar/gkaa958. URL: https://hal.archives-ouvertes.fr/hal-02988562 (cit. on p. 10).

[53] M. Gebser, R. Kaminski, B. Kaufmann and T. Schaub. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2012 (cit. on p. 6).

[54] F. Gondret, I. Louveau, M. Houee, D. Causeur and A. Siegel. ‘Data integration’. In: *Meeting INRA-ISU*. Ames, United States, Mar. 2015, p. 11. URL: https://hal.archives-ouvertes.fr/hal-01210940 (cit. on p. 11).

[55] U. Guyet, N. T. Nguyen, H. Doré, J. Haguait, J. Pittera, M. Conan, M. Ratin, E. Corre, G. Le Corguillé, L. A. Brillet-Guéguen, M. M. Hoebeke, C. Six, C. Steglich, A. Siegel, D. Eveillard, F. Partensky and L. Garczarek. ‘Synergic Effects of Temperature and Irradiance on the Physiology of the Marine Synechococcus Strain WH7803’. In: *Frontiers in Microbiology* 11 (July 2020). DOI: 10.3389/fmicb.2020.01707. URL: https://hal.sorbonne-universite.fr/hal-02929424 (cit. on p. 10).

[56] F. Herault, A. Vincent, O. Dameron, P. Le Roy, P. Cherel and M. Damon. ‘The longissimus and semimembranosus muscles display marked differences in their gene expression profiles in pig’. In: *PLoS ONE* 9.5 (2014). e96491. DOI: 10.1371/journal.pone.0096491. URL: https://hal.inria.fr/hal-00989635 (cit. on p. 11).

[57] H. Kleinjan, C. Frioux, G. Califano, M. Aite, E. Fremy, E. Karimi, E. Corre, T. Wichard, A. Siegel, C. Boyen and S. M. Dittami. 'Insights into the potential for mutualistic and harmful host-microbe interactions affecting brown alga freshwater acclimation'. In: *Molecular Ecology* 32.3 (2022), pp. 703–723. DOI: 10.1111/mec.16766. URL: https://hal.science/hal-03868898 (cit. on p. 11).

[58] D. Mandakovic, Á. Cintolesi, J. Maldonado, S. Mendoza, M. Aite, A. Gaete, F. Saitua, M. Allende, V. Cambiazo, A. Siegel, A. Maass, M. Gonzalez and M. Latorre. 'Genome-scale metabolic models of Microbacterium species isolated from a high altitude desert environment'. In: *Scientific Reports* 10.1 (Dec. 2020), pp. 1–12. DOI: 10.1038/s41598-020-62130-8. URL: https://hal.inria.fr/hal-02524471 (cit. on p. 11).

[59] D. Nègre, M. Aite, A. Belcour, C. Frioux, L. Brillet-Guéguen, X. Liu, P. Bordron, O. Godfroy, A. P. Lipinska, C. Leblanc, A. Siegel, S. Dittami, E. Corre and G. V. Markov. 'Genome–Scale Metabolic Networks Shed Light on the Carotenoid Biosynthesis Pathway in the Brown Algae Saccharina japonica and Cladosiphon okamuranus'. In: *Antioxidants* 8.11 (Nov. 2019), p. 564. DOI: 10.3390/antiox8110564. URL: https://hal.inria.fr/hal-02395080 (cit. on p. 10).

[60] M. Popova, I. Fakih, E. Forano, A. Siegel, R. Muñoz-Tamayo and D. Morgavi. 'Rumen microbial genomics: from cells to genes (and back to cells)'. In: *CAB Reviews Perspectives in Agriculture Veterinary Science Nutrition and Natural Resources* 2022 (Aug. 2022). DOI: 10.1079/cabireviews202217025. URL: https://hal.inrae.fr/hal-03929845 (cit. on p. 11).

[61] S. Prigent, G. Collet, S. M. Dittami, L. Delage, F. Ethis de Corny, O. Dameron, D. Eveillard, S. Thiele, J. Cambefort, C. Boyen, A. Siegel and T. Tonon. 'The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond'. In: *Plant Journal* (Sept. 2014), pp. 367–81. DOI: 10.1111/tpj.12627. URL: https://hal.archives-ouvertes.fr/hal-01057153 (cit. on pp. 9, 10).

[62] S. Prigent, C. Frioux, S. M. Dittami, S. Thiele, A. Larhlimi, G. Collet, G. Fabien, 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 (Jan. 2017), p. 32. DOI: 10.1371/journal.pcbi.1005276. URL: https://hal.inria.fr/hal-01449100 (cit. on p. 11).

[63] M. H. Saier, V. S. Reddy, B. V. Tsu, M. S. Ahmed, C. Li and G. Moreno-Hagelsieb. 'The Transporter Classification Database (TCDB): recent advances'. In: *Nucleic Acids Res.* 44.D1 (Jan. 2016), pp. D372–379 (cit. on p. 8).

[64] D. B. Searls. 'String variable grammar: A logic grammar formalism for the biological language of DNA'. In: *The Journal of Logic Programming* 24.1 (1995). Computational Linguistics and Logic Programming, pp. 73–102. DOI: http://dx.doi.org/10.1016/0743-1066(95)00034-H. URL: http://www.sciencedirect.com/science/article/pii/074310669500034H (cit. on p. 9).

[65] Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, R. Iyer, M. C. Schatz, S. Sinha and G. E. Robinson. 'Big Data: Astronomical or Genomical?' In: *PLoS biology* 13.7 (2015), e1002195 (cit. on p. 6).

[66] H. Talibart. 'Comparison of homologous protein sequences using direct coupling information by pairwise Potts model alignments'. Theses. Université Rennes 1, Feb. 2021. URL: https://theses.hal.science/tel-03376771 (cit. on p. 8).

[67] H. Talibart and F. Coste. 'PPalign: optimal alignment of Potts models representing proteins with direct coupling information'. In: *BMC Bioinformatics* 22.317 (Dec. 2021), pp. 1–22. DOI: 10.1186/s12859-021-04222-4. URL: https://hal.inria.fr/hal-03264248 (cit. on p. 8).

[68] N. R. Tartaglia, A. Nicolas, V. DE REZENDE RODOVALHO, B. S. R. d. Luz, V. Briard-Bion, Z. Krupova, A. Thierry, F. Coste, A. Burel, P. P. Martin, J. Jardin, V. Azevedo, Y. Le Loir and E. Guédon. 'Extracellular vesicles produced by human and animal Staphylococcus aureus strains share a highly conserved core proteome'. In: *Scientific Reports* 10.1 (Apr. 2020), pp. 1–13. DOI: 10.1038/s41598-020-64952-y. URL: https://hal.inrae.fr/hal-02638124 (cit. on p. 8).

[69] N. Theret, F. Bouezzeddine, F. Azar, M. Diab-Assaf and V. Legagneux. 'ADAM and ADAMTS Proteins, New Players in the Regulation of Hepatocellular Carcinoma Microenvironment'. In: *Cancers* 13.7 (2021), p. 1563. DOI: 10.3390/cancers13071563. URL: https://hal.archives-ouvertes.fr/hal-03215892 (cit. on p. 12).

[70] N. Theret, J. Feret, A. Hodgkinson, P. Boutillier, P. Vignet and O. Radulescu. 'Integrative models for TGF-beta signaling and extracellular matrix'. In: *Extracellular Matrix Omics*. Ed. by S. Ricard-Blum. Vol. 7. Biology of Extracellular Matrix. Springer, Dec. 2020, p. 17. DOI: 10.1007/978-3-030-58330-9_10. URL: https://hal.inria.fr/hal-02458073 (cit. on p. 12).

[71] R. Verborgh, M. Vander Sande, O. Hartig, J. Van Herwegen, L. De Vocht, B. De Meester, G. Haesendonck and P. Colpaert. 'Triple Pattern Fragments: a Low-cost Knowledge Graph Interface for the Web'. In: *Journal of Web Semantics* 37–38 (Mar. 2016), pp. 184–206. DOI: doi:10.1016/j.websem.2016.03.003. URL: http://linkeddatafragments.org/publications/jws2016.pdf (cit. on p. 7).