IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2018

Project-Team DYLISS

Dynamics, Logics and Inference for biological Systems and Sequences

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Biology

Table of contents

Project-Team DYLISS

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 July 01

Keywords:

Computer Science and Digital Science:

A3.1.1. - Modeling, representation
A3.1.2. - Data management, quering and storage
A3.1.7. - Open data
A3.1.10. - Heterogeneous data
A3.2.3. - Inference
A3.2.4. - Semantic Web
A3.2.5. - Ontologies
A3.2.6. - Linked data
A3.3.3. - Big data analysis
A7.2. - Logic in Computer Science
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A9.1. - Knowledge
A9.2. - Machine learning
A9.7. - AI algorithmics
A9.8. - Reasoning

Other Research Topics and Application Domains:

B1.1.2. - Molecular and cellular biology
B1.1.7. - Bioinformatics
B1.1.10. - Systems and synthetic biology
B2.2.3. - Cancer
B2.2.5. - Immune system diseases

1. Team, Visitors, External Collaborators

Research Scientists

Samuel Blanquart [Inria, Researcher, from May 2018]
François Coste [Inria, Researcher]
Jacques Nicolas [Inria, Senior Researcher, until May 2018, HDR]
Anne Siegel [Team Leader, CNRS, Senior Researcher, HDR]

Faculty Members

Catherine Belleannée [Univ de Rennes I, Associate Professor]
Olivier Dameron [Univ de Rennes I, Associate Professor, HDR]

Post-Doctoral Fellows

Célia Biane-Fourati [Inria, from Dec 2018]
Maxime Folschette [Univ de Rennes I, UMR IRSET laboratory, until Oct 2018]
Vijay Ingalalli [Inria]
Clémence Frioux [Inria, phD until Nov 2018, CNRS non permanent researcher since Dec 2018]

PhD Students

Lucas Bourneuf [Univ de Rennes I]
Mael Conan [Univ de Rennes I, UMR IRSET laboratory]
Nicolas Guillaudeux [Univ de Rennes I, from Oct 2018]
Marine Louarn [Inserm, Inria-Inserm funding]
Yann Rivault [Ecole des hautes études en santé publique, until Sep 2018]
Hugo Talibart [Univ de Rennes I]
Méline Wery [Univ de Rennes I]
Pierre Vignet [CNRS, until Nov 2018, PhD since Dec. 2018, UMR IRSET laboratory]

Technical staff

Jeanne Got [CNRS, permanent position]
Meziane Aite [CNRS]
Guillaume Alviset [INRA, from Oct 2018]
Arnaud Belcour [CNRS]
Marie Chevallier [CNRS, UMR Ecobio laboratory, until Sep 2018]
Xavier Garnier [Inria since Oct 2018, previously INRA]

Administrative Assistant

Marie Le Roic [Univ de Rennes I]

External Collaborators

Emmanuelle Becker [Univ de Rennes 1, UMR IRSET laboratory, Assistant professor, from Feb 2018]
François Moreews [INRA, Engineer]
Denis Tagu [INRA, Scientist, HDR]
Nathalie Théret [INSERM, UMR IRSET laboratory, Scientist, HDR]

2. Overall Objectives

2.1. 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 and with genomics 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 to analyze the dynamical response of a biological system, model and classify families of gene products with sensitive and expressive languages, identify 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, 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. Although "omics" technologies have allowed the number of measured compounds in a systems to increase tremendously, 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 position of developing experimental strategies to progressively shrink the space of hypotheses and gain in 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 Sec. 3.2) (2) Understanding metabolism (see Sec. 3.3) (3) Characterizing regulatory and signaling phenotypes (see Sec. 3.4).

Implementing methods in software and platforms Seven platforms have been developped in the team for the last five years: Askomics, AuReMe, FinGoc, Caspo, Cadbiom, Logol, 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, dynamical models) which are compatible with both 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. Computer science – 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 objectives 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 knowledge, that is, the explicitation of relations occurring in structured data. In this setting, our main goal is to facilitate and optimize the integration 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) [43], 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 and expression of genes in non-model species, such as enzymes and isoforms functions in 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, abstract gene models) enabling us to represent sequence conservation, sets of short and degenerated patterns and crossing or distant dependencies. This requires both to determine classes of formal syntactic models allowing to handle 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 [47]. In our opinion, life sciences cumulates 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** [42]. To explore this research field, we use techniques from symbolic data mining (Semantic Web technologies, symbolic clustering, constraint satisfaction and grammatical modelling) 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 provide a relevant framework, as demonstrated by the success of Linked (Open) Data [34]. 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 the combination of *linked data fragments* [48], 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 functional static map of biological models that can be analyzed and/or compressed. This requires to linking 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, 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 [url]. *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 to (1) take advantage of the information readily available in the LOD cloud for analyzing there own data and (2) 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.

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 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 [package]. (2) The PowerGrasp python package implements graph compression methods oriented toward visualization, and based on power graph analysis [package]. (3) The iggy package enables the repairing of an interaction graph with respect to expression data. [Python package]

3.3. Metabolism: from enzyme sequences to systems ecology

Our researches in bioinformatics in relation with metabolic processes are driven by the understanding of 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 enzymatic 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 taking a sample of functional sequences as input to infer a grammar representing their key syntactical characteristics, including dependencies between residues. Our first goal is to enable an automatic semi-supervised refinement of enzymes classification [5] by combining the Protomata-Learner [38] framework - which captures local dependencies - with formal concept analysis. More challenging, we are exploring the learn of grammars representing long-distance dependencies such as those exhibited by contacts of amino-acids that are far in the sequence but close in the 3D protein folding.

System level: enriching and comparing metabolic networks for non-model organisms Non-model organisms are associated with often incomplete and poorly annotated sequences, leading to draft networks of their metabolism which largely suffer from incompleteness. in former studies, the team has developed several methods to improve the quality of eukaryotes metabolic networks, by solving several variants of the so-called *Metabolic Network gap-filling problem* with logical programming approaches [9], [8]. 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 A new emerging field is system ecology, which aims at building predictive models of species interactions within an ecosystem for deciphering cooperative and competitive relationships between species [41].This field raise 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 from transporter identification, based on our expertise in metabolic network gap-filling. A second very challenging focus is the prediction of transporters families by obtaining refined characterization of transporters, which are quite unexplored apart from specific databases [45].

3.3.2. Associated software tools

Protomata is a machine learning suite for the inference of automata characterizing (functional) families of proteins from available sequences by modeling alternative local dependencies. They are well suited to predict new family members with a high specificity [url]. The tool builds sequences alignments (partial and local), learns automata and searches for new family members in sequence databases. Applications of Protomata tools include automatic updating of the cyanolase database [38] and the refinement of the classification of HAD enzymes [5].

AuReMe workspace is designed for tractable reconstruction of metabolic networks [url]. The toolbox allows for the Automatic Reconstruction of Metabolic networks based on the combination of multiple heterogeneous data and knowledge sources [12]. The main added-values are the inclusion of graph-based tools relevant for the study of non-classical organisms (Meneco, Menetools, Shogen packages), the possibility to trace the reconstruction and curation procedures (Padmet package), and the exploration of reconstructed metabolic networks with wikis (wiki-export package). It has been used for reconstructing metabolic networks of micro and macro-algae [44], extremophile bacteria [39] and communities of organisms [3].

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 involves 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. A 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 subject 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...), that is, short, degenerated and overlapping functional sequences, are modeled by relying on the high capability of SVG grammars to deal with structure and ambiguity [46].

System level: extracting causal signatures of complex phenotypes with systems biology frameworks The main challenge we address 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 [10], multi-layer reactions in interaction graphs [36], and multi-layer information in large-scale Petri nets [33]. 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 modelling and matching [url] 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) [1]. *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 software Cell ASP Optimizer (*Caspo*) constitutes a pipeline for automated reasoning on logical signaling networks (learning, classifying, designing experimental perturbations, identifying controllers, take time-series into account) [url]. The software handles inherent experimental noise but enumerating all different logical networks which are compatible with a set of experimental observations [10]. The main advantage is that is 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 [url] It is based on Guarded transition semantic and allows synchronization events to be investigated in large-scale biological networks [33]. For instance, it was designed to allow controler of phenotypes in large-scale knowledge databases (PID) to be curated and analyzed [4].

4. Application Domains

4.1. Application fields in biology

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

Marine Biology: seaweed enzymes and metabolism Our main field of field **marine biology**, in close collaborations with the Roscoff Biological Station, in the framework of the Idealg project. Our goal is to apply our methods based on combinatorial optimization to the reconstruction of genome-scale metabolic networks, the understanding of microbial consortia, and classification of enzyme families. A main application model is *E. Siliculosus*, for which we reconstructed a metabolic network, predicted *HAD* proteins, and suggested new annotations of 56 genes based on metabolic network considerations. These approaches also shed light on evolution of metabolic processes. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. For instance, our analyses suggest that the bacterium *Ca. P. ectocarpi* is able to provide both β-alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway. These studies are now extended to the understanding of full host-microbial interactions.

Micro-biology: elucidating the functioning of extremophile consortiums of bacteria In this application field, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile. In order to elucidate the main characteristics of these bacteria. In particular, genome-scale metabolic network have been reconstructed for bacteria involved in biomining processes and in Salmon pathogenicity, already leading to a better understanding of bacterial interactions and growth.

Agriculture and environmental sciences: upstream controllers of pork and pea-aphid metabolism and regulation In this application field, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues, in collaboration with the INRA centers of Rennes (Pegase, Igepp, Scribe). This is a relevant application field for our researches work on the identification of upstream regulators within large-scale knowledge databases and on semantic-based analysis of metabolic networks, in order to interpreting differences of gene expression in pork meat and figure out the main gene-regulators of the response of porks to several diets, or to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants.

Health: deciphering pathways involved in the TGF-β signalling network This topic is studied with the IRSET laboratory of Rennes. TGF-β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis with both benefical anti-tumor effect and pro-tumor effect. Deciphering protumor versus antitumor signaling requires to take into account a system-wide view and develop predictive models for therapeutic benefit. We are developping *Cadbiom* in order to identify gene networks associated with innate immune response to viral infection that combine TGF-β and interleukine signaling pathways.

5. Highlights of the Year

5.1. Highlights of the Year

The main novelty of the year is the publication associated with software AuReMe for metabolic network reconstruction in Plos Computational Biology [12], and the development of Miscoto, a tool to design synthetic microbial communities [14], presented at the ECCB conference.

6. New Software and Platforms

6.1. AskOmics

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).
NEWS OF THE YEAR: (1) Improvements: Bugfixes and ui improvements in response to user feedback. (2) Versionning: Regular development cycle: a new version of AskOmics will be available every 3 months (currently 18.10, next 19.01) (3) Deployment: Deployment has been improved with docker and docker-compose. Virtual machine images are available on genostack to easy deploy AskOmics and Virtuoso (4) Federation: AskOmics can perform federated queries on mutliple triplestores, including other AskOmics endpoints, but also external endpoints like uniprot or dbpedia. (still in development)

- Authors: Charles Bettembourg, Xavier Garnier, Anthony Bretaudeau, Fabrice Legeai, Olivier Dameron, Olivier Filangi and Yvanne Chaussin
- Partners: Université de Rennes 1 - CNRS - INRA
- Contact: Olivier Dameron
- URL: https://github.com/askomics/askomics

6.2. CADBIOM

Computer Aided Design of Biological Models
KEYWORDS: Health - Biology - Biotechnology - Bioinformatics - Systems Biology
FUNCTIONAL DESCRIPTION: The Cadbiom software provides a formal framework to help the modeling of biological systems such as cell signaling network with Guarder Transition Semantics. It allows synchronization events to be investigated in biological networks among large-scale network in order to extract signature of controllers of a phenotype. Three modules are composing Cadbiom. 1) The Cadbiom graphical interface is useful to build and study moderate size models. It provides exploration, simulation and checking. For large-scale models, Cadbiom also allows to focus on specific nodes of interest. 2) The Cadbiom API allows a model to be loaded, performing static analysis and checking temporal properties on a finite horizon in the future or in the past. 3) Exploring large-scale knowledge repositories, since the translations of the large-scale PID repository (about 10,000 curated interactions) have been translated into the Cadbiom formalism.

NEWS OF THE YEAR: - Comprehensive command line to run the calculations and analyze the generated results. - Module designed to produce models through the interpretation of various databases or ontologies, formalized according to the BioPAX standard. - Update of the site and the documentation.

- Participants: Geoffroy Andrieux, Michel Le Borgne, Nathalie Theret, Nolwenn Le Meur, Pierre Vignet and Anne Siegel
- Contact: Anne Siegel
- URL: http://cadbiom.genouest.org

6.3. MiSCoTo

Microbiota Screening and COmmunity Selection with TOpology
KEYWORDS: Metabolic networks - ASP - Answer Set Programming - Logic programming
SCIENTIFIC DESCRIPTION: MiSCoTo solves combinatorial problems using Answer Set Programming. It aims at minimizing either the number of selected species or both the number of selected species and the cost of the interaction between them, characterized by the number of metabolic exchanges. In the first case, the level of modeling is called lumped or mixed-bag, in the latter, it is compartmentalized.
FUNCTIONAL DESCRIPTION: Metabolic networks are composed of biochemical reactions and gather the expected metabolic capabilities of species. For organisms that live in interaction altogether (microbiotas), complementarity between these networks can be exploited to predict cooperation events. This software takes as inputs metabolic networks for various species (host, symbionts of the microbiota), components of the growth medium and a metabolic objective (metabolites to be produced), and aims at selecting a minimal set of symbionts to ensure the metabolic objective can be achieved. The software can use two types of modelings: a simplified one and another that takes into account the cost of metabolic exchanges and aims at minimizing it.
NEWS OF THE YEAR: Release of the first version of the software

- Participants: Clémence Frioux, Anne Siegel, Enora Fremy and Camille Trottier
- Contact: Anne Siegel
- Publication: Scalable and exhaustive screening of metabolic functions carried out by microbial consortia
- URL: https://github.com/cfrioux/miscoto

6.4. Pathmodel

KEYWORDS: ASP - Answer Set Programming - Metabolic networks - Metabolic Pathway Drift - Bioinformatics - Systems Biology - Metabolomics
SCIENTIFIC DESCRIPTION: This tool is a prototype of the Metabolic Pathway Drift concept. This concept states that metabolic pathways undergo substantial turnover. The reactions involved in a pathway can change between species (change in reaction order or replacement of an enzyme by another one). Another goal of this tool is to linked genomics and metabolomics data. To implement this concept, Pathmodel uses the Answer Set Programming language. The input are the reactants and products involved in the pathway, known reactions occurring between these molecules, known m/z ratio, known domains shard by these molecules, an initial molecule and a goal molecule. Using these data, Pathmodel will infer reactions between molecules to reach the goal molecule using the known reactions. The result consist of potential alternative pathways for the studied organism.
FUNCTIONAL DESCRIPTION: A metabolic pathway is a series of biochemical reactions. These reactions modify metabolites in order to synthesize a new metabolite or to produce energy. One difficulties when dealing with pathways in non-model organism is their incomplete conservation during evolution. To deal with this problem, we developed a prototype inferring new biochemical reactions using reactions and metabolites from known metabolic pathways and metabolomics data. This method produce alternative pathway that could occurred in the species of interest.

RELEASE FUNCTIONAL DESCRIPTION: First version.

NEWS OF THE YEAR: Development of the tool. First release.

- Participants: Arnaud Belcour, Jacques Nicolas, Gabriel Markov and Anne Siegel
- Partner: Station Biologique de Roscoff
- Contact: Anne Siegel
- Publication: Inferring biochemical reactions and metabolite structures to cope with metabolic pathway drift
- URL: https://gitlab.inria.fr/DYLISS/PathModel

6.5. AuReMe

Automatic Reconstruction of Metabolic networks
KEYWORDS: Workflow - Bioinformatics - Metabolic networks - Omic data - Toolbox - Data management
FUNCTIONAL DESCRIPTION: AuReMe enables the reconstruction of metabolic networks from different sources based on sequence annotation, orthology, gap-filling and manual curation. The metabolic network is exported as a local wiki allowing to trace back all the steps and sources of the reconstruction. It is highly relevant for the study of non-model organisms, or the comparison of metabolic networks for different strains or a single organism.

Five modules are composing AuReMe: 1) The Model-management PADmet module allows manipulating and traceing all metabolic data via a local database. 2) The meneco python package allows the gaps of a metabolic network to be filled by using a topological approach that implements a logical programming approach to solve a combinatorial problem 3) The shogen python package allows genome and metabolic network to be aligned in order to identify genome units which contain a large density of genes coding for enzymes, it also implements a logical programming approach. 4) The manual curation assistance PADmet module allows the reported metabolic networks and their metadata to be curated. 5) The Wiki-export PADmet module enables the export of the metabolic network and its functional genomic unit as a local wiki platform allowing a user-friendly investigation.
NEWS OF THE YEAR: (1) Creation of a python package for the multiprocessing of Pathway-Tools. (2) Moving from Python 2 to Python 3. (3) Application of AuReMe to microbiome reconstructions and wikis.

- Participants: Marie Chevallier, Meziane Aite, Guillaume Collet, Nicolas Loira, Sylvain Prigent, Jeanne Got, Anne Siegel and Alejandro Maass
- Partner: University of Chile
- Contact: Meziane Aite
- Publication: Traceability, reproducibility and wiki-exploration for ”à-la-carte” reconstructions of genome-scale metabolic models
- URL: http://aureme.genouest.org/

6.6. biseau

KEYWORDS: ASP - Answer Set Programming - Graph - Formal concept analysis
SCIENTIFIC DESCRIPTION: Use ASP as a Domain Specific Language to specify dot-based visualizations.
NEWS OF THE YEAR: First release.

- Contact: Lucas Bourneuf
- Publication: An Answer Set Programming Environment for High-Level Specification and Visualization of FCA
- URL: https://gitlab.inria.fr/lbourneu/biseau

7. New Results

7.1. Scalable methods to query data heterogenity

Participants: Guillaume Alviset, Olivier Dameron, Xavier Garnier, Vijay Ingalalli, Marine Louarn, Yann Rivault, Anne Siegel, Denis Tagu.

Ontology design and integration [*O. Dameron, Y. Rivault*] We have contribued to several technics improving data integration in onthologies

- The ATOL ontology [link to ontology] supports the annotation of phenotype traits in livestock. It was extended with health-related traits. For each organism, livestock diseases are organized according to their type (infectious, genetic, metabolic,...), their transmission and their symptoms. [32]
- queryMed is an R package [url] that provides both high-level and low-level functions for facilitating the integration of reference ontologies and datasets represented in RDF as Linked Data. It currently focuses on drugs indications, interactions and contra-indications by integrating the Drug Indication Database (DID) and the Drug Interaction Knowledge Base (DIKB). Typical applications concern public health and pharmaco-epidemiology. [27], [26]

Using AskOmics to integrate heterogeneous data [*O. Dameron, A. Siegel*]

- We contributed to the conversion of an Alzheimer's disease map into a heavyweight ontology, the Alzheimer's Disease Map Ontology (ADMO, [url]), an ontological upper model based on systems biology terms. It provides the ontological formalization for the existing disease map AlzPathway that gives a detailed and broad account of Alzheimer's Disease pathophysiology [25], [20].
- We also contributed to decipher the role of small non-coding RNAs in the regulation of animal reproduction, especially the role of miR-202 in female fecundity by regulating medaka oncogenesis [16].

Graph compression and analysis [*L. Bourneuf*]. We introduced a general approach combining procedural and logical languages to specify graph objects. This is a generalization of previous work [37], using the reconstruction of Formal Concept Analysis framework example to target the AI community [23].

7.2. Metabolism: from enzyme sequences to systems ecology

Participants: Meziane Aite, Arnaud Belcour, Marie Chevallier, Mael Conan, François Coste, Olivier Dameron, Clémence Frioux, Jeanne Got, Jacques Nicolas, Anne Siegel, Hugo Talibart.

Efficient identification of substitutable context-free grammars by reduction [*F. Coste, J. Nicolas*] To study more formally the approach by reduction initiated by ReGLiS [40], we introduced a formal characterization of the grammars in reduced normal form (RNF) which can be learned by this approach. Modifying the core of ReGLiS to ensure polynomial running time, we show that local substitutable languages represented by RNF context-free grammars are identifiable in polynomial time and thick data (IPTtD) from positive examples by reduction [19].

Learning grammars capturing 3D structural features of proteins [*F. Coste, H. Talibart*] With the team of Witold Dyrka in Polland, we investigated the problem of learning context-free grammars modeling well protein sequences with respect to their 3D structures.

- A preliminary step is to be able to quantify the relevance of a grammar with respect to a structure. In [21], we introduced and assessed quantitative measures for comparing the topology of the parse tree of a protein sequence analyzed by a context-free grammar with the topology of the protein structure.
- In [24], we established a new framework for learning probabilistic context-free grammars for protein sequences using predicted or experimentally assessed amino acid 3D contacts. We relied on maximum-likelihood and contrastive estimators of parameters in this setting and an implementation for simple yet practical grammars. Tested on samples of protein motifs, grammars developed within the framework showed improved precision in recognition and higher fidelity to protein structures.

Metabolic pathway inference from non genomic data [*A. Belcour, M. Aite, J. Nicolas, A. Siegel, N. Théret, V. Dellannée, M. Conan*] We designed methods for the identification of metabolic pathways for which enzyme information is not precise enough.

- Heterocyclic Aromatic Amines (HAAs) are environmental and food contaminants classified as probable carcinogens. Our approach based on a refinement of molecular predictions with enzyme activity scores allowed us to accurately predict HAAs biotransformation and their potential DNA reactive compounds [13].
- We designed a prototype (*Pathmodel*) implementing inference methods to reconstruct biochemical reactions and metabolite structures to cope with metabolic pathway drift mechanisms. Using known metabolic pathways and metabolomics data, the tool infers alternative pathways compatible with the species known metabolites [29].

Large-scale eukaryotic metabolic network reconstruction [*A. Siegel, M. Chevallier, C. Frioux, M. Aite, J. Cambefort*] Metabolic network reconstruction has attained high standards but is still challenging for complex organisms such as eukaryotes.

- In this direction, we developed AuReMe for a flexible and reproducible reconstruction of these models. Together with a convenient mean for exploration through a local wiki, AuReMe is well-suited for the study of non-model organisms [12].
- In addition, a new gap-filling method satisfying the two main semantics of activation in metabolism is available. It enables to refine the models by pinpointing reactions such that metabolic objectives are met [15].

Systems ecology: design of microbial consortia [*C. Frioux, A. Siegel*]. Finding key elements among hundreds or thousands in microbiota to explain metabolic behaviours or prepare biological experimentations is a highly combinatorial problem. We introduced a two-step approach, MiSCoTo to screen the metabolic capabilities of microbiotas and exhaustively select members of interest by solving optimization problems with logic programming. We applied these methods to data from the Human Microbiome Project and a system composed of the Human metabolic network and 773 models for gut bacteria [14], [11].

7.3. Regulation and signaling: detecting complex and discriminant signatures of phenotypes

Participants: Catherine Belleannée, Samuel Blanquart, Célia Biane-Fourati, Nicolas Guillaudeux, Marine Louarn, Maxime Folschette, François Moreews, Anne Siegel, Nathalie Théret, Pierre Vignet, Méline Wery.

Comparative-genomics based prediction of non-model transcriptomes [*C. Belleannée, S. Blanquart, N. Guillaudeux*] In order to annotate the transcriptome of a non-model species, *Canis lupus familiaris*, we developed a method to predict whether or not a transcript known in a given species/gene could be expressed in an other species/gene. Exploiting knowledge in human, mouse and dog , we predicted a total of 7201 unknown yet transcripts and interpreted the evolutionary dynamics of gene's isoform sets. [30]

Signaling network identification [*M. Folschette, A. Siegel*] [22], [17]

- We introduced a new method to learn an interaction graph from the knowledge of its state space, without assumption on the semantics that was used to produce it. Proofs and characterizations are given for the synchronous, asynchronous and generalized semantics.
- We also used the caspo time-series software to integrate large-scale time series phosphoproteomic data (HPN-DREAM Breast Cancer challenge) into protein signaling networks and infer a family of Boolean Networks. The method highlights commonalities and discrepancies between the four cell lines.

Static analysis of ruled-based models [*P. Vignet, N. Théret*] We used a model of TGF-β to illustrate the main features of Kasa, a static analyzer for Kappa models. Kappa is a rule based language that describes systems of mechanistic interactions between proteins by the means of site-graph rewriting rules. The cornerstone of KaSa is a fix-point engine which detects some patterns that may never occur whatever the evolution of the system may be [18].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. SANOFI: co-supervised phD

Participants: Emmanuelle Becker, Olivier Dameron, Anne Siegel, Méline Wery.

This collaboration project is focused on the implementation of an integrative analysis framework based on semantic web technologies and reasoning in the framework of scleroderma pathology. **CIFRE co-supervised Grant: Ph.D. funding. 2017-2020**

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Ecosyst (Brittany region)

Participants: Marie Chevallier, Clémence Frioux, Anne Siegel.

This project aims at creating a regional research network related to the emerging topic of *systems ecology*, together with the development of tools to be shared by the researchers. It is co-leaded by OSUR (environmental lab) and Dyliss. 2016-2018. Total grant: 100k€

9.1.2. Ph.D. fundings from Université, Inria Rennes and Inserm

The team benefits from the funding of Ph.D. theses by Univ. Rennes (L. Bourneuf, 2016-2019 – H. Talibart, 2017-2020 – N. Guillaudeux, 2018-2021), by Inria (C. Frioux, 2015-2018), by Inserm (M. Louarn, 2017-2019, Inria-Inserm PhD Grant program), and by our collaborators from IRSET (M. Conan, 2017-2020 - P. Vignet, 2018-2020).

9.2. National Initiatives

9.2.1. IDEALG (ANR/PIA-Biotechnology and Bioresource)

Participants: Meziane Aite, Arnaud Belcour, Marie Chevallier, François Coste, Clémence Frioux, Jeanne Got, Jacques Nicolas, Anne Siegel.

The project gathers 18 partners from Station Biologique de Roscoff (coordinator), CNRS, IFREMER, UEB, UBO, UBS, ENSCR, University of Nantes, INRA, AgroCampus, and the industrial field in order to foster biotechnology applications within the seaweed field. Dyliss is co-leader of the WP related to the establishment of a virtual platform for integrating omics studies on seaweed and the integrative analysis of seaweed metabolism. Major objectives are the building of brown algae metabolic maps, metabolic flux analysis and the selection of symbiotic bacteria for brown algae. We will also contribute to the prediction of specific enzymes (sulfatases and haloacid dehalogenase) [More details]. 2012-20. Total grant: 11M€. Dyliss grant: 534k€.

9.2.2. PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases (ANSM)

Participants: Olivier Dameron, Yann Rivault.

The project involves EHESP (coordinator) (public health, Rennes), Univ. Rennes 1 (including the Dyliss), INSERM, CESP and CHU Rennes. The project goal is to develop generic methods supporting efficient and semantically-rich queries for pharmaco-epidemiology studies on medico-administrative databases. 2015-2018. Total grant: 3,6M€. Lacodam & Dyliss grant: 145k€.

9.2.3. TGFSysBio (ITMO Cancer)

Participants: Olivier Dameron, Maxime Folschette, Vijay Ingalalli, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

Partners are INSERM (coordinator) (IRSET, Univ. Rennes 1) CNRS (Dyliss team) and Inria (Antique, Paris). The TGFSYSBIO project aims at developing the first model of extracellular and intracellular TGF-beta system by combining a ruled-based modelling approach (kappa) and a Petri net modelling approach (cadbiom). 2015-18. Total grant: 418k€. Dyliss grant: 129k€.

9.2.4. Programs funded by Inria

9.2.4.1. IPL Algae in silico

Participants: Meziane Aite, Arnaud Belcour, François Coste, Jeanne Got, Anne Siegel.

This project involves mainly the inria teams BIOCORE (coordinator), ANGE and DYLISS. Microalgae are recognized for the extraordinary diversity of molecules they can contain: proteins, lipids (for biofuel or long chain polyunsaturated fatty acids for human health), vitamins, antioxidants, pigments. The project aims at predicting and optimizing the productivity of microalgae. Dyliss is in charge of the identification of physiological functions for microalgae based on their proteomes, which is undergone through the reconstruction of the metabolic network of the *T. lutea* microalgae. Dyliss is also working with the the inria team PLEIADE on learning and predicting the specificities of desaturase enzymes in *Ostreococcus tauri* green algae. 2014-18.

9.2.4.2. IPL Neuromarkers

Participants: Olivier Dameron, Anne Siegel.

This project involves mainly the inria teams Aramis (coordinator) Dyliss, Genscale and Bonsai. The project aims at identifying the main markers of neurodegenerative pathologies through the production and the integration of imaging and bioinformatics data. Dyliss is in charge of facilitating the interoperability of imaging and bioinformatics data. 2017-20.

9.2.4.3. FederatedQueryScaler (Exploratory Research Action)

Participants: Olivier Dameron, Xavier Garnier, Vijay Ingalalli.

This project is coordinated by Dyliss and is a common project with the WIMMICS Inria team. This project aims at developing automatic generation of abstractions for biological data and knowledge in order to scale federated queries in the context of semantic web technologies. 2017-2018.

9.2.4.4. Askomics (ADT)

Participants: Olivier Dameron, Xavier Garnier, Guillaume Alviset, Anne Siegel.

AskOmics [url] is a visual SPARQL query interface supporting both intuitive data integration and querying while avoiding the user to face most of the technical difficulties underlying RDF and SPARQL. The underlying motivation is that even though Linked (Open) Data now provide the infrastructure for accessing large corpora of data and knowledge, life science end-users seldom use them, nor contribute back their data to the LOD cloud by lack of technical expertise. AskOmics aims at bridging the gap between end users and the LOD cloud. 2018-2020.

9.3. European Initiatives

9.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: Polish National Science Center

Project acronym: NCN 2016/21/B/ST6/02158

Project title: Grammatical inference methods in classification of amyloidogenic proteins

Duration: January 2017 - January 2020

Coordinator: Olgierd Unold, Politechnika Wroclawska

Other partners: Politechnika Wroclawska (Polland)

Abstract: The objective is to develop the methods for induction of context-free and probabilistic grammars to describe a language matching amyloidogenic protein sequences.

9.3.2. Collaborations with Major European Organizations

Partner: University of Potsdam, Computer science department (Germany)

Title: Modeling combinatorial and hybrid problems with Answer Set Programming

9.4. International Initiatives

9.4.1. IIL projects

We have a cooperation with Univ. of Chile (MATHomics, A. Maass) on methods for the identification of biomarkers and software for biochip design. It aims at combining automatic reasoning on biological sequences and networks with probabilistic approaches to produce biomarkers of extremophile bacteria. In this context, IntegrativeBioChile was an Associate Team between Dyliss and the Laboratory of Bioinformatics and Mathematics of the Genome hosted at Univ. of Chile funded from 2011 to 2016. The collaboration is now supported by Chilean programs.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- **Niger**. University of Maradi [O. Abdou-Arbi]
- **Poland**. Politechnika Wroclawska [W. Dyrka]

9.5.2. Visits to International Teams

- **Chile**. University of Chile [A. Siegel, C. Frioux, M. Aite, M. Louarn]
- **Poland**. Politechnika Wroclawska [F. Coste]

9.5.2.1. Research Stays Abroad

- **Germany**. University of Potsdam [L. Bourneuf, 3 months (nov 2017 - jan 2018)]

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Workshop on Systems Biology Chile-France, Co-organization (A. Siegel) [More details]

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- SWAT4HCLS (2018) Semantic Web and Tools for Health Care and Life Sciences (O. Dameron)
- BBCC (2018): Bioinformatica e Biologia Computazionale in Campania (O. Dameron)
- JFO (2018): Journées Francophones sur les Ontologies (O. Dameron)
- Journée IA et santé, organized by AFIA and AIM (O. Dameron)
- JOBIM (2018): French conference of Bioinformatics (A. Siegel)
- ISCB (2018) International Conference on Systems Biology (A. Siegel)
- WCB (2018) Workshop on Constraint-based methods in Bioinformatics (A. Siegel)

- BIOINFORMATICS (2018) (A. Siegel)
- ICGI 2016: The 13th International Conference on Grammatical Inference [F. Coste]

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- O. Dameron is an associate editor of the Journal of Biomedical Semantics
- A. Siegel is an academic editor of Genes (MDPI)

10.1.3.2. Reviewer - Reviewing Activities

- Artificial Intelligence in Medicine
- Briefings in Bioinformatics
- Journal of Biomedical Semantics
- Journal on Data Semantics
- Journal of Theoretical Biology
- Natural Computing
- PLOS One

10.1.4. Invited Talks

- Meziane Aite. Workshop France-Chile on Systems Biology. *Tools for metabolic network comparison at genome- scale: application on algaes*, Valparaiso, Chile.
- Maxime Folschette. 4^{th} Annual Meeting of the Bioss Workgroup (Journées annuelles du GT Bioss). *GULA: Semantics-Free Learning of a Biological Regulatory Networks from a Synchronous, Asynchronous or Generalized State Graph*. Marseille.
- Clémence Frioux. Invited speaker of weekly seminar: Bioinformatics group at Laboratoire de Recherche en Informatique (LRI). *Investigating host-microbiota cooperation with gap-filling optimization problems*. Orsay.
- Clémence Frioux. Invited speaker at UMR NUtriomics, Institute of Cardiometabolism and Nutrition (ICAN). *Scalable and exhaustive screening of metabolic functions in microbiotas*. Paris.
- Clémence Frioux. Seminar in the Ecology group led by Pr. Pablo Marquet Universida Catolica de Chile. *Scalable and exhaustive screening of metabolic functions in microbiotas*. Santiago, Chile.
- Clémence Frioux. Workshop France-Chile on Systems Biology. *Scalable and exhaustive screening of metabolic functions in microbiotas*. Valparaiso, Chile.
- Clémence Frioux. Journées Scientifiques de Nantes. *Scalable and exhaustive screening of metabolic functions in carried out by microbial consortia*. Nantes.
- Marine Louarn. Workshop France-Chile on Systems Biology. *Integration of omics data for the specification of regulation networks in health; identification of regulatory mechanisms*. Valparaiso, Chile.
- Anne Siegel. National meeting of the French Society of Microbiology. *Several aspects of artificial intelligence : which perspectives for microbiology?*. Paris.
- Anne Siegel. Century Seminar. *Reasoning over biological systems with large-scale and incomplete data*. Marseille.
- Anne Siegel. Journées annuelles du GT Bioss. *Learning boolean rules for the regulatory control of metabolism : a case study*. Marseille.
- Anne Siegel. Séminaire INRA-CATI. *Metabolic network reconstruction for non-model organisms*. INRA.
- Anne Siegel. ECCB worshop Logical modeling of cellular networks. *Learning boolean rules for the regulatory control of metabolism : a case study*. Athens, Greece.

- Anne Siegel. Workshop France-Chile on Systems Biology. *Integrating and querying data with AskOmics*. Valparaiso, Chile.
- Nathalie Théret. *Causality analysis of TGF-β-dependent signaling pathways*. Workshop "Computational System Biology for Cancer", Paris.
- Nathalie Théret. *Computational modeling to identify biomarkers and targets* Workshop DHU2020, Nantes, France

10.1.5. Leadership within the Scientific Community

- Member of the steering committee of the International Conference on Grammatical Inference [F. Coste]
- The team was involved in the foundation of a national working group on the symbolic study of dynamical systems named bioss [web access]. The group gathers more than 170 scientists, from computer science to biology. Three meetings were organized this year (Rennes, Marseille, Paris). The group is supported by two French National Research Networks: bioinformatics (GDR BIM : bioinformatique moléculaire) and informatics-mathematics (GDR IM : Informatique Mathématique).

10.1.6. Scientific Expertise

10.1.6.1. International expertise

- Luxembourg, FNR. Core program [A. Siegel]
- Senegal, CEA-MITIC (Centre d'Excellence Africain en Mathématiques, Informatique et TIC [O. Dameron, reviewed 1 project]

10.1.6.2. National responsibilities

- **Scientific program committee** National Research Agency, CES 45 [A. Siegel]
- **Institutional boards for the recruitment and evaluation of researchers**.
  - National Council of Universities, section 65 [O. Dameron, nominated member].
- **Recruitment committees**.
  - Inria Senior Researchers [national committee, A. Siegel]
  - Inria Junior Researchers [Saclay, A. Siegel]
  - Assistant Professor [Univ. Marseille, A. Siegel]

10.1.6.3. National scientific boards

- Animation of the Systems Biology working group of national infrastructure GDR IM and GDR BIM (Bioss, details above) [A. Siegel].
- Board of directors of the French Society for biology of the extracellular matrix [N. Théret].

10.1.6.4. Local responsibilities

- Scientific Advisory Board of Biogenouest [N. Théret]
- IRISA laboratory (computer science department of Univ. Rennes 1) council [A. Siegel]
- Responsability of the IRISA laboratory "Health-biology" cross-cutting axis [O. Dameron] [More details]
- Elected member of the social action services (ASUR) of Univ. Rennes 1 [C. Belleannée]
- Scientific committee of Univ. Rennes 1 school of medicine [O. Dameron, A. Siegel].

10.1.7. Research Administration

10.1.7.1. Inria Instances

- Inria National evaluation board [A. Siegel, nominated member].
- Equality and diversity Committee, Inria - Responsible of the working group focusing on recruitment procedures [A. Siegel].

10.1.7.2. Prospective working groups

- "Big & Open Data en recherche à l'horizon 2040" foresight working group of PROSPER network [F. Coste].

10.1.7.3. Inria local instances

- Gender equality commission, IRISA & Inria Rennes [A. Siegel, coordinator]
- Inria-Rennes Computer science infrastructure users commission - CUMI-R commission [F. Coste]
- Inria Rennes PHD recruitment (CORDIs) [C. Belleannée]

10.2. Teaching - Supervision - Juries

10.2.1. Teaching track responsibilities

- Coordination of the doctoral school "Biology and Health" of University of Bretagne Loire, Rennes Site 1 [N. Théret]
- Coordination of the master degree "Bioinformatics and genomics", Univ. Rennes1 [O. Dameron]

10.2.2. Course responsibilities

- "Internship", Master 1 in Computer Sciences, Univ. Rennes 1 [C. Belleannée]
- "Supervised machine learning", Master 2 in Computer Sciences, Univ Rennes 1 [F. Coste]
- "Atelier bioinformatique", Licence 2 informatique, Univ. Rennes 1 [O. Dameron]
- "Bioinformatique pour la génomique", 2nd year school of medicine, Univ. Rennes 1 [O. Dameron]
- "Bases de mathématiques et probablité" and "Méthodes en informatique", Master1 in public health, Univ. Rennes 1 [O. Dameron]
- "Intégration: Remise à niveau en informatique", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Programmation en Python", Master 1 in Public Health, Univ. Rennes 1 [O. Dameron]
- "Programmation impérative en Python", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Système informatique GNU/Linux", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Semantic Web and bio-ontologies", Master 2 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Internship", Master 2 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- Master: A. Siegel, Integrative and Systems biology, Master 2 in bioinformatics, Univ. Rennes 1 [A. Siegel]
- Micro-environnement Cellulaire normal & pathologique, Master 2 in Biologie cellulaire et Moléculaire, Univ. Rennes 1 [N. Théret]

10.2.3. Teaching

Licence: C. Belleannée, Langages formels, 20h, L3 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Projet professionnel et communication, 16h, L1 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Enseignant référent, 20h, L1 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Spécialité informatique : Functional and immutable programming , 42h, L1 informatique, Univ. Rennes1, France

Licence: O. Dameron, Biostatistiques, 12h, 1st year school of medicine, Univ. Rennes 1, France

Licence: O. Dameron, Bioinformatique pour la génomique, 2h, 2nd year school of medicine, Univ. Rennes 1, France

Licence: O. Dameron, "Atelier bioinformatique", 8h, Licence 2 informatique, Univ. Rennes 1, France

Master: O. Dameron, "Intégration: Remise à niveau en informatique", 14h Master 1 in bioinformatics, Univ. Rennes 1, France

Licence: C. Frioux, Projet Python, 18h, L3 ENSAI, France.

Licence: C. Frioux, Techniques d'optimsiation, 12h, L3 ENSAI, France.

Licence: M. Louarn, Introduction à la BioInformatique, 2h, L2 Informatique, Univ. Rennes 1, France.

Licence: M. Louarn, TPs Python, 8h, L1 Biologie, Univ. Rennes 1, France.

Licence: H. Talibart, Programmation scientifique Python, 24h, L1, Rennes1, France.

Licence: M. Wery, TPs Python, 36h, L1 Biologie, Univ. Rennes 1, France

Master: O. Dameron, "Programmation impérative en Python", 82.25h Master 1 in bioinformatics, Univ. Rennes 1, France

Master: C. Belleannée, Programmation logique avec contraintes et algorithmes génétiques, 40h, M1 informatique, Univ. Rennes1, France.

Master: C. Belleannée, Algorithmique du texte et bioinformatique, 10h, M1 informatique, Univ. Rennes1, France

Master: L. Bourneuf, Projet, 20h, M1 Santé Publique, France.

Master: F. Coste, Supervised machine learning, 15h, M2 Computer Sciences, Univ Rennes 1, France

Master: O. Dameron, "Object-oriented programming", 20h Master 1 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, "Système informatique GNU/Linux", 20h, Master 1 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, 8h, "Internship", Master 1 in bioinformatics, Univ. Rennes 1

Master: O. Dameron, "Bases de mathématiques et probablité", 18.75h, Master1 in public health, Univ. Rennes 1

Master: O. Dameron, "Programmation impérative en Python", 3h Master 1 in public health, Univ. Rennes 1, France

Master: O. Dameron, 21h, "Semantic Web and bio-ontologies", Master 2 in bioinformatics, Univ. Rennes 1

Master: O. Dameron, 10h, "Internship", Master 2 in bioinformatics, Univ. Rennes 1

Master: C. Belleannée, 20h, "Internship", Master 1 Computer Sciences, Univ. Rennes 1, France

Master: M. Louarn, Informatique Médicale Avancée, 2h, M1 Médecine, Univ. Rennes 1, France.

Master: M. Louarn, Object-oriented programming, 25h, M2 bioinformatique et génomique, Univ. Rennes 1, France.

Master: A. Siegel, Integrative and Systems biology, 25h, M2, Univ. Rennes 1, France.

Doctorat: A. Siegel, Reasoning over biological systems with large-scale and incomplete data, 3h, Thematic Research School on "advances in Systems & Synthetic Biology", Evry'18.

10.2.4. Supervision

PhD : Clémence Frioux, *Investigating host-microbiota cooperation with gap-filling optimization problems.*, started in Oct. 2015, defended on Nov. 19 2018, supervised by A. Siegel [11].

PhD in progress : Yann Rivault, *Analyse de trajectoires de soins à partir de bases de données médico-administratives : apport d'un enrichissement pas des connaissances biomédicales issues du Web des Données*, started in Oct. 2015, defense expected 28th January 2019, supervised by O. Dameron and N. Lemeur.

PhD in progress : Lucas Bourneuf, *Justifiable graph decomposition to assist biological network understanding*, started in Oct. 2016, supervised by J. Nicolas.

PhD in progress : Hugo Talibart, *Learning grammars with long-distance correlations on proteins*, started in Nov. 2017, supervised by F. Coste and J. Nicolas.

PhD in progress : Mael Conan, *Predictive approach to assess the genotoxicity of environmental contaminants during liver fibrosis*, started in Oct. 2017, supervised by S. Langouet and A. Siegel.

PhD in progress: Marine Louarn, *Intégration de données génomiques massives et hétérogènes, application aux mutations non-codantes dans le lymphome folliculaire*, started in Oct. 2017, supervised by A. Siegel, T. Fest (CHU) and O. Dameron.

PhD in progress : Méline Wery, *Methodology development in disease treatment projects.* , started in Oct. 2017, supervised by O. Dameron, C. Bettembourg (Sanofi) and A. Siegel.

PhD in progress : Nicolas Guillaudeux, *Comparer des structures de gènes pour la prédiction de transcrits isoformes codants et non-codants*, started in Oct. 2018, supervised by O. Dameron, C. Belleannée and S. Blanquart.

PhD in progress : Pierre Vignet, *Identification et conception expérimentale de nouveaux agents thérapeutiques à partir d'un modèle informatique des réseaux d'influence du TGF-β dans les pathologies hépatiques chroniques* , started in Dec. 2018, supervised by N. Théret and A. Siegel.

10.2.5. Juries

- *Referee of Ph-D thesis*. C. Biane, Univ Paris Saclay [A. Siegel]. A. Bonnafoux, ENS Lyon [A. Siegel]. B. Miraglio, Univ Nice [A. Siegel]. G Personeni, Univ. Lorraine [O. Dameron].
- *Member of Ph-D thesis juries*. C. Marchet, Univ Rennes [A. Siegel]. M. Pichene, Univ. Rennes [A. Siegel].
- *Refererr of habilitation thesis juries*. A. Baudot, Univ. Marseille.
- *Member of medicine doctorate juries* A. El Matoutat [O. Dameron].

10.2.6. Interns

- Internship, from Jan 2018 until Jun 2018. Supervised by I. Vigalatti and O. Dameron. Student: Guillaume Alviset. Subject: Selecting SPARQL endpoints for faster processing of federated queries.
- Internship, from Jan 2018 until Jun 2018. Supervised by A. Siegel. Student: Aurelien Cornet. Subject: Inférence de réseaux biologiques multi-échelles.
- Internship, from Apr 2018 until Jul 2018. Supervised by A. Siegel, F. Legeai and J. Bobe. Student: Clara Delahaye. Subject: Intégration des réseaux géniques de Oryzias latipes impliqués dans la fécondité et régulés par le microARN-202-5p.
- Internship, from Apr 2018 until Jul 2018. Supervised by A. Siegel and R. Munoz. Student: Emile Dumont. Subject: Reconstruction du réseau métabolique de bactéries cellulolytiques du rumen : exploration des voies de dégradation de la cellulose et de l'hémicellulose.
- Internship, from Jan 2018 until Jun 2018. Supervised by C. Frioux. Student: Enora Fremy. Subject: Etude du potentiel de coopération métabolique entre hôte et microbiote à l'échelle fonctionnelle.
- Internship, from Jan 2018 until Jun 2018. Supervised by C. Belleannée and S. Blanquart. Student: Nicolas Guillaudeux. Subject: Prédiction de transcriptome : analyse comparative multi-gènes, orthologues.
- Internship, from Apr 2018 until Jul 2018. Supervised by J. Got. Student: Teo Lemane. Subject: Curation du réseau métabolique de T. Lutea via des méthodes hétérogènes.
- Internship, from Jan 2018 until Jul 2018. Supervised by N. Théret, J. Nicolas and L. Bourneuf. Student: Linh Chi Nguyen. Subject: Compression of large interaction graphs for the identification of protein modules in the extracellular matrix.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Science en Cour[t]s (*http://sciences-en-courts.fr/*) Many of our on-going and former Ph-D students (M.Wery, L. Bourneuf, A. Antoine-Lorquin, C. Bettembourg, J. Coquet, V. Delannée, G. Garet, S. Prigent) have been heavily involved in the organization of a local Popularization Festival where Ph.D. students explain their thesis via short movies. The movies are presented to a professional jury composed of artists and scientists, and of high-school students. Previous years films can be viewed on the festival web-site [More details]

10.3.2. Interventions

- "J'peux pas j'ai informatique", Welcoming of high-school students to leverage stereotypes about computer sciences, Apr. 2018 [A. Siegel]
- Atelier Google Numerics, Rennes, table ronde *Enseigner le code dès le plus jeune âge*, Oct. 2018 [A. Siegel]
- "Rencontre des jeunes mathématiciennes", ENS Rennes, 2-days workshop for high-school girl students in mathematics. Introduction to gender stereotypes in mathematics, Nov. 2018 [A. Siegel]

10.3.3. Creation of media or tools for science outreach

- Le Language comme inspiration [youtube link] [H. Talibart]
- Les aventuriers du remède perdu [youtube link] [M. Louarn and M. Wery]

11. Bibliography

Major publications by the team in recent years

[1] C. BELLEANNÉE, O. SALLOU, J. NICOLAS. *Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling*, in "PRIB2014 - Pattern Recognition in Bioinformatics, 9th IAPR International Conference", Stockholm, Sweden, M. COMIN, L. KALL, E. MARCHIORI, A. NGOM, J. RAJAPAKSE (editors), Springer International Publishing, August 2014, vol. 8626, pp. 34-47 [*DOI :* 10.1007/978-3-319-09192-1_4], https://hal.inria.fr/hal-01059506

[2] C. BETTEMBOURG, C. DIOT, 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, 30 p. [*DOI :* 10.1371/JOURNAL.PONE.0133579], https://hal.inria.fr/hal-01184934

[3] P. BORDRON, M. LATORRE, M.-P. CORTÉS, M. GONZALES, S. THIELE, A. SIEGEL, A. MAASS, D. EVEILLARD. *Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach*, in "MicrobiologyOpen", 2015, vol. 5, n^{o} 1, pp. 106-117 [*DOI :* 10.1002/MBO3.315], https://hal.inria.fr/hal-01246173

[4] J. COQUET, N. THÉRET, V. LEGAGNEUX, 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", Darmstadt, Lecture Notes in Computer Sciences, September 2017, 17 p. , https://hal.archives-ouvertes.fr/hal-01559249

[5] F. COSTE, G. GARET, A. GROISILLIER, J. NICOLAS, T. TONON. *Automated Enzyme classification by Formal Concept Analysis*, in "ICFCA - 12th International Conference on Formal Concept Analysis", Cluj-Napoca, Romania, Springer, June 2014, https://hal.inria.fr/hal-01063727

[6] F. COSTE, 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", Wrocław, Poland, September 2018, vol. 93, pp. 155 - 168, https://hal.inria.fr/hal-01872266

[7] C. FRIOUX, E. FREMY, C. TROTTIER, A. SIEGEL. *Scalable and exhaustive screening of metabolic functions carried out by microbial consortia*, in "Bioinformatics", September 2018, vol. 34, n° 17, pp. i934 - i943 [*DOI :* 10.1093/BIOINFORMATICS/BTY588], https://hal.inria.fr/hal-01871600

[8] C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO. *Hybrid Metitebolic Network Completion*, in "Theory and Practice of Logic Programming", November 2018, pp. 1-23, https://hal.inria.fr/hal-01936778

[9] 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, A. SIEGEL. *Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks*, in "PLoS Computational Biology", January 2017, vol. 13, n° 1, 32 p. [*DOI :* 10.1371/JOURNAL.PCBI.1005276], https://hal.inria.fr/hal-01449100

[10] S. VIDELA, J. SAEZ-RODRIGUEZ, C. GUZIOLOWSKI, A. SIEGEL. *caspo: a toolbox for automated reasoning on the response of logical signaling networks families*, in "Bioinformatics", 2017 [*DOI :* 10.1093/BIOINFORMATICS/BTW738], https://hal.inria.fr/hal-01426880

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] C. FRIOUX. *Investigating host-microbiota cooperation with gap-filling optimization problems*, Université de Rennes 1, November 2018, https://hal.inria.fr/tel-01945853

Articles in International Peer-Reviewed Journals

[12] 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, A. SIEGEL. *Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models*, in "PLoS Computational Biology", 2018, vol. 14, n° 5, e1006146 p. [*DOI :* 10.1371/JOURNAL.PCBI.1006146], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01807842

[13] V. DELANNÉE, S. LANGOUËT, A. SIEGEL, N. THÉRET. *In silico prediction of Heterocyclic Aromatic Amines metabolism susceptible to form DNA adducts in humans*, in "Toxicology Letters", January 2019, vol. 300, pp. 18-30 [*DOI :* 10.1016/J.TOXLET.2018.10.011], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01903264

[14] C. FRIOUX, E. FREMY, C. TROTTIER, A. SIEGEL. *Scalable and exhaustive screening of metabolic functions carried out by microbial consortia*, in "Bioinformatics", September 2018, vol. 34, n° 17, pp. i934 - i943 [*DOI :* 10.1093/BIOINFORMATICS/BTY588], https://hal.inria.fr/hal-01871600

[15] C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO. *Hybrid Metabolic Network Completion*, in "Theory and Practice of Logic Programming", November 2018, pp. 1-23 [*DOI :* 10.1017/S1471068418000455], https://hal.inria.fr/hal-01936778

[16] S. GAY, J. BUGEON, A. BOUCHAREB, L. HENRY, C. DELAHAYE, F. LEGEAI, J. MONTFORT, A. LE CAM, A. SIEGEL, J. BOBE, V. THERMES. *MiR-202 controls female fecundity by regulating medaka oogene-*

sis, in "PLoS Genetics", September 2018, vol. 14, n^{o} 9, pp. 1-26 [*DOI :* 10.1371/JOURNAL.PGEN.1007593], https://hal.archives-ouvertes.fr/hal-01871468

[17] M. RAZZAQ, L. PAULEVÉ, A. SIEGEL, J. SAEZ-RODRIGUEZ, J. BOURDON, C. GUZIOLOWSKI. *Computational Discovery of Dynamic Cell Line Specific Boolean Networks from Multiplex Time-Course Data*, in "PLoS Computational Biology", 2018, vol. 14, pp. 1-23 [*DOI :* 10.1371/JOURNAL.PCBI.1006538], https://hal.archives-ouvertes.fr/hal-01897020

International Conferences with Proceedings

[18] P. BOUTILLIER, F. CAMPORESI, J. COQUET, J. FERET, K. Q. LÝ, N. THÉRET, P. VIGNET. *KaSa A Static Analyzer for Kappa*, in "CMSB 2018 - 16th International Conference on Computational Methods in Systems Biology", Brno, Czech Republic, M. ČEŠKA, D. ŠAFRÁNEK (editors), LNCS, Springer Verlag, September 2018, vol. 11095, pp. 285-291 [*DOI :* 10.1007/978-3-319-99429-1_17], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01888951

[19] F. COSTE, 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", Wrocław, Poland, September 2018, vol. 93, pp. 155 - 168, https://hal.inria.fr/hal-01872266

[20] V. J. HENRY, I. MOSZER, O. DAMERON, M.-C. POTIER, M. HOFMANN-APITIUS, O. COLLIOT. *Converting Alzheimer's disease map into a heavyweight ontology: a formal network to integrate data*, in "DILS 2018 - 13th International Conference on Data Integration in the Life Sciences", Hannover, Germany, DILS 2018 - 13th International Conference on Data Integration in the Life Sciences, November 2018, pp. 1-9, https://hal.archives-ouvertes.fr/hal-01917742

[21] M. PYZIK, F. COSTE, W. DYRKA. *How to measure the topological quality of protein parse trees?*, in "ICGI 2018 - 14th International Conference on Grammatical Inference", Wroclaw, Poland, September 2018, vol. 93, pp. 118 - 138, https://hal.inria.fr/hal-01938608

[22] T. RIBEIRO, M. FOLSCHETTE, M. MAGNIN, O. ROUX, K. INOUE. *Learning Dynamics with Synchronous, Asynchronous and General Semantics*, in "28th International Conference on Inductive Logic Programming", Ferrara, Italy, Inductive Logic Programming, Springer, Cham, September 2018, vol. LNCS, volume 11105 [*DOI :* 10.1007/978-3-319-99960-9_8], https://hal.archives-ouvertes.fr/hal-01826564

Conferences without Proceedings

[23] L. BOURNEUF. *An Answer Set Programming Environment for High-Level Specification and Visualization of FCA*, in "FCA4AI'18 - 6th Workshop «What can FCA do for Artificial Intelligence?»", Stockholm, Sweden, July 2018, pp. 1-12, https://hal.archives-ouvertes.fr/hal-01945938

[24] W. DYRKA, F. COSTE, H. TALIBART. *Where the context-free grammar meets the contact map: a probabilistic model of protein sequences aware of contacts between amino acids*, in "ISMB 2018 - 3DSIG: Structural Bioinformatics and Computational Biophysics", Chicago, United States, July 2018, https://hal.inria.fr/hal-01939021

[25] V. HENRY, I. MOSZER, O. DAMERON, M.-C. POTIER, M. HOFMANN-APITIUS, O. COLLIOT. *Integrating ontological representation and reasoning into a disease map: application to Alzheimer's disease*, in "DMCM 2018 - 3rd Disease Maps Community Meeting", Paris, France, June 2018, pp. 1-2, https://hal.archives-ouvertes.fr/hal-01873474

[26] Y. RIVAULT, O. DAMERON, N. LE MEUR. *Ontologies biomédicales et Web Sémantique pour la réutilisation des bases de données médico-administratives en pharmaco-épidémiologie*, in "JFO 2018 - 7ème Journées Francophones sur les Ontologies", Hammamet, Tunisia, November 2018, pp. 1-6, https://hal.archives-ouvertes.fr/hal-01912046

[27] Y. RIVAULT, O. DAMERON, N. LE MEUR. *queryMed : enrichissement des données de recherche en pharmaco-épidémiologie*, in "Septièmes rencontres R", Rennes, France, July 2018, pp. 1-2, https://hal.archives-ouvertes.fr/hal-01834137

Scientific Popularization

[28] C. FRIOUX, E. FREMY, C. TROTTIER, A. SIEGEL. *Scalable and exhaustive screening of metabolic functions carried out by microbial consortia*, in "European Conference on Computational Biology", Athènes, Greece, September 2018, https://hal.inria.fr/hal-01946860

Other Publications

[29] 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, G. V. MARKOV. *Inferring biochemical reactions and metabolite structures to cope with metabolic pathway drift*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01943880

[30] N. GUILLAUDEUX. *Prédiction de transcriptome : analyse comparative multi-gènes, orthologues*, Univ Rennes ; Inria ; CNRS ; Irisa, June 2018, 42 p. , https://hal.inria.fr/hal-01948461

[31] C. MARCHET, L. LECOMPTE, C. DA SILVA, C. CRUAUD, J.-M. AURY, J. NICOLAS, P. PETERLONGO. *CARNAC-LR: De novo Clustering of Gene Expressed Variants in Transcriptomic Long Reads Data Sets*, April 2018, pp. 1-2, RECOMB-seq 2018 - Eighth RECOMB Satellite Workshop on Massively Parallel Sequencing, Poster, https://hal.archives-ouvertes.fr/hal-01929963

[32] J. YON, N. BAREILLE, B. BED'HOM, X. BERTHELOT, S. COMBES, O. DAMERON, G. FOUCRAS, A. JOURNAUX, F. LAUNAY, P.-Y. LE BAIL, N. LE FLOC'H, M. MARTIGNON, C. NÉDELLEC, M.-H. PINARD - VAN DER LAAN, E. QUILLET, M.-C. MEUNIER-SALAÜN. *Ontologie ATOL : amélioration de l'outil par l'intégration des caractères de santé*, April 2018, Journées d'Animation Scientifique du département Phase (JAS Phase 2018), Poster, https://hal.archives-ouvertes.fr/hal-01768063

References in notes

[33] G. ANDRIEUX, M. LE BORGNE, N. THÉRET. *An integrative modeling framework reveals plasticity of TGF-Beta signaling*, in "BMC Systems Biology", 2014, vol. 8, n° 1, 30 p. [*DOI :* 10.1186/1752-0509-8-30], http://www.hal.inserm.fr/inserm-00978313

[34] T. BERNERS LEE, W. HALL, J. A. HENDLER, K. O'HARA, N. SHADBOLT, D. J. WEITZNER. *A Framework for Web Science*, in "Foundations and Trends in Web Science", 2007, vol. 1, n° 1, pp. 1–130

[35] S. BLANQUART, J.-S. VARRÉ, P. GUERTIN, A. PERRIN, A. BERGERON, K. M. SWENSON. *Assisted transcriptome reconstruction and splicing orthology*, in "BMC Genomics", Nov 2016, vol. 17, n° 10, 786 p. , https://doi.org/10.1186/s12864-016-3103-6

[36] P. BLAVY, F. GONDRET, S. LAGARRIGUE, J. VAN MILGEN, 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", 2014, vol. 8, n° 1, 32 p. [*DOI :* 10.1186/1752-0509-8-32], https://hal.inria.fr/hal-00980499

[37] L. BOURNEUF, J. NICOLAS. *FCA in a Logical Programming Setting for Visualization-oriented Graph Compression*, in "International Conference on Formal Concept Analysis 2017", Rennes, France, Lecture Notes in Computer Sciences, June 2017, vol. 10308 [*DOI :* 10.1007/978-3-319-59271-8_6], https://hal.archives-ouvertes.fr/hal-01558302

[38] A. BRETAUDEAU, F. COSTE, F. HUMILY, L. GARCZAREK, G. LE CORGUILLÉ, C. SIX, M. RATIN, O. COLLIN, W. M. SCHLUCHTER, F. PARTENSKY. *CyanoLyase: a database of phycobilin lyase sequences, motifs and functions*, in "Nucleic Acids Research", November 2012, 6 p. [*DOI :* 10.1093/NAR/GKS1091], https://hal.inria.fr/hal-01094087

[39] M. P. CORTÉS, S. N. MENDOZA, D. TRAVISANY, A. GAETE, A. SIEGEL, V. CAMBIAZO, A. MAASS. *Analysis of Piscirickettsia salmonis Metabolism Using Genome-Scale Reconstruction, Modeling, and Testing*, in "Frontiers in Microbiology", December 2017, vol. 8, 15 p. [*DOI :* 10.3389/FMICB.2017.02462], https://hal.inria.fr/hal-01661270

[40] F. COSTE, G. GARET, J. NICOLAS. *A bottom-up efficient algorithm learning substitutable languages from positive examples*, in "ICGI (International Conference on Grammatical Inference)", Kyoto, Japan, A. CLARK, M. KANAZAWA, R. YOSHINAKA (editors), Proceedings of Machine Learning Research, September 2014, vol. 34, pp. 49–63, https://hal.inria.fr/hal-01080249

[41] K. FAUST, J. RAES. *Microbial interactions: from networks to models*, in "Nat. Rev. Microbiol.", Jul 2012, vol. 10, n° 8, pp. 538–550

[42] M. Y. GALPERIN, D. J. RIGDEN, X. M. FERNÁNDEZ-SUÁREZ. *The 2015 nucleic acids research database issue and molecular biology database collection*, in "Nucleic acids research", 2015, vol. 43, n° D1, pp. D1–D5

[43] M. GEBSER, R. KAMINSKI, B. KAUFMANN, T. SCHAUB. *Answer Set Solving in Practice*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers, 2012

[44] 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, T. TONON. *The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond*, in "Plant Journal", September 2014, pp. 367-81 [*DOI :* 10.1111/TPJ.12627], https://hal.archives-ouvertes.fr/hal-01057153

[45] M. H. SAIER, V. S. REDDY, B. V. TSU, M. S. AHMED, C. LI, G. MORENO-HAGELSIEB. *The Transporter Classification Database (TCDB): recent advances*, in "Nucleic Acids Res.", Jan 2016, vol. 44, n° D1, pp. D372–379

[46] D. B. SEARLS. *String variable grammar: A logic grammar formalism for the biological language of DNA*, in "The Journal of Logic Programming", 1995, vol. 24, n° 1, pp. 73 - 102, Computational Linguistics and Logic Programming [*DOI :* 10.1016/0743-1066(95)00034-H], http://www.sciencedirect.com/science/article/pii/074310669500034H

[47] Z. D. STEPHENS, S. Y. LEE, F. FAGHRI, R. H. CAMPBELL, C. ZHAI, M. J. EFRON, R. IYER, M. C. SCHATZ, S. SINHA, G. E. ROBINSON. *Big Data: Astronomical or Genomical?*, in "PLoS biology", 2015, vol. 13, n^{o} 7, e1002195 p.

[48] R. VERBORGH, M. VANDER SANDE, O. HARTIG, J. VAN HERWEGEN, L. DE VOCHT, B. DE MEESTER, G. HAESENDONCK, P. COLPAERT. *Triple Pattern Fragments: a Low-cost Knowledge Graph Interface for the Web*, in "Journal of Web Semantics", March 2016, vol. 37–38, pp. 184–206 [*DOI :* DOI:10.1016/J.WEBSEM.2016.03.003], http://linkeddatafragments.org/publications/jws2016.pdf