Activity Report 2017

Team DYLISS

Dynamics, Logics and Inference for Biological Systems and Sequences

Joint team with Inria Rennes – Bretagne Atlantique

D7 – Data and Knowledge Management

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
A3.2.3. - Inference
A3.2.4. - Semantic Web
A3.2.5. - Ontologies
A3.3. - Data and knowledge 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

Other Research Topics and Application Domains:
B1.1.2. - Molecular biology
B1.1.3. - Cellular biology
B1.1.9. - Bioinformatics
B1.1.11. - Systems biology
B1.1.14. - Microbiology
B2.2.3. - Cancer
B2.2.5. - Immune system diseases

1. Personnel

Research Scientists
Anne Siegel [Team leader, CNRS, Senior Researcher, HDR]
François Coste [Inria, Researcher]
Jacques Nicolas [Inria, 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
Maxime Folschette [Univ de Rennes I, from Nov 2017]
Vijay Ingalalli [CNRS, from Jun 2017]

PhD Students
Lucas Bourneuf [Univ de Rennes I]
Mael Conan [Univ de Rennes I, from Oct 2017]
Jean Coquet [Univ de Rennes I]

Victorien Delannée [Univ de Rennes I, until Nov 2017]
Clémence Frioux [Inria]
Julie Laniau [Inria, until Jul 2017 and CNRS from Sep 2017 until Nov 2017]
Marine Louarn [Univ de Rennes I, from Oct 2017]
Yann Rivault [Univ de Rennes I]
Juliette Talibart [Univ de Rennes I, from Nov 2017]
Méline Wery [Univ de Rennes I, from Oct 2017]

Technical staff
Meziane Aite [Inria]
Xavier Garnier [INRA]
Camille Trottier [CNRS, until Sep 2017]
Pierre Vignet [CNRS]
Jeanne Got [CNRS]
Arnaud Belcour [CNRS, from Dec 2017]

Administrative Assistant
Marie Le Roic [Univ de Rennes I]

Visiting Scientist
Oumarou Abdou Arbi [Université de Maradi, from Sep 2017 to Dec 2017]

External Collaborators
Jérémie Bourdon [Univ de Nantes, HDR]
Marie Chevallier [CNRS]
Damien Eveillard [Univ de Nantes]
François Moreews [INRA]
Denis Tagu [INRA, HDR]
Nathalie Théret [INSERM, HDR]

2. Overall Objectives

2.1. Overall objectives

The research domain of the bioinformatics Dyliss team is sequence analysis and systems biology. Our main goal in biology is to characterize groups of genetic actors that control the phenotypic answer of species when challenged by their environment. The team explores methods in the field of formal systems, more precisely in knowledge representation, constraints programming, multi-scale analysis of dynamical systems, and machine learning. Our goal is to identify key regulators of the environmental response by structuring and reasoning on information which combines physiological responses measured with omics technologies (RNA-seq, metabolomics, proteomics), genetic information from their long-distant cousins and knowledge about regulation and metabolic pathways stored in public repositories.

The main challenges we face are data incompleteness and heterogeneity. We favor the construction and study of a "space of feasible models or hypotheses" including known constraints and facts on a living system *rather than searching for a single optimized model*. We develop methods allowing a precise investigation of this space of hypotheses. Therefore, we are in position of developing experimental strategies to progressively shrink the space of hypotheses and gain in the understanding of the system. Importantly, one should notice that our models span a quite large spectrum of discrete structures: oriented graphs, boolean networks, automata, or expressive grammars.

More concretely, the steps of the analysis are to (*i*) 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 perturbation, sequencing...); (*ii*) investigate the space of admissible models and exhibit its main features by solving combinatorial optimization problems; (*iii*) identify corresponding genomic products within sequences. At each of these steps, we rely on symbolic methods for model space exploration (ontologies and formal concepts analysis).

We target applications for which large-scale heterogeneous data about a specific but complex physiological phenotype are available. Existing long-term partnerships with biological labs give strong support to this choice. In marine biology, we collaborate closely with the Station biologique de Roscoff (*Idealg*, Investissement avenir "Bioressources et Biotechnologies"). We also collaborate with other teams of Inria in the *IPL Algae In Silico project to understand the metabolism of a micro-algae.* In environmental microbiology we collaborate both with the CRG in Chile in the framework of the Ciric Chilean Inria center (*Ciric-Omics*). In agriculture, our main partners are within the INRA institute in Rennes, with a focus on the understanding of pea-aphids reproduction mode and of breeding animals metabolism (porc, chicken, cow). More recently, we have introduced health as a new application field of the team, especially through the study of large-scale boolean networks and their confrontation with knowledge repositories (collaboration with Inserm, CHU Rennes and Sanofi).

3. Research Program

3.1. Modeling knowledge integration with combinatorial constraints

Biological networks are built with data-driven approaches aiming at translating genomic information into a functional map. Most methods are based on a probabilistic framework which defines a probability distribution over the set of models. The reconstructed network is then defined as the most likely model given the data.

Our team has investigated an alternative perspective where each data induces a set of constraints – related to the steady state response of the system dynamics – on the set of possible values in a network of fixed topology. The methods that we have developed complete the network with product states at the level of nodes and influence types at the level of edges, able to globally explain experimental data. In other words, the selection of relevant information in the model is no more performed by selecting *the* network with the highest score, but rather by exploring the complete space of models satisfying constraints on the possible dynamics supported by prior knowledge and observations. In the (common) case when there is no model satisfying all the constraints, we relax the problem by introducing new combinatorial optimization problems that introduce the possibility of correcting the data or the knowledge. Common properties to all solutions are considered as a robust information about the system, as they are independent from the choice of a single solution to the optimization problem [77].

Solving these computational issues requires addressing NP-hard qualitative (non-temporal) issues. We have developed a long-term collaboration with Potsdam University in order to use a logical paradigm named **Answer Set Programming** (ASP) [46], [73] to solve these constraint satisfiability and combinatorial optimization issues. Applied on transcriptomic or cancer networks, our methods identified which regions of a large-scale network shall be corrected [47], and proposed robust corrections [6]. This result suggested that this approach was compatible with efficiency, scale and expressivity needed by biological systems.

During the last years, our goal was to provide **formal models of queries on biological networks** with the focus of integrating dynamical information as explicit logical constraints in the modeling process. Using these technologies requires to revisit and reformulate constraint-satisfiability problems at hand in order both to decrease the search space size in the grounding part of the process and to improve the exploration of this search space in the solving part of the process. Concretely, getting logical encoding for the optimization problems forces to clarify the roles and dependencies between parameters involved in the problem. This paves the way to a refinement approach based on a fine investigation of the space of hypotheses in order to make it smaller

and gain in the understanding of the system. Our studies confirmed that logical paradigms are a powerful approach to build and query reconstructed biological systems, in complement to discriminative ("black-box") approaches based on statistical machine-learning. Based on these technologies, we have developed a panel of methods allowing the integration of muli-scale data knowledge, linking genomics, metabolomics, expression data and protein measurement of several phenotypes.

Notice that our main issue is in the field of knowledge representation. More precisely, we do not wish to develop new solvers or grounders, a self-contained computational issue which is addressed by specialized teams such as our collaborator team in Potsdam. Our goal is rather to investigate how the constant progresses in the field of constraint logical programming, shown by the performance of ASP-solvers, are sufficient to address the complexity of constraint-satisfiability and combinatorial optimization issues explored in systems biology. In this direction, we work in close interaction with Potsdam university to feed their research activities whith challenging issues from bioinformatics and, as a feed-back, take benefit of the prototypes they develop.

By exploring the complete space of models, our approach typically produces numerous candidate models compatible with the observations. We began investigating to what extent domain knowledge can further refine the analysis of the set of models by identifying classes of similar models, or by selecting a subset of models that satisfy an additional constraint (for instance, best fit with a set of experiments, or with a minimal size). We anticipate that this will be particularly relevant when studying non-model species for which little is known but valuable information from other species can be transposed or adapted. These efforts consist in developing reasoning methods based on ontologies as formal representation of symbolic knowledge. We use Semantic Web tools such as SPARQL for querying and integrating large sources of external knowledge, and measures of semantic similarity and particularity for analyzing data.

3.2. Modeling the dynamical response of biological systems with logical and (non)-linear constraints

As explained below, Answer Set Programming technologies enable the identification of key controllers based on the integration of static data. As a natural follow-up, we also develop optimization techniques to learn models of the dynamics of a biological system. As before, our strategy is not to select a single model fitting with experimental data but rather to decipher the complete set of families of models which are compatible with the observed response. Our main research line in this field is to decipher the appropriate level of expressivity (in terms of constraints) allowing both to properly report the nature of data and knowledge and to allow for an exhaustive study of the space of feasible models. To implement this strategy, we rely on several constraint programming frameworks, which depend on the model scale and the nature of time-points kinetic measurements. Logical programming (Answer Set Programming) is used to decipher the combinatorics of synchrone Boolean networks explaining static or dynamics response of signaling networks to perturbations (such as measured by phosphoproteomics technologies) [7]. SAT-based approaches are used to decipher the combinatorics of large-scale asynchronous boolean networks. In order to gain in expressivity, we model these networks as guarded-transition network, an extension of Petri nets [40],. Finally, classical learning methods are used to build ad-hoc parameterized numerical models that provide the most parsimonious explanations to experimental measurements.

3.3. Modeling sequences with formal grammars

Once groups of genome products involved in the answer of the species have been identified with integrative or dynamical methods, it remains to characterize the biological actors within genomes. To that goal, we both learn, model and parse formal patterns within DNA, RNA or protein sequences. More precisely, our research on modeling biomolecular sequences with expressive formal grammars focuses on learning such grammars from examples, helping biologists to design their own grammar and providing practical parsing tools.

On the development of **machine learning** algorithms for the induction of grammatical models [66], we have a strong expertise on learning finite state automata. We have proposed an algorithm that learns successfully automata modeling families of (non homologous) functional families of proteins [5], leading to a tool named Protomata-learner. The algorthim is based on a similar fragment merging heuristic approach which reports partial and local alignments contained in a family of sequences. As an example, this tool allowed us to properly model the TNF protein family, a difficult task for classical probabilistic-based approaches. It was also applied successfully to model important enzymatic families of proteins in cyanobacteria [4]. Our future goal is to further demonstrate the relevance of formal language modeling by addressing the question of a fully automatic prediction from the sequence of all the enzymatic families, aiming at improving even more the sensitivity and specificity of the models. As enzyme-substrate interactions are very specific central relations for integrated genome/metabolome studies and are characterized by faint signatures, we shall rely on models for active sites involved in cellular regulation or catalysis mechanisms. This requires to build models gathering both structural and sequence information in order to describe (potentially nested or crossing) long-term dependencies such as contacts of amino-acids that are far in the sequence but close in the 3D protein folding. Our current researches is focused on the inference of Context-Free Grammars including the topological information coming from the structural characterization of active sites.

Using context-free grammars instead of regular patterns increases the complexity of **parsing** issues. Indeed, efficient parsing tools have been developed to identify patterns within genomes but most of them are restricted to simple regular patterns. Definite Clause Grammars (DCG), a particular form of logical context-free grammars have been used in various works to model DNA sequence features [86]. An extended formalism, String Variable Grammars (SVGs), introduces variables that can be associated to a string during a pattern search [110], [109]. This increases the expressivity of the formalism towards mildly context sensitive grammars. Thus, those grammars model not only DNA/RNA sequence features but also structural features such as repeats, palindromes, stem/loop or pseudo-knots. A few years ago, we have designed a first tool, STAN (suffix-tree analyser), in order to make it possible to search for a subset of SVG patterns in full chromosome sequences [92]. This tool was used for the recognition of transposable elements in *Arabidopsis thaliana* [113]. We have enlarged this experience through a new modeling language, called Logol [1]. Generally, a suitable language for the search of particular components in languages has to meet several needs : expressing existing structures in a compact way, using existing databases of motifs, helping the description of interacting components. In other words, the difficulty is to find a good tradeoff between expressivity and complexity to allow the specification of realistic models at genome scale. The Logol language and associated framework have been built in this direction. The Logol specificity besides other SVG-like languages mainly lies in a systematic introduction of constraints on string variables.

3.4. Symbolic methods for model space exploration: Semantic web for life sciences and Formal Concepts Analysis

All the methods presented in the previous sections usually result in pools of candidates which equivalently explain the data and knowledge. These candidates can be dynamical systems, compounds, biological sequences, proteins... In any case, the output of our formal methods generally requires *a posteriori* investigation and filtering by domain experts. In order to assist them, we rely on two classes of symbolic technics: Semantic Web technologies and Formal Concept Analysis (FCA). They both aim at the formalization and management of knowledge, that is, the explicitation of relations occuring in structured data. These technics complement each other: the production of relevant concepts in FCA highly depends on the availability of semantic annotations using a controlled set of terms and conversely, building and exploiting ontologies is a complex process that can be made much easier with FCA.

Integrating heterogenous data with semantic web technologies The emergence of ontologies in biomedical informatics and bioinformatics happened in parallel with the development of the **Semantic Web** in the computer science community [108]. Let us recall that the Semantic Web is an extension of the current Web that provides an infrastructure integrating data and ontologies in order to support unified reasoning. Since the beginning, life sciences have been a major application domain for the Semantic Web [48]. This was

motivated by the joint evolution of data acquisition capabilities in the biomedical field, and of the methods and infrastructures supporting data analysis (grids, the Internet...), resulting in an explosion of data production in complementary domains [58], [49]. Consequently, Semantic Web technologies have become an integral part of translational medicine and translational bioinformatics [63]. The Linked Open Data project promotes the integration of data sources in machine-processable formats compatible with the Semantic Web [57], with a strong involvement of life sciences in this initiative.

However, a specificity of life sciences "data deluge" is that the proportion of generated data is much higher than in the more general "big data phenomenon", and that these data are highly connected [112]. **The bottleneck that once was data scarcity now lies in the lack of adequate methods supporting data integration, processing and analysis** [88]. Each of these steps typically hinges on domain knowledge, which is why they resist automation. This knowledge can be seen as the set of rules representing in what conditions data can be used or can be combined for inferring new data or new links between data.

In this setting, we are working on the integration of Semantic Web resources with our data analysis methods in order to take existing biological knowledge into account. We have introduced several methods to interpret semantic similarities and particularities [56], [55]. We now focus our attention on the semi-automated construction of RDF abstractions of heterogeneous datasets which can be handled by non-expert users. This allows both to automatically prepare input datasets for the other methods developed in the team and to analyse the output of the methods in a wide knowledge context.

Using Formal concept analysis to explore the results of bioinformatics analyses Formal concept analysis aims at the development of conceptual structures which can be logically activated for the formation of judgments and conclusions [117]. It is used in various domains managing structured data such as knowledge processing, information retrieval or classification [90]. In its most simple form, one considers a binary relation between a set of objects and a set of attributes. In this setting, formal concept analysis formalizes the semantic notions of extension and intension. Concepts are related within a lattice structure (Galois connection) by subconcept-superconcept relations, and this allows drawing causality relations between attribute subsets. In bioinformatics, it has been used to derive phylogenetic relations among groups of organisms [87], a classification task that requires to take into account many-valued Galois connections. We have proposed in a similar way a classification scheme for the problem of protein assignment in a set of protein families [67].

One of the most important issue with concept analysis is due to the fact that current methods remain very sensitive to the presence of uncertainty or incompleteness in data. On the other hand, this apparent defect can be reversed to serve as a marker of incompleteness or inconsistency [68]. Following this inspiration, we have proposed a methodology to tackle the problem of uncertainty on biological networks where edges are mostly predicted links with a high level of false positives [118]. The general idea consists to look for a tradeoff between the simplicity of the conceptual representation and the need to manage exceptions. As a very prospective challenge, we are exploring the idea of using ontologies to help this or to help ontology refinement using concept analysis [93], [52], [102].

More generally, common difficult tasks in this context are visualization, search for local structures (graph mining) and network comparison. Network compression is a good solution for an efficient treatment of all these tasks. This has been used with success in power graphs, which are abstract graphs where nodes are clusters of nodes in the initial graph and edges represent bicliques between two sets of nodes [104]. In fact, concepts are maximal bicliques and we are currently developing the power graph idea in the framework of concept analysis.

3.5. Implementing methods in software and platforms

Seven platforms have been developed in the team for the last five years: Askomics, AuReMe, FinGoc, Caspo, Cadbiom, Logol, Protomata. Indeed, one of the team's goals is to facilitate interplays between the tools for biological data analysis and integration. Improvements and novelties of these platforms are described in the "software" section. Our platforms 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]. It worths considering them into larger dedicated environments to benefit from the expertise of other research groups. The **BioShadock** repository of the GenOuest platform allows one to share the different docker containers that we are developing [website]. The **GenOuest galaxy portal** of the GenOuest platform now provides access to most tools for integrative biology and sequence annotation (access on demand).

3.5.1. *AskOmics platform*

Goal Integration and interrogation software for linked biological data based on semantic web technologies [url].

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 or directly in RDF) to be loaded into a Triple Store system using a user-friendly web interface. *AskOmics* also provides an intuitive graph-based user interface supporting the creation of complex queries that currently require hours of manual searches across tens of spreadsheet files. The elements of interest selected in the graph are then automatically converted into a SPARQL query that is executed on the users's data.

Originality Our experience is that end users (i) do not benefit for all the information available in the LOD cloud repositories by lack of SPARQL expertise (understandably: they are biologists and most of them do not have an interest in either learning SPARQL nor in learning how to integrate data); (ii) do not contribute their data back to the LOD cloud. Again, they do not have the expertise nor the resources to produce and maintain datasets and the associated metadata as linked data, nor to maintain the underlying server infrastructure. Therefore there is a need for helping 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. In this context, the main originality is the graphical interface that allows any SPARQL query to be built transparently and iteratively by a non-expert user.

Application This software was developed in the context of the MirnAdapt (pea-aphid) project in 2016. The tool has been presented to the agriculture communities in conferences [53], [84] and to the Galaxy community [29]. Up to now, more than 10 biological partners team are actually testing and using the prototype software (colza, pea-aphids, copper microbiology, marine biology), and SANOFI has shown its interest to co-develop the tool. Even if its current user base belongs to the bioinformatics community, the scope of *AskOmics* is domain-independent and has the potential to reach a wider audience related to the Semantic Web community.

3.5.2. *AuReMe workspace*

Goal Tracable reconstruction of metabolic networks [url].

Description The toolbox *AuReMe* allows for the Automatic Reconstruction of Metabolic networks based on the combination of multiple heterogeneous data and knowledge sources [64]. It is available as a Docker image. Five modules are composing *AuReMe*: 1) The Model-management PADmet module allows manipulating and traceing all metabolic data via a local database. [package] 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 [107], [65] and [21] [python package] 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 [60] [python package]. 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 [package].

Originality The main added-values are the inclusion of graph-based gap-filling tools that are particularly relevant for the study of non-classical organisms, the possibility to trace the reconstruction and curation procedures, and the representation and exploration of reconstructed metabolic networks with wikis.

Application The tools included in *AuReMe* have been used for reconstructing metabolic networks of micro and macro-algae [97], extremophile bacteria [16] and communities of organisms [61] in the context of the Idealg, Ciric-omics and IPL Algae-In-Silico projects.

3.5.3. *FinGoc-tools*

Goal Filtering interaction networks with graph-based optimization criteria.

Description The goal is to offer a set of tools for the reconstruction of networks from genome, literature and large-scale observation data (expression data, metabolomics...) in order to elucidate the main regulators of an observed phenotype. Most of the optimization issues are addressed with Answer Set Programming. 1) The **lombarde package** enables the filtering of transcription-factor/binding-site regulatory networks with mutual information reported by the response to environmental perturbations. The high level of false-positive interactions is filters according to graph-based criteria. Knowledge about regulatory modules such as operons or the output of the shogen package can be taken into account [39], [38] [web server]. 2) The **KeyRegulatorFinder package** allows searching key regulators of lists of molecules (like metabolites, enzymes or genes) by taking advantage of knowledge databases in cell metabolism and signaling. The complete information is transcribed into a large-scale interaction graph which is filtered to report the most significant upstream regulators of the considered list of molecules [59] [package]. 3) The **powerGrasp python package** provides an implementation of graph compression methods oriented toward visualization, and based on power graph analysis. [package]. 4) The **iggy package** enables the repairing of an interaction graph with respect to expression data. It proposes a range of different operations for altering experimental data and/or a biological network in order to re-establish their mutual consistency, an indispensable prerequisite for automated prediction. For accomplishing repair and prediction, we take advantage of the distinguished modeling and reasoning capacities of Answer Set Programming. [6] [114] [Python package]

Originality The main added-value of these tools is to make explicit the criteria used to highlight the role of the main regulators: the underlying methods encode explicit graph-based criteria instead of relying on statistical approaches. This makes it possible to explain local relationships and patterns within interaction graphs by explicit biological relationships.

Application The tools have been used to figure out the main gene-regulators of the response of porks to several diets in [74], [76] and [18]. The tools were also used to to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants [85], [119].

3.5.4. *Caspo software*

Participant: Anne Siegel.

Goal Studying synchronous boolean networks [url]

Description Cell ASP Optimizer (*Caspo*) constitutes a pipeline for automated reasoning on logical signaling networks. The main underlying issue is that inherent experimental noise is considered, many different logical networks can be compatible with a set of experimental observations (see [106] and [22]). It is available as a Docker container. Five modules are composing *Caspo*: 1) the *Caspo-learn* module performs an automated inference of logical networks from experimental data allows for identifying admissible large-scale logic models saving a lot of efforts and without any a priori bias [115] and [78]. 2) The *Caspo-classify*, predict and visualize modules allows for classifying a family of boolean networks with respect to their input-output predictions [78]. 3) The *Caspo-design* module designs experimental perturbations which would allow for an optimal discrimination of rival models in a family of boolean networks [116]. 4) The *Caspo-control* module identifies key-players of a family of networks: it computes robust intervention strategies that force a set of target species or compounds into a desired steady state [80]. 5) The *Caspo-timeseries* module to take into account time-series observation datasets in the learning procedure [94] [python package and docker container].

Originality The *Caspo* modules provide friendly and efficient solutions to problems that were previously addressed in theoretical papers with MILP programs. The main advantage is that is enables a complete study of logical network without requiring any linear constraint programs.

Application The *Caspo* tool was initiated in the framework of the BioTempo project. *Caspo-learn* has been included as a module to learn logical networks from early steady-state data in CellNopt, a generic platform which implements several methods for learning and studying signaling networks are different modeling levels (from logical models to numerical models).

3.5.5. *Cadbiom package*

Goal Building and analyzing the asynchronous dynamics of enriched logical networks [url]

Description Based on Guarded transition semantic, the *Cadbiom* software provides a formal framework to help the modeling of biological systems such as cell signaling network. It allows synchronization events to be investigated in biological networks [40]. It is available as a Docker image. 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, the translations of the large-scale PID repository (about 10,000 curated interactions) have been translated into the *Cadbiom* formalism.

Originality Model-checking approaches applied to Boolean networks [81] or multivalued networks [91] allow the trajectories of the system to be entirely studied but they can only be applied to small-size networks. On the contrary, *Cadbiom* is able to handle large-scale knowledge databases.

Application The *Cadbiom* tool was applied to study the regulators of the TGF-β, a gene that controls liver fibrosis [40] in the framework of the *TGFSysBio* project. The study of its predictions also enabled large-scale knowledge databases (PID) to be curated [25].

3.5.6. *Logol software*

Goal Complex pattern modelling and matching [url]

Description The *Logol* toolbox is a swiss-army-knife for pattern matching on DNA/RNA/Protein sequences, using a high-level grammatical formalism to permit a large expressivity for patterns [50]. A *Logol* pattern can consist in a complex combination of motifs (such as degenerated strings) and structures (such as imperfect stem-loop ou repeats). *Logol* key features are the possibilities to divide a pattern description into several sub-patterns, to model long range dependencies, to enable the use of ambiguous models or to permit the inclusion of negative conditions in a pattern definition. The *LogolMatch* parser takes as input a biological sequence and a grammar file. It returns a XML file containing all the occurrences of the pattern in the sequence with their parsing details. The input sequences can be genomes from biological banks.

Originality Many pattern matching tools exist to efficiently model specific types of patterns: vmatch [82], patmatch [121], cutadapt [89], scoring matrix or profile HMMs [44], [71]. The main advantage of *Logol* is its very large expressivity. It encompasses most of the features of these specialized tools and enables interplays between several classes of patterns (motifs and structures).

Application The *Logol* tool was applied to the detection of mutated primers in a metabarcoding study [41], [42] or to stem-loop identification (e.g. in CRISPR [1] [103], [50]). Ongoing application is the search for transposable elements in the human genome in the context of a colorectal cancer study [2]. *Logol* strongly supported the study of the LXR-α targets in the framework of the FatInteger project.

3.5.7. *Protomata-suite*

Goal Expressive pattern discovery on protein sequences [url]

---

[1] http://crispi.genouest.org/, https://hal.inria.fr/hal-00643408
[2] Preprint: http://www.biorxiv.org/content/early/2017/03/09/115030

Description*Protomata* is a machine learning suite for the inference of automata characterizing (functional) families of proteins from available sequences. Based on partial and local alignments, *Protomata* learns precise characterizations of the families of proteins, allowing new family members to be predicted with a high specificity. Three main modules are integrated in the *Protomata-learner* workflow are available as well as stand-alone programs: 1) *paloma* builds partial local multiple alignments, 2) *protobuild* infers automata from these alignements and 3) *protomatch* and *protoalign* scans, parses and aligns new sequences with learnt automata. The suite is completed by tools to handle or visualize data and can be used online by the biologists via a web interface on Genouest Platform. It is actively maintained (version v2.1 was released in April 2017) and we are scheduling a new major version with enhanced scoring schemes that we have proposed [105].

Originality The main specificity is that the power of characterization is beyond the scope of classical sequence patterns such as PSSM (e.g. MEME suite [43]), Profile HMM (e.g. HMMER package [71]), or Prosite Patterns [111] allowing new family members to be predicted with a high specificity.

Application The *Protomata* tool is used both to update automatically the Cyanolase database [62] and, when combined to Formal Concept Analysis, to automated enzyme classification, such as the HAD superfamily of proteins [67] in the framework of the *Idealg project*.

4. Application Domains

4.1. Application fields in biology

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco, memap and memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [97], [69]. On the same application model, the pattern discovery tool *protomata learner* combined with supervised bi-clustering based on formal concept analysis allows for the classification of sub-families of specific proteins [67]. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [4]. At the larger level of 4D structures, classification technics have also allowed us to introduce new methods for the characterization of viruses in marine metagenomic sample [72]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [51] [95]. We are currently involved in two new applications in this domain: the team participates to a Inria Project Lab program with the Biocore and Ange Inria teams, focused on the understanding on green micro-algae; and we are involved in the deciphering of phytoplancton variability at the system biology level in collaboration with the Station Biologique de Roscoff (ANR Samosa).

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *meneco bioquali, ingranalysis, shogen, lombarde* were designed in this context [77]. In 2016, two applications focused on the study of extremophile consortium of bacteria have been performed with these tools [61], [38]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archebacteria [96]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [10] [50].

Similarly, in **environmental sciences**, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. In collaboration with researchers from Inra/Pegase laboratory, we develop methods to distinguish the response of breeding animals to different diaries or treatments [37] and characterize upstream transcriptional regulators [59], applied to porks [74], [75], [76]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [79].

In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants [85], [120]. This was performed in collaboration with Inra/Igepp. This paved the way to the recent research track initiated in the team about integration of heterogeneous data with RDF-technologies (see AskOmics software) [84], [70] and about graph-compression (see powergrasp software).

In **bio-medical applications**, we focus our attention on the confrontation of large-scale measurements with large-scale knowledge repositories about regulation pathways such as Transpath, PID or pathway commons. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [47]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [115] and N. Theret group at Inserm/Irset (Rennes) [40]. The dynamical system tools *caspo and cadbiom* were designed within these collaborations. Ongoing studies focus on the understanding of the metabolism of xenobiotics (mecagenotox program) and the filtering of sets of regulatory compounds within large-scale signaling network (TGFSysBio project).

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

The team received a best paper award at the conference ICFCA and a best student paper award at the conference LPNMR .

BEST PAPERS AWARDS:

[24]
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

[26]
C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO. *Hybrid Metabolic Network Completion*, in "14th International Conference on Logic Programming and Nonmonotonic Reasoning - LPNMR 2017", Espoo, Finland, M. BALDUCCINI, T. JANHUNEN (editors), Logic Programming and Nonmonotonic Reasoning, Springer, July 2017, vol. 10377, pp. 308-321, https://hal.inria.fr/hal-01557347

6. New Software and Platforms

6.1. AskOmics

KEYWORDS: RDF - SPARQL - Querying - Graph - LOD - Linked open data
FUNCTIONAL DESCRIPTION: AskOmics allows to load heterogeneous bioinformatics data (formatted as tabular files) into a Triple Store system using a user-friendly web interface. AskOmics also provides an intuitive graph-based user interface supporting the creation of complex queries that currently require hours of manual searches across tens of spreadsheet files. The elements of interest selected in the graph are then automatically converted into a SPARQL query that is executed on the users' data.

NEWS OF THE YEAR: Several functionalities have been developed: 1) capacity of integrating genomics data (import of GFF and BED files and generation of RDF compliant with the FALDO ontology), 2) integration of data and knowledge in the OWL format to exploit biological information from external repositories, particularly from EBI and NCBI. Notably, this functionality allows AskOmics to support the Gene Ontology, the Taxonomy ontology as well as BioPAX biological networks. 3) improved user interface expressivity for generating SPARQL queries, 4) implementation of a support for multiple concurrent user sessions, with the distcintion between public and user-specific datasets 5) deployment of AskOmics on the GenOuest cloud infrastructure to facilitate its release and diffusion 6) interoperability between AskOmics and the Galaxy workflow environment.

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

6.2. PADMet-utils

KEYWORDS: Metabolic networks - Bioinformatics - Workflow - Omic data - Toolbox - Data management - LOD - Linked open data
FUNCTIONAL DESCRIPTION: The main concept underlying padmet-utils is to provide solutions that ensure the consistency, the internal standardization and the reconciliation of the information used within any workflow that combines several tools involving metabolic networks reconstruction or analysis.
NEWS OF THE YEAR: In 2017, Padmet-utils was enriched with a RDF export to allow the interoperability of the AuReMe workspace for the reconstruction of metabolic networks with the Askomics Tool for querying heterogeneous data. Padmet-utils was also extended to handle metabolic networks in the SBML3 format.

- Participants: Alejandro Maass, Meziane Aite and Anne Siegel
- Partner: University of Chile
- Contact: Anne Siegel
- URL: https://gitlab.inria.fr/maite/padmet-utils

6.3. CADBIOM

Computer Aided Design of Biological Models
KEYWORDS: Health - Biology - Biotechnology - Bioinformatics - Systems Biology
FUNCTIONAL DESCRIPTION: Based on Guarded transition semantic, this software provides a formal framework to help the modeling of biological systems such as cell signaling network. It allows investigating synchronization events in biological networks. Software development has been restarted since November 2016. The source code is available at the following address: https://gitlab.irisa.fr/0000B8EG/Cadbiom

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

6.4. conquests

Crossroads in Metabolic Network from Stoechiometric and Topologic Studies
KEYWORDS: Bioinformatics - ASP - Answer Set Programming - Constraint-based programming

FUNCTIONAL DESCRIPTION: This Python package in systems biology allows the identification of essential metabolites with respect to the production of targeted elements in a metabolic network, by comparing flux and graph-based analysis. Conquests's inputs are a sbml file corresponding to a metabolic network and the biomass reaction name. The outputs are three sets of essential metabolites. They are computed according to three complementary criteria: graph-based accessibility of targeted metabolites, the presence of flux in the biomass reaction and the maximisation of flux in the biomass reaction.

NEWS OF THE YEAR: Conquest was released in 2017.

- Contact: Julie Laniau

7. New Results

7.1. Data integration and pre-processing with semantic-based technologies

Participants: Olivier Dameron, Xavier Garnier, Yann Rivault, Anne Siegel, Denis Tagu.

Interoperable infrastructure and implementation of a health data model for remote monitoring of chronic diseases with comorbidities In the context of **telemedecine**, we worked on a numerical application for monitoring patients with chronic diseases. We have developed a system based on a formal ontology that integrates the alert information and the patient data extracted from the electronic health record in order to better classify the importance of alerts. A pilot study was conducted on atrial fibrillation alerts. The results suggest that this approach has the potential to significantly reduce the alert burden in telecardiology [101], [100]. In 2017, we proposed an architecture supporting data exchange in the context of multiple chronic diseases [*O. Dameron, Y. Rivault*] [27].

AskOmics, a web tool to integrate and query biological data using semantic web technologies The software AksOmics has been adapted to two types of scientific topics important in agronomical and environmental sciences: plant genomic data and insect pest genomic data. With *AskOmics*, plant genomicists (from academic and private labs from the Rapsodyn project - Investment for the future) working on the rapeseed (Brassica napus) are able to tackle the understanding of which gene copy is active or repressed in key developmental processes in relation with seed quality and oil production, in the frame of plant breeding. Additionally, entomologists use this tool to extract valuable knowledge on the way insect pests such as aphids are able to rapidly disseminate on crops, in the frame of free-pesticide methods for plant protection. *AskOmics* has been presented to the international community of insect genomics (i5k: http://i5k.github.io/) by web-seminars and *AskOmics* developers have been invited at international workshops. For facilitating AskOmics's adoption by end-user, it has recently been integrated within the Galaxy workflow engine [*O. Dameron, X. Garnier, A. Siegel, D. Tagu*] [28], [29], [23]

7.2. Data and knowledge integration based on combinatorial optimization

Participants: Meziane Aite, Lucas Bourneuf, Marie Chevallier, Damien Eveillard, Clémence Frioux, Jeanne Got, Julie Laniau, François Moreews, Jacques Nicolas, Anne Siegel.

A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [59] was valuable for figuring out the main gene-regulators of the response of porks to several diets [*F. Moreews, A. Siegel*] [18]

FCA in a Logical Programming Setting for Visualization-oriented Graph Compression We have explored the underlying idea of lossless **network compression** to address the problem of uncertainty in biological networks built from predictions, to help to visualize the networks and to classify their nodes in accordance with available annotations [119]. Network compression has been used with success in Dresden (M. Schroeder) with a heuristic approach called Power Graph analysis building abstract graphs where nodes are clusters of nodes in the initial graph and edges represent bicliques between two sets of nodes. First encouraging results have been presented (best paper award) showing that it is possible to mimic the Power Graph behaviour while opening the possibility to achieve better compression levels compared to alternative compression schema. [*L. Bourneuf, J. Nicolas*] [24]

Metabolic network completion and analysis We released the application paper of the tool *Meneco*, a tool dedicated to the topological gap-filling of genome-scale draft metabolic networks. The tool reformulates gap-filling as a qualitative combinatorial optimization problem, omitting constraints raised by the stoichiometry, and solves this problem using Answer Set Programming. Run on an artificial test set of 10,800 degraded *Escherichia coli* networks, we evidenced that *Meneco* outperforms the stoichiometry-based tool *Gapfill* in terms of precision. In addition, *Meneco* reports 10 times less putative reactions than MILP-based tool *Fastgapfill* for an equivalent precision. This is a strong advantage for manual curation post-processing, since curating 50 to 80 reactions is still possible whereas manually-curating 800 reactions is out-of-range. Meneco was applied to the reconstruction and understanding of a pathogeneic strain of salmon. [*C. Frioux, J. Got, A. Siegel*] [21], [16]

Toward the study of metabolic functions in communities of organisms In [21], we provided a first example on how to use topological metabolic modeling to assess the complementarity between two members of an algal ecosystem. Since this study, we generalized the selection of subcommunities of interest and propose likely interactions that could occur between seaweeds and their associated bacteria. A focus has also been done on plant microbiota and the reasons underlying the organization of the community. Altogether, these on-going works enable a better understanding of holobiont organizations and functioning. [*M. Aite, M. Chevallier, C. Frioux, J. got, A. Siegel, C. Trottier*] [21], [31], [30]

Hybrid Metabolic Network Completion In order to improve the precision of gap-filling approaches, we introduced a hybrid approach to formally reconcile existing stoichiometric and topological approaches to network completion in a unified formalism. An hybrid ASP encoding based on MILP constraint propagator was developed. It relies upon the theory reasoning capacities of the ASP system Clingo to solve the resulting logic program with linear constraints over reals. For short, this technology made it possible to combine the best of the combinatorial problem solver Clingo with the MILP solver CPlex. Run on the artificial test set of 10,800 degraded *Escherichia coli* networks introduced in [21], our approach yielded greatly superior results than obtainable from purely qualitative or MILP approaches. [*C. Frioux, A. Siegel*] [26], [19]

Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks Whenever flux or graph-based criteria are used to study metabolic networks, these analyses are generally centered on the outcome of the network and considers all metabolic compounds to be equivalent in this respect. We generalized the concept of essentiality to metabolites and introduced the concept of the phenotypic essential metabolite (PEM) which influences the growth phenotype according to sustainability, producibility or optimal-efficiency criteria. The exhaustive study of phenotypic essential metabolites in six genome-scale metabolic models suggests that the combination and the comparison of graph, stoichiometry and optimal flux-based criteria allow some features of the metabolic network functionality to be deciphered by focusing on a small number of compounds. [*C. Frioux, J. Laniau, A. Siegel*] [19]

7.3. Systems biology

Participants: Jérémie Bourdon, Jean Coquet, Victorien Delannée, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

A modeling approach to evaluate the balance between bioactivation and detoxification of MeIQx in human hepatocytes Heterocyclic aromatic amines (HAA), including MeIQx, are environmental and food contaminants that are potentially carcinogenic for humans. Using a computational approach, we developed a numerical model for MeIQx metabolism that predicts the MeIQx biotransformation into detoxification or bioactivation pathways according to the concentration of MeIQx. Our results demonstrate that CYP1A2 is a key enzyme in the system that regulates the balance between bioactivation and detoxification. This highlights the importance of complex regulations of enzyme competitions that should be taken into account in any multi-organ model [*V. Delannée, A. Siegel, N. Théret*] [17]

caspo: a toolbox for automated reasoning on the response of logical signaling networks families The accompanying paper of the complete family of modules introduced in the caspo software was published in 2017 (see software section for details) [*A. Siegel*] [22]

Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering: Application to TGF-β Signaling At a dynamical level, in [40], reaction-based and regulatory information was transpoed in a unified formalism of enriched Petri Nets (discrete dynamical systems), namely a simplified version of guarded transitions in which we introduced temporal parameters for each transition to manage competition and cooperation between parts of the models. This allowed integrating the 137 human signaling maps from the Pathway Interaction Database (PID) into a single unified large-scale dynamic model. Simulation and model checking analyses evidence that 15,934 different sets of molecules are able to regulate 159 of TGF-β target genes (TGF-β is a multifunctional cytokine that regulates mammalian cell development, differentiation, and homeostasis). Further analysis of these 15,934 sets of molecules by biological experts is obviously impractical. Our study identified five clusters of sets of molecules for which enrichment analysis highlighted the over-represented molecules as well as the specific biological processes they are associated with. These results are biologically-relevant and consistent with the pleiotropic nature of TGF-β [*J. Coquet, N. Théret, O. Dameron*] [25]

A Logic for checking the probabilistic steady-state properties of reaction networks. We have constructed a probabilistic analog to flux balance analysis of reaction networks to enable a formal verification of logical constrains about the stationary regime of a system by using information from experimental variances and co-variances. This is mainly based on a stationary analysis of the probabilistic dynamics relying on a Bernoulli approximation of a reaction network. The analysis requires solving non linear optimization problems [*J. Bourdon, A. Siegel*] [20]

7.4. Sequence and structure annotation

Participants: Catherine Belleannée, François Coste, Jacques Nicolas.

Better scoring schemes for the recognition of functional proteins by protomata The machine learning algorithm included in *Protomata-learner* learns weighted automata representing both functional families from the sequences of amino acids, and the possible disjunctions between members. We investigated alternative sequence weighting strategies and null-models. We introduced a normalization of the score, and a method to assess the significance of scores, to simplify the prediction. Preliminary results show a good improvement of the prediction power of the computed models. [*F. Coste*] [36]

Detection of mutated primers and impact on targeted metagenomics results In targeted metagenomics, an initial task is the detection in each sequence of the primers used for amplifying the targeted region. The selected sequences are then trimmed and clustered in order to inventory the species present in the sample. Common pratices consist in retaining only the sequences with perfect primers (i.e. non-mutated by sequencing error). In the context of a study characterizing the biodiversity of tropical soils in unicellular eukaryotes, we have implemented the search for mutated primers, using the grammatical pattern matching tool Logol, and shown that retrieving sequences with mutated primers has a significant impact on targeted metagenomics results, as it makes possible to detect more species (7% additional OTUs in our study). [*C. Belleannée*] [34].

First landscape of binding to chromosomes for a domesticated mariner transposase in the human genome. In order to study the diversity of genomic targets of the SETMAR protein in two colorectal cell lines, a first task was to massively discover the Made1 80-bp transposon element in the human genome. For that, we used our Logol grammar-like approach to look for non perfect Made1 instances. In Logol, a pattern can be divided into several sub-patterns. The Made1 model took advantage of this feature to strengthen the most conserved regions. Cumulating this search with the Blast alignment search permitted to significantly increase the Made1 annotation in the human genome.[*C. Belleannée*] [33]

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Participants: Olivier Dameron, Anne Siegel, Méline Wery.

Our software *AskOmics* was considered as relevant by the SANOFI bio-medical company in order to facilitate the integration and the query of the data produced by their scientists. A former Ph.D. of Dyliss who designed the first prototypes of *AskOmics* was recruited by SANOFI. Since then, SANOFI is included in the developer's team of *AskOmics* and a joint Dyliss–SANOFI CIFRE Ph.D. thesis started about the integration of complementary reasoning features to SPARQL queries in Oct. 2017.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. *Regional initiative: the Ecosyst project*
Participants: Damien Eveillard, Marie Chevallier, Clémence Frioux, Anne Siegel, Camille Trottier.

EcoSyst is a Biogenouest inter-regional federating project (Brittany & Pays de la Loire) aiming at the emergence of Systems Ecology at the level of western France regions. Drawing on the strengths and skills involved, EcoSyst targets the incubation of new ideas and new projects at disciplinary interfaces. Thanks to this community project, we want to develop the skills of Ecology, Environment, Modeling, Bioinformatics and Systems Biology and their application to organisms and ecosystems of interest in agronomy, sea and health. EcoSyst includes also the identification of the major issues and concerns, the fundamental and essential methods and the very real needs of the community (training, tools, ...); this in order to consider the construction of a community platform (or an offer of service within an existing platform) on complex systems modeling, meeting expectations of the community as fully as possible.

9.1.2. *Regional partnership with computer science laboratories in Nantes*
Participants: Anne Siegel, Jérémie Bourdon, Damien Eveillard, François Coste, Maxime Folschette, Jacques Nicolas.

Methodologies are developed in close collaboration with the LS2N (fusion of LINA and IRCCyN) located at University of Nantes and École centrale de Nantes. This is acted through the Biotempo and Idealg ANR projects and co-development of common software toolboxes within the Renabi-GO platform support. C. Trottier is a co-supervised bioianalysis and software development engineer within the Idealg project. M. Chevallier is a co-supervised development and animation engineer within the regional initiative "Ecosyst". In addition, the ongoing Ph-D student J. Laniau is co-supervised with a member of the LS2N laboratory. Finally, M. Folschette is a PostDoc working on on a project aiming at analyzing TGF-beta-related pathways evolutions after epithelial-mesenchymal transition in liver cancer, which is a recognized biological process leading to metastasis. This project is based on a topic shared with the LS2N: the use of graph coloring and reconstruction to witness expression changes, and is funded by the Université Bretagne Loire.

9.1.3. *Regional partnership in Marine Biology*
Participants: Meziane Aite, Arnaud Belcour, Catherine Belleannée, Jérémie Bourdon, Jean Coquet, François Coste, Damien Eveillard, Olivier Dameron, Clémence Frioux, Jeanne Got, Julie Laniau, Jacques Nicolas, Camille Trottier, Anne Siegel.

A strong application domain of the Dyliss project is marine Biology. This application domain is co-developed with the station biologique de Roscoff and their three UMR and involves several contracts. Our approach based on parcimonious modelling allowed an in silico characterization of processes required within sea urchin translation [83], [95]. We are also strongly involved in the the IDEALG consortium, a long term project (10 years, ANR Investissement avenir) aiming at the development of macro-algae biotechnology. Among the research activities, we are particularly interested in the analysis and reconstruction of metabolism and the characterization of key enzymes. 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 *E. Siliculosus* **seaweed metabolism** to be deciphered [97], [67]. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* [69] and used this reconstructed network to decipher interactions within the **algal-bacteria holobiont** [21].

9.1.4. Regional partnership in agriculture and environmental sciences

Participants: Catherine Belleannée, François Coste, Olivier Dameron, Xavier Garnier, François Moreews, Jacques Nicolas, Anne Siegel, Denis Tagu.

We have a strong and long term collaboration with biologists of INRA in Rennes : PEGASE and IGEPP units. F. Morrews is a permanent engineer from PEGASE center hosted in the team to develop methods for integrative biology applied to species of interest in agriculture. D. Tagu is a research director at INRA/IGEPP who spends 20% of his time in the team to develop collaborative projects. This partnership has been supported by the co-supervision of phDs, post-docs and engineers. This collaboration was also reinforced by collaboration within ANR contracts (MirNadapt, FatInteger).

In collaboration with researchers from the PEGASE center (INRA) focused on breeding animals, we have contributed to several studies aiming at better integrating and investigating data in order to facilitate animal selection and alimentation. The *NutritionAnalyzer* prototype was developed to understand better the impact of several diaries or treatments for lactary cows over the composition of milk [37]. Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [59] and on semantic-based analysis of metabolic networks [54] was also very valuable for interpreting differences of gene expression in pork meat [79] and figure out the main gene-regulators of the response of porks to several diets (see [74], [76] and [18]).

In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants in the framework of the *MirnAdapt* project. In terms of biological output of the network studies on the pea aphid microRNAs, we have identified one new microRNA (apmir-3019, not present in any known species other than the pea aphid) who has more than 900 putative mRNA targets. All these targets, as well as apmir3019, are differentially expressed between sexual and asexual embryos [85], [119].

9.1.5. Regional partnership in health

Participants: Jean Coquet, Olivier Dameron, Victorien Delannée, Marine Louarn, Anne Siegel, Nathalie Théret, Pierre Vignet.

We also have a strong and long term collaboration in health, namely with the IRSET laboratory at Univ. Rennes 1. N. Théret, research director at INSERM, is hosted in the team to strenghen our collaborative projects. Our collaborations are acted by the co-supervised Ph-D theses of V. Delannée [14], M. Conan (Metagenotox project, funded by Anses) and J. Coquet [12]. This partnership was reinforced by the ANR contract Biotempo ended at the end of 2014. In 2015, the project of combining semantic web technologies and bi-clustering classification based on formal concept analysis was applied to systems biology within the PEPS CONFOCAL project. This scientific project has been recently pushed forward in the recent TGFSYSBio project funded by Plan Cancer on the modelling of the microenvironment of TGFbeta signaling network (P. Vignet has been recruited on this contract at the end of 2016).

A new application was initiated in 2017 through a collaboration with Rennes hospital, supported by a Inria-INSERM Ph-D thesis (M. Louarn).

9.2. National Initiatives

9.2.1. ANR Idealg

Participants: Meziane Aite, Arnaud Belcour, Jérémie Bourdon, Marie Chevallier, François Coste, Damien Eveillard, Clémence Frioux, Jeanne Got, Julie Laniau, Jacques Nicolas, Anne Siegel.

IDEALG is one of the five laureates from the national call 2010 for Biotechnology and Bioresource and will run until 2020. It gathers 18 different partners from the academic field (CNRS, IFREMER, UEB, UBO, UBS, ENSCR, University of Nantes, INRA, AgroCampus), the industrial field (C-WEED, Bezhin Rosko, Aleor, France Haliotis, DuPont) as well as a technical center specialized in seaweeds (CEVA) in order to foster biotechnology applications within the seaweed field. We are participating to the tasks related to the establishment of a virtual platform for integrating omics studies on seaweeds and the integrative analysis of seaweed metabolism, in cooperation with SBR Roscoff. Major objectives are the building of brown algae metabolic maps, flux analysis and the selection of symbiotic bacteria to brown algae. We will also contribute to the prediction of specific enzymes (sulfatases) [More details].

9.2.2. *Programs funded by research institutions*

9.2.2.1. PEPS PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases
Participants: Olivier Dameron, Yann Rivault.

As a partner of the PEPS platform, several teams at Inria Rennes develop generic methods supporting efficient and semantically-rich queries for pharmaco-epidemiology studies on medico-administrative databases. The leader is Thomas Guyet (Inria team Lacodam). We showed that Semantic Web technologies are technically suited for representing patients' data from medico-administrative databases as RDF and querying them using SPARQL. We also demonstrated that this approach is relevant as it supports the combination of patients' data with hierarchical knowledge in order to address the problem of reconciling precise patients data with more general query criteria [45], [99], [98]. This work is mostly conducted by Yann Rivault, whose PhD thesis is supervized by Olivier Dameron and Nolwenn LeMeur (Ecole des Hautes Etudes en Santé Publique).

9.2.2.2. Cancer Plan: TGFSysBio
Participants: Jean Coquet, Olivier Dameron, Maxime Folschette, Vijay Ingalalli, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

The TGFSYSBIO project aims to develop the first model of extracellular and intracellular TGF-beta system that might permit to analyze the behaviors of TGF-beta activity during the course of liver tumor progression and to identify new biomarkers and potential therapeutic targets. Based on collaboration with Jérôme Feret from ENS, Paris, we will combine a rule-based model (Kappa language) to describe extracellular TGF-beta activation and large-scale state-transition based (Cadbiom formalism) model for TGF-beta-dependent intracellular signaling pathways. The multi-scale integrated model will be enriched with a large-scale analysis of liver tissues using shotgun proteomics to characterize protein networks from tumor microenvironment whose remodeling is responsible for extracellular activation of TGF-beta. The trajectories and upstream regulators of the final model will be analyzed with symbolic model checking techniques and abstract interpretation combined with causality analysis. Candidates will be classified with semantic-based approaches and symbolic bi-clustering technics. The project is funded by the national program "Plan Cancer - Systems biology" from 2015 to 2018.

9.2.2.3. ANR Samosa
Participants: Mael Conan, Damien Eveillard, Jeanne Got, Anne Siegel.

Oceans are particularly affected by global change, which can cause e.g. increases in average sea temperature and in UV radiation fluxes onto ocean surface or a shrinkage of nutrient-rich areas. This raises the question of the capacity of marine photosynthetic microorganisms to cope with these environmental changes both at short term (physiological plasticity) and long term (e.g. gene alterations or acquisitions causing changes in fitness in a specific niche). *Synechococcus* cyanobacteria are among the most pertinent biological models to tackle this question, because of their ubiquity and wide abundance in the field, which allows them to be studied at all levels of organization from genes to the global ocean.

The SAMOSA project is funded by ANR from 2014 to 2018, coordinated by F. Gaczarek at the Station Biologique de Roscoff/UPMC/CNRS. The goal of the project is to develop a systems biology approach to characterize and model the main acclimation (i.e., physiological) and adaptation (i.e. evolutionary) mechanisms involved in the differential responses of *Synechococcus* clades/ecotypes to environmental fluctuations, with the goal to better predict their respective adaptability, and hence dynamics and distribution, in the context of global change. For this purpose, following intensive omics experimental protocol driven by our colleagues from − Station Biologique de Roscoff −, we aim at constructing a gene network model sufficiently flexible to allow the integration of transcriptomic and physiological data.

9.2.2.4. ANSES Mecagenotox
Participants: Victorien Delannée, Mael Conan, Anne Siegel, Nathalie Théret.

The objective of Mecagenotox project is to characterize and model the human liver ability to bioactivate environmental contaminants during liver chronic diseases in order to assess individual susceptibility to xenobiotics. Indeed, liver pathologies which result in the development of fibrosis are associated with a severe dysfunction of liver functions that may lead to increased susceptibility against contaminants. In this project funded by ANSES and coordinated by S. Langouet at IRSET/inserm (Univ. Rennes 1), we will combine cell biology approaches, biochemistry, biophysics, analytical chemistry and bioinformatics to 1) understand how the tension forces induced by the development of liver fibrosis alter the susceptibility of hepatocytes to certain genotoxic chemicals (especially Heterocyclic Aromatic Amines) and 2) model the behavior of xenobiotic metabolism during the liver fibrosis. Our main goal is to identify "sensitive" biomolecules in the network and to understand more comprehensively bioactivation of environmental contaminants involved in the onset of hepatocellular carcinoma.

9.2.3. Programs funded by Inria

9.2.3.1. ADT Complex-biomarkers and ADT Proof of concept
Participants: Jeanne Got, Marie Chevallier, Meziane Aite, Anne Siegel.

These projects started in Oct. 2014 and aims at designing a working environment based on workflows to assist molecular biologists to integrate large-scale omics data on non-classical species. The main goal of the workflows will be to facilitate the identification of set of regulators involved in the response of a species when challenged by an environmental stress. Applications target extremophile biotechnologies (biomining) and marine biology (micro-algae).

9.2.3.2. IPL Algae in silico
Participants: Meziane Aite, Jeanne Got, Julie Laniau, Anne Siegel.

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. It involves mainly the inria teams BIOCORE (PI), ANGE and DYLISS. 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.

9.2.3.3. IPL Neuromarkers
Participants: Olivier Dameron, Anne Siegel.

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

9.2.3.4. FederatedQueryScaler (Exploratory Research Action)
Participants: Olivier Dameron, Xavier Garnier, Vijay Ingalalli.

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. It is a common project with the WIMMICS Inria team.

9.3. European Initiatives

9.3.1. Collaborations with Major European Organizations

Partner: Aachen university (Germany)

Title: Modeling the logical response of a signalling network with constraints-programming.

9.4. International Initiatives

9.4.1. Inria International Labs

9.4.1.1. Other IIL projects

We have a cooperation with Univ. of Chile (MAHTomics, 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 manage, explore and integrate large sets of heterogeneous omics data into networks of interactions allowing to produce biomarkers, with a main application to biomining bacteria. The program is co-funded by Inria and CORFO-chile from 2012 to 2016. In this context, Integrative-BioChile 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]
- **India**. VIT University, Vellore [K. Lakshmanan]

9.5.2. Visits to International Labs

- **Chile**. University of Chile [A. Siegel, C. Frioux]

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 Selection

10.1.1.1. Member of the Conference Program Committees

- SWAT4HCLS (2017) Semantic Web and Tools for Health Care and Life Sciences (O. Dameron)
- BBCC (2017): Bioinformatica e Biologia Computazionale in Campania (O. Dameron)
- JOBIM (2017): French conference of Bioinformatics (A. Siegel)
- SIIM (2017) Symposium sur l'Ingénierie des Informations Médicales (O. Dameron)

10.1.1.2. Review

- ISMB/ECCB 2017.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- O. Dameron is an associate editor of the Journal of Biomedical Semantics

- J. Bourdon in an academic editor of PLoS One

10.1.2.2. Reviewer - Reviewing Activities

- Briefings in Bioinformatics, (O. Dameron)
- Journal of Biomedical Semantics (O. Dameron)
- Journal on Data Semantics (O. Dameron)
- Molecular Cancer (J. Nicolas),
- Plos One (J. Nicolas)

10.1.3. Invited Talks

10.1.3.1. French seminars

Paris, Hopital Lariboisière (Seminar, 2017) – SANOFI (Gentilly, 3 Invited seminars, 2017) – Clermont-Ferrand (Insect team, 2017) – Nantes (Université of Nantes, 2017).

10.1.3.2. Workshops and meetings

Conference on Boolean networks (Marseille, 2017) – Bioss Meeting on artificial intelligence (Gif, 2017)

10.1.4. Leadership within the Scientific Community

- Member of the steering committee of the International Conference on Grammatical Inference.
- 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. The group is supported by two French National Research Networks: bioinformatics (GDR BIM : bioinformatique moléculaire) and informatics-mathematics (GDR IM : Informatique Mathématique). It gathered twice in 2017: for a general meeting in Montpellier (Mar. 2017) and for a workshop focused on links between systems biology and artificial intelligence in Orsay (June 2017).

10.1.5. Scientific Expertise

10.1.5.1. International responsibilities

- Evaluation panel of the "Europe-USA Call Strengthening Transnational Research in Molecular Plant Sciences" launched by ERA-CAPS.

10.1.5.2. National responsibilities

- **Institutional boards for the recruitment and evaluation of researchers**. Inria National evaluation board (A. Siegel, nominated member). National Council of Universities, section 65 (O. Dameron, nominated member).
- **Evaluation committees of French laboratories or doctoral schools**. Bioinformatics groups of Institut Curie (Paris, presidency of the committee, A. Siegel) – Doctoral school of Nice University (N. Théret).
- Presidency of the expert panel for the call *Systems biology applied to Cancer* of the National Cancer Plan 2017 (A. Siegel).
- **Recruitment committees**. Inria Senior Researchers (national committee, A. Siegel) – Inria Junior Researchers (Nice, National Committee, A. Siegel)
- **Scientific Advisory boards** GDR BIM " Molecular Bioinformatics" (J. Nicolas).

10.1.5.3. National scientific boards

- Scientific Advisory Board of the French National Research Network GDR BIM *Molecular Bioinformatics* (J. Nicolas).
- Operational Legal and Ethical Risk Assessment Committee (COERLE) at Inria (J. Nicolas).
- Animation of the Bioss working group (A. Siegel).

- Board of directors of the French Society for biology of the extracellular matrix (N. Théret).

10.1.5.4. Prospective working groups

- "Big & Open Data" foresight working group of PROSPER network (F. Coste).
- "Prospectives in predictive toxicology" working group at INRA (A. Siegel)

10.1.5.5. Local responsibilities

- Scientific Advisory Board of Biogenouest (J. Bourdon, 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) [3]
- SCAS (Service Commun d'Action Sociale) of Univ. Rennes 1 (C. Belleannée)
- Scientific committee of Univ. Rennes 1 school of medicine (O. Dameron, A. Siegel).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching track responsibilities

- Coordination of the doctoral school "Life, Agronomy and Health" of University of Rennes 1 [N. Théret]
- Coordination of the master degree "Bioinformatics and genomics", Univ. Rennes1 [O. Dameron]
- Coordination of the sub-domain "From Data to Knowledge: Machine Learning, Modeling and Indexing Multimedia Contents and Symbolic Data", Master in Computer Science, University of Rennes 1, France [F. Coste].

10.2.2. Course responsibilities

- "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]
- "Big data and Semantic Web", Master 2 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]
- "e-Santé et réseaux hospitaliers", last year in engineering school ESIR, Univ. Rennes 1, [O. Dameron]
- "Equilibre Dynamique de la communication Cellulaire" Master 2 in Sciences cellulaire et Moléculaire du Vivant, Univ. Rennes 1 [N. Theret]

10.2.3. Teaching

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

Licence: C. Belleannée, Algorithmique et Programmation Fonctionnelle, 60, L1 informatique, Univ. Rennes1, France.

Licence: J. Coquet, Module Programmation Scientifique 1, 20h, L1 informatique, Rennes1, France.

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

---

[3] *http://www.irisa.fr/en/page/crosscutting-axis*

Licence: O. Dameron, C2i niveau 2, 2.5h, 2nd year school of medicine, Univ. Rennes 1, France

Licence: O. Dameron, Bioinformatique pour la génomique, 5h, 2nd year school of medicine, Univ. Rennes 1, France

Licence: C. Frioux, Programmation scientifique Python, 12h, L1, Rennes1, France.

Licence: C. Frioux, LaTeX, 12h, L3 ENSAI, France.

Licence: C. Frioux, Outils bureautiques pour le statisticien , 6h, L3 ENSAI, France.

Licence: C. Frioux, Algorithmique et programmation Python, 6h, L3 ENSAI, France.

Licence: L. Bourneuf, Ingénieurerie Systèmes et Réseaux, 10h, L3 INFO, France.

Licence: L. Bourneuf, Algorithmique des graphes, 8h, L3 INFO, France.

Licence: L. Bourneuf, Algorithmique des graphes, 2h, L3 MIAGE, France.

Master: L. Bourneuf, Principes de Programmation et d'Algorithmique, 6h, M1 BIG, France.

Master: L. Bourneuf, Projet, 10h, M1 BIG, 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: F. Coste, Apprentissage Automatique Supervisé, 10h, M2 Informatique, Univ. Rennes 1, France

Master: O. Dameron, Object-oriented programing, 20h, M1 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Gestion de projet en informatique, 12h, M1 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Ontologies biomédicales, 6h, Engineering school Institut Mines-Télécom Bretagne-Atlantique Brest, France

Master: O. Dameron, Internship jury, 25h, M1 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Internship jury, 7.5h, M2 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Intégration : remise à niveau en informatique, 14h, M1 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Programmation impérative en Python, 39.5h, M1 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Système informatique GNU/Linux, 12h, M1 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Programmation en Python, 24h, M1 in Public Health, Univ. Rennes 1 [O. Dameron]

Master: O. Dameron, Semantic Web and bio-ontologies, 14h, M2 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Bases de mathématiques et probabilités, 15h, M1 santé publique, Univ. Rennes 1, France

Master: A. Siegel, Integrative and Systems biology, 20h, M2, Univ. Rennes 1, France

Master: A. Siegel, Introduction to integrative biology, 2h, M2, Univ. Rennes 1, France

10.2.4. Supervision

PhD : Victorien Delannée, *Intégrer les échelles moléculaires et cellulaires dans l'inférence de réseaux métaboliques. Application aux xénobiotiques.*, started in Oct. 2014, defended in Nov. 2017, supervised by A. Siegel and N. Théret [14].

PhD : Julie Laniau, *Structure de réseaux biologiques : rôle des noeuds internes vis-à-vis de la production de composés*, started in Oct. 2013, defended in Oct. 2017, supervised by A. Siegel and D. Eveillard. [15]

PhD : Jean Coquet, *Semantic-based reasoning for biological pathways analysis*, started in Oct. 2014, defended in Dec. 2017, supervised by O. Dameron and N. Théret [12]/

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

PhD in progress : Clémence Frioux, *Using preferences in Answer Set Programming to decipher interactions within the species of an ecosystem at the genomic scale*, started in Oct. 2015, supervised by A. Siegel.

PhD in progress : Yann Rivault, *Analyse de parcours de soins à partir de bases de données médico-administratives en utilisant des outils du Web Sémantique: identification de complications et de leurs déterminants suite à la pose chirurgicale de dispositif médical implantable en ambulatoire* , started in Oct. 2015, supervised by O. Dameron and N. Lemeur.

PhD in progress : Juliette 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.

10.2.5. Juries

- *Member of Ph-D thesis juries*. J. Mercier, Univ. Evry/CEA [A. Siegel, reviewer]. C. Franay, INRA Toulouse [A. Siegel, reviewer]. W Bedhiafi, Univ. Pierre et Marie Curie Paris + UTM Tunis [O. Dameron]. V. Delannée, Univ. Rennes 1 [N. Theret, O. Dameron]. J. Coquet, Univ. Rennes 1 [O. Dameron, N. Theret]. P. Finet, Univ. Rennes 1 [O. Dameron]

- *Member of habilitation thesis juries*. E. Remy, Univ. Marseille [A. Siegel, president].

- *Member of medicine doctorate juries* G. Lebailly, Univ. Rennes 1 [O. Dameron].

10.2.6. Internships

- Internship, from Jun 2017 until Jul 2017. Supervised by J. Nicolas. Student: Alexis Baudin. Subject: Recherche d'attracteurs dans les réseaux booléens synchrones en ASP.

- Internship, from Jan until Jun 2017. Supervised by A. Siegel. Student: Mael Conan. Subject: Modélisation et caractérisation de la réponse au stress de la cyanobactérie marine Synechococcus sp. WH7803.

- Internship, from from Apr 2017 until Jul 2017. Supervised by Nathalie Théret and Olivier Dameron. Student: Kevin Courtet. Subject: Integration of genic regulatory interaction network by miRNAs from patients' macrophages with cystic fibrosis.

- Internship, from Apr 2017 until Jul 2017. Supervised by J. Got. Student: Nicolas Guillaudeux Subject: Vérifications du réseau métabolique entier de Tisochrysis lutea.

- Internship, from Apr 2017 until Jul 2017. Supervised by J. Nicolas and F. Coste. Student: AliHassan Kachalo Subject: Annotation automatique en familles des séquences d'une superfamille d'enzymes, les HAD (haloacides déhalogénases) par Analyse de Concepts Formels (FCA).

- Internship, from from Apr 2017 until Jul 2017. Supervised by C. Frioux and C. Trottier. Student: Claire Lippold. Subject: Exploration et caractérisation du microbiome associé à Ectocarpus subulatus str. BFT.

- Internship, from Jan 2017 until Jun 2017. Supervised by O. Dameron and A. Siegel. Student: Marine Louarn. Subject: Analysis and integration of heterogeneous large-scale genomics data.

- Internship, from Apr 2017 until Jul 2017. Supervised by . Théret and J. Nicolas. Student: Aurelie Nicolas. Subject: Modeling of interaction networks from extracellular matrix components using formal concept analyses.

- Internship, from Apr 2017 until Jul 2017. Supervised by C. Belleannée. Student: Dimitri Pedron. Subject: Annotation et prédiction de transcriptome: validation d'ORF alternatifs prédits. Application au gène CREM chez l'humain, la souris et le chien.

- Internship, from Feb 2017 until Jun 2017. Supervised by F. Coste. Student: Manon Ruffini, Subject: Better scoring schemes for the recognition of functional proteins by protomata

- Internship, from Feb 2017 until Jun 2017. Supervised by J. Nicolas. Student: MarieSalmon, Subject: Biclustering: quantitative formal concept analysis in Answer Set Programming

- Internship, from Jan 2017 until Jun 2017. Supervised by A. Siegel and O. Dameron. Student: Meline Wery Subject: Formalizing and computing signatures of phenotypes within a biological network

- Internship, from May 2017 until Jul 2017. Supervised by C. Belleannée. Student: Mohamed Zemmouri Subject: Analyse de texte en bioinformatique : Modélisation grammaticale d'un site ADN, et recherche du site, même dégénéré, sur l'intégralité du génome humain

10.3. Popularization

10.3.1. *General audience paper*

We have written a contribution to the collaborative book edited by CNRS on the main issues of data-mining. Our contribution was specifically related to modeling issues arising in ecology with the development of NGS technologies [32].

10.3.2. *Science en Cour[t]s*

(*http://sciences-en-courts.fr/*) Many of our on-going and former Ph-D students (A. Antoine-Lorquin, C. Bettembourg, J. Coquet, V. Delannée, G. Garet, S. Prigent) have been heavily involved in 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 [4] In this context, the following scientific film has been created by the team members in 2017 *Helene et les cartons* [5].

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

---

[4]*https://www.youtube.com/embed/3DRgLLITKUc*
[5]*https://www.youtube.com/embed/yheO8Y0nWu0*

[2] J. BOURDON, D. EVEILLARD, A. SIEGEL. *Integrating quantitative knowledge into a qualitative gene regulatory network*, in "PLoS Computational Biology", September 2011, vol. 7, nᵒ 9 [*DOI :* 10.1371/JOURNAL.PCBI.1002157], http://hal.archives-ouvertes.fr/hal-00626708

[3] 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

[4] 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, vol. 41 [*DOI :* 10.1093/NAR/GKS1091], http://hal.inria.fr/hal-00760946

[5] F. COSTE, G. KERBELLEC. *A Similar Fragments Merging Approach to Learn Automata on Proteins*, in "ECML:Machine Learning: ECML 2005, 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005, Proceedings", J. GAMA, R. CAMACHO, P. BRAZDIL, A. JORGE, L. TORGO (editors), Lecture Notes in Computer Science, Springer, 2005, vol. 3720, pp. 522-529

[6] M. GEBSER, C. GUZIOLOWSKI, M. IVANCHEV, T. SCHAUB, A. SIEGEL, P. VEBER, S. THIELE. *Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming*, in "Principles of Knowledge Representation and Reasoning", AAAI Press, 2010

[7] C. GUZIOLOWSKI, S. VIDELA, F. EDUATI, S. THIELE, T. COKELAER, A. SIEGEL, J. SAEZ-RODRIGUEZ. *Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming*, in "Bioinformatics", August 2013, vol. 29, nᵒ 18, pp. 2320-2326 [*DOI :* 10.1093/BIOINFORMATICS/BTT393], http://hal.inria.fr/hal-00853704

[8] 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

[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, nᵒ 1, 32 p. [*DOI :* 10.1371/JOURNAL.PCBI.1005276], https://hal.inria.fr/hal-01449100

[10] C. ROUSSEAU, M. GONNET, M. LE ROMANCER, J. NICOLAS. *CRISPI: a CRISPR interactive database*, in "Bioinformatics", 2009, vol. 25, nᵒ 24, pp. 3317-3318

[11] 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

[12] J. COQUET. *Comprehensive study of large signaling pathways by clustering trajectories and characterization by semantic analysis*, Université de Rennes 1, December 2017, https://hal.inria.fr/tel-01670730

[13] V. DELANNÉE. *Integrate molecular and cellular scales in the inference of metabolic networks : application to xenobiotics*, Université Rennes 1, November 2017, https://tel.archives-ouvertes.fr/tel-01690005

[14] V. DELANNÉE. *Integrate molecular and cellular scales in the inference of metabolic networks. Application to xenobiotics*, Univ. Rennes 1, November 2017, https://hal-univ-rennes1.archives-ouvertes.fr/tel-01659375

[15] J. LANIAU. *Biological networks structure : role of internal nodes regarding production of target elements*, Inria Rennes - Bretagne Atlantique, October 2017, https://hal.archives-ouvertes.fr/tel-01656474

Articles in International Peer-Reviewed Journals

[16] 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

[17] V. DELANNÉE, S. LANGOUËT, N. THÉRET, A. SIEGEL. *A modeling approach to evaluate the balance between bioactivation and detoxification of MeIQx in human hepatocytes*, in "PeerJ", 2017, vol. 5, e3703 p. [*DOI :* 10.7717/PEERJ.3703], https://hal.inria.fr/hal-01575579

[18] F. GONDRET, A. VINCENT, M. HOUÉE-BIGOT, A. SIEGEL, S. LAGARRIGUE, D. CAUSEUR, H. H. GILBERT, I. LOUVEAU. *A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs*, in "BMC Genomics", 2017, vol. 18, n° 1, 244 p. [*DOI :* 10.1186/S12864-017-3639-0], https://hal.archives-ouvertes.fr/hal-01494107

[19] J. LANIAU, C. FRIOUX, J. NICOLAS, C. BAROUKH, M. P. CORTÉS, J. GOT, C. TROTTIER, D. EVEILLARD, A. SIEGEL. *Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks*, in "PeerJ", 2017, vol. 5, e3860 p. [*DOI :* 10.7717/PEERJ.3860], https://hal.archives-ouvertes.fr/hal-01635688

[20] V. PICARD, A. SIEGEL, J. BOURDON. *A Logic for Checking the Probabilistic Steady-State Properties of Reaction Networks*, in "Journal of computational biology : a journal of computational molecular cell biology", August 2017, vol. 24, n° 8, pp. 1–12 [*DOI :* 10.1089/CMB.2017.0099], https://hal.archives-ouvertes.fr/hal-01552190

[21] 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

[22] 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

Invited Conferences

[23] V. WUCHER, F. LEGEAI, L. BOURNEUF, T. DERRIEN, A. GALLOT, S. HUDAVERDIAN, S. JAUBERT-POSSAMAI, N. LETERME-PRUNIER, J. NICOLAS, H. SEITZ, A. SIEGEL, S. TANGUY, G. LE TRIONNAIRE, D. TAGU. *Integrative genomics and gene networks for studying phenotypic plasticity in the pea aphid*, in "10th Arthropod Genomics Symposium", Notre Dame, United States, June 2017, https://hal.inria.fr/hal-01566438

International Conferences with Proceedings

[24] *Best Paper*

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.

[25] 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, France, Lecture Notes in Computer Sciences, September 2017, 17 p. , https://hal.archives-ouvertes.fr/hal-01559249

[26] *Best Paper*

C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO. *Hybrid Metabolic Network Completion*, in "14th International Conference on Logic Programming and Nonmonotonic Reasoning - LPNMR 2017", Espoo, Finland, M. BALDUCCINI, T. JANHUNEN (editors), Logic Programming and Nonmonotonic Reasoning, Springer, July 2017, vol. 10377, pp. 308-321, https://hal.inria.fr/hal-01557347.

National Conferences with Proceedings

[27] P. FINET, B. GIBAUD, O. DAMERON, R. LE BOUQUIN JEANNES. *Interoperable infrastructure and implementation of a health data model for remote monitoring of chronic diseases with comorbidities*, in "Journées d'Etude sur la TéléSANté, 6ème edition", Bourges, France, Pôle Capteurs, Université d'Orléans, May 2017, https://hal.archives-ouvertes.fr/hal-01565009

Conferences without Proceedings

[28] X. GARNIER, A. BRETAUDEAU, O. FILANGI, F. LEGEAI, A. SIEGEL, O. DAMERON. *AskOmics, a web tool to integrate and query biological data using semantic web technologies*, in "JOBIM 2017 - Journées Ouvertes en Biologie, Informatique et Mathématiques", Lille, France, July 2017, 1 p. , https://hal.inria.fr/hal-01577425

[29] X. GARNIER, O. DAMERON, O. FILANGI, F. LEGEAI, A. BRETAUDEAU. *Integration of Linked Data into Galaxy using Askomics* , in "Galaxy Community Conference", Montpellier, France, June 2017, https://hal.inria.fr/hal-01576870

[30] C. LIPPOLD, C. TROTTIER, C. FRIOUX, M. CHEVALLIER, S. M. DITTAMI, A. SIEGEL. *Toward the study of metabolic functions in algal holobionts*, in "4ème Colloque de Génomique Environnementale", Marseille, France, September 2017, https://hal.inria.fr/hal-01576484

[31] V. POTDEVIN, M. CHEVALLIER, R. GARRIDO OTER, A. SIEGEL, S. HACQUARD, D. EVEILLARD, P. VANDENKOORNHUYSE. *Toward a better understanding of the plant microbiota*, in "4ème colloque de

Génomique Environnementale", Marseille, France, GDR Génomique Environnementale, September 2017, https://hal.inria.fr/hal-01576743

Scientific Books (or Scientific Book chapters)

[32] D. EVEILLARD, A. SIEGEL, P. VANDENKOORNHUYSE. *L'écologie des systèmes*, in "Les Big Data à Découvert", M. BOUZEGHOUB, R. MOSSERI (editors), CNRS éditions, 2017, https://hal.inria.fr/hal-01575603

Other Publications

[33] A. ANTOINE-LORQUIN, A. ARNAOTY, S. ASGARI, M. BATAILLER, L. BEAUCLAIR, C. BELLEANNÉE, S. BIRE, N. BUISINE, V. COUSTHAM, A. GIRAULT, S. GUYETANT, T. LECOMTE, B. PIÉGU, B. PITARD, I. STÉVANT, Y. BIGOT. *First landscape of binding to chromosomes for a domesticated mariner transposase in the human genome: diversity of genomic targets of SETMAR isoforms in two colorectal cell lines*, July 2017, working paper or preprint [*DOI :* 10.1101/115030], https://hal.archives-ouvertes.fr/hal-01558001

[34] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE. *Detection of mutated primers and impact on targeted metagenomics results*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01564093

[35] C. MARCHET, L. LECOMPTE, C. DA SILVA, C. CRUAUD, J.-M. AURY, J. NICOLAS, P. PETERLONGO. *De novo Clustering of Gene Expressed Variants in Transcriptomic Long Reads Data Sets*, November 2017, working paper or preprint [*DOI :* 10.1101/170035], https://hal.archives-ouvertes.fr/hal-01643156

[36] M. RUFFINI. *Better scoring schemes for the recognition of functional proteins by protomata*, Rennes 1, June 2017, https://hal.inria.fr/hal-01557941

References in notes

[37] O. ABDOU-ARBI, S. LEMOSQUET, J. VAN MILGEN, A. SIEGEL, J. BOURDON. *Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs*, in "BMC Systems Biology", 2014, vol. 8, n⁰ 1, 8 p. [*DOI :* 10.1186/1752-0509-8-8], https://hal.inria.fr/hal-00947219

[38] V. ACUÑA, A. ARAVENA, C. GUZIOLOWSKI, D. EVEILLARD, A. SIEGEL, A. MAASS. *Deciphering transcriptional regulations coordinating the response to environmental changes*, in "BMC Bioinformatics", January 2016, vol. 17, n⁰ 1, pp. 129-42, https://hal.archives-ouvertes.fr/hal-01260866

[39] V. ACUÑA, A. ARAVENA, A. MAASS, A. SIEGEL. *Modeling parsimonious putative regulatory networks: complexity and heuristic approach*, in "15th conference in Verification, Model Checking, and Abstract Interpretation", San Diego, United States, Springer, 2014, vol. 8318, pp. 322-336 [*DOI :* 10.1007/978-3-642-54013-4_18], https://hal.inria.fr/hal-00926477

[40] G. ANDRIEUX, M. LE BORGNE, N. THÉRET. *An integrative modeling framework reveals plasticity of TGF-Beta signaling*, in "BMC Systems Biology", 2014, vol. 8, n⁰ 1, 30 p. [*DOI :* 10.1186/1752-0509-8-30], http://www.hal.inserm.fr/inserm-00978313

[41] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE. *Detection of mutated primers and impact on targeted metagenomics results*, in "RCAM'16 "Recent Computational Advances in Metagenomics"", The Hague, Netherlands, September 2016, https://hal.inria.fr/hal-01576304

[42] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE. *Impact de la recherche d'amorces mutées sur les résultats d'analyses métagénomiques*, in "Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM)", Rennes, France, Jobim-2016, Société Française de Bioinformatique (SFBI), June 2016, https://hal.archives-ouvertes.fr/hal-01343121

[43] T. L. BAILEY, M. BODEN, F. A. BUSKE, M. FRITH, C. E. GRANT, L. CLEMENTI, J. REN, W. W. LI, W. S. NOBLE. *MEME Suite: tools for motif discovery and searching*, in "Nucleic Acids Research", 2009, vol. 37, n° suppl_2, W202 p. , http://dx.doi.org/10.1093/nar/gkp335

[44] T. L. BAILEY, J. JOHNSON, C. E. GRANT, W. S. NOBLE. *The MEME Suite*, in "Nucleic Acids Research", 2015, vol. 43, n° W1, pp. W39-W49, http://dx.doi.org/10.1093/nar/gkv416

[45] F. BALUSSON, M.-A. BOTREL, O. DAMERON, Y. DAUXAIS, E. DREZEN, A. DUPUY, T. GUYET, D. GROSS-AMBLARD, A. HAPPE, N. LE MEUR, B. LE NAUTOUT, E. LERAY, E. NOWAK, C. RAULT, E. OGER, E. POLARD. *PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases*, in "International Congress on e-Health Research", Paris, France, October 2016, https://hal.inria.fr/hal-01380939

[46] C. BARAL. *Knowledge Representation, Reasoning and Declarative Problem Solving*, Cambridge University Press, 2010

[47] T. BAUMURATOVA, D. SURDEZ, B. DELYON, G. STOLL, O. DELATTRE, O. RADULESCU, A. SIEGEL. *Localizing potentially active post-transcriptional regulations in the Ewing's sarcoma gene regulatory network*, in "BMC Systems Biology", 2010, vol. 4, n° 1, 146 p. [*DOI :* 10.1186/1752-0509-4-146], http://www.hal.inserm.fr/inserm-00984711

[48] R. BELLAZZI. *Big Data and Biomedical Informatics: A Challenging Opportunity*, in "Yearbook of medical informatics", 2014, vol. 9, n° 1, In press

[49] R. BELLAZZI, M. DIOMIDOUS, I. N. SARKAR, K. TAKABAYASHI, A. ZIEGLER, A. T. MCCRAY. *Data analysis and data mining: current issues in biomedical informatics*, in "Methods of information in medicine", 2011, vol. 50, n° 6, pp. 536–544

[50] 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

[51] R. BELLÉ, S. PRIGENT, A. SIEGEL, P. CORMIER. *Model of cap-dependent translation initiation in sea urchin: a step towards the eukaryotic translation regulation network*, in "Molecular Reproduction and Development", 2010, vol. 77, n° 3, pp. 257-64

[52] R. BENDAOUD, A. NAPOLI, Y. TOUSSAINT. *Formal Concept Analysis: A Unified Framework for Building and Refining Ontologies*, in "Knowledge Engineering: Practice and Patterns, 16th International Conference, EKAW 2008, Acitrezza, Italy, September 29 - October 2, 2008. Proceedings", A. GANGEMI, J. EUZENAT (editors), Lecture Notes in Computer Science, Springer, 2008, vol. 5268, pp. 156-171, http://dx.doi.org/10.1007/978-3-540-87696-0_16

[53] C. BETTEMBOURG, O. DAMERON, A. BRETAUDEAU, F. LEGEAI. *AskOmics : Intégration et interrogation de réseaux de régulation génomique et post-génomique*, in "IN OVIVE (INtégration de sources/masses de données hétérogènes et Ontologies, dans le domaine des sciences du VIVant et de lâEnvironnement)", Rennes, France, June 2015, 7 p. , https://hal.inria.fr/hal-01184903

[54] C. BETTEMBOURG, C. DIOT, O. DAMERON. *Semantic particularity measure for functional characterization of gene sets using gene ontology*, in "PLoS ONE", 2014, vol. 9, n° 1, e86525 [*DOI :* 10.1371/JOURNAL.PONE.0086525], https://hal.inria.fr/hal-00941850

[55] C. BETTEMBOURG, C. DIOT, O. DAMERON. *Optimal threshold determination for interpreting semantic similarity and particularity*, in "PLoS ONE", 2015, vol. 10, n° 7, e0133579 [*DOI :* 10.1371/JOURNAL.PONE.0133579], https://hal.archives-ouvertes.fr/hal-01207763

[56] 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

[57] C. BIZER, T. HEATH, T. BERNERS LEE. *Linked Data–The story so far*, in "International Journal on Semantic Web and Information Systems", 2009, vol. 5, n° 3, pp. 1–22

[58] J. A. BLAKE, C. J. BULT. *Beyond the data deluge: Data integration and bio-ontologies*, in "Journal of Biomedical Informatics", 2006, vol. 39, n° 3, pp. 314–320

[59] 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

[60] P. BORDRON, D. EVEILLARD, A. MAASS, A. SIEGEL. *An ASP application in integrative biology: identification of functional gene units*, in "LPNMR - 12th Conference on Logic Programming and Nonmonotonic Reasoning - 2013", Corunna, Spain, September 2013, http://hal.inria.fr/hal-00853762

[61] 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° 1, pp. 106-117 [*DOI :* 10.1002/MBO3.315], https://hal.inria.fr/hal-01246173

[62] 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

[63] H. CHEN, T. YU, J. Y. CHEN. *Semantic Web meets Integrative Biology: a survey*, in "Briefings in bioinformatics", 2012, vol. 14, n° 1, pp. 109–125

[64] M. CHEVALLIER, M. AITE, J. GOT, G. COLLET, N. LOIRA, M. P. CORTÉS, C. FRIOUX, J. LANIAU, C. TROTTIER, A. MAASS, A. SIEGEL. *Handling the heterogeneity of genomic and metabolic networks data*

within flexible workflows with the PADMet toolbox, in "Jobim 2016 : 17ème Journées Ouvertes en Biologie, Informatique et Mathématiques", Lyon, France, June 2016, https://hal.inria.fr/hal-01377844

[65] G. COLLET, D. EVEILLARD, M. GEBSER, S. PRIGENT, T. SCHAUB, A. SIEGEL, S. THIELE. *Extending the Metabolic Network of Ectocarpus Siliculosus using Answer Set Programming*, in "LPNMR - 12th Conference on Logic Programming and Nonmonotonic Reasoning - 2013", Corunna, Spain, September 2013, http://hal.inria.fr/hal-00853752

[66] F. COSTE. *Learning the Language of Biological Sequences*, in "Topics in Grammatical Inference", J. HEINZ, J. M. SEMPERE (editors), Springer-Verlag, 2016 [*DOI :* 10.1007/978-3-662-48395-4_8], https://hal.inria.fr/hal-01244770

[67] 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

[68] S. DAMINELLI, V. J. HAUPT, M. REIMANN, M. SCHROEDER. *Drug repositioning through incomplete bicliques in an integrated drug-target-disease network*, in "Integr. Biol.", 2012, vol. 4, pp. 778-788, http://dx.doi.org/10.1039/C2IB00154C

[69] 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, 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", 2014, vol. 5, 241 p. [*DOI :* 10.3389/FGENE.2014.00241], https://hal.inria.fr/hal-01079739

[70] A. EVRARD, C. BETTEMBOURG, M. JUBAULT, O. DAMERON, O. FILANGI, A. BRETAUDEAU, F. F. LEGEAI. *Integration and query of biological datasets with Semantic Web technologies: AskOmics*, June 2016, Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM 2016), Poster, https://hal.inria.fr/hal-01391087

[71] R. D. FINN, J. CLEMENTS, W. ARNDT, B. L. MILLER, T. J. WHEELER, F. SCHREIBER, A. BATEMAN, S. R. EDDY. *HMMER web server: 2015 update*, in "Nucleic Acids Research", 2015, vol. 43, n⁰ W1, W30 p. , http://dx.doi.org/10.1093/nar/gkv397

[72] C. GALIEZ, M. CHRISTOPHE, F. COSTE, P. BALDI. *VIRALpro: a tool to identify viral capsid and tail sequences*, in "Bioinformatics", 2016 [*DOI :* 10.1093/BIOINFORMATICS/BTV727], https://hal.archives-ouvertes.fr/hal-01242251

[73] 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

[74] F. GONDRET, I. LOUVEAU, M. HOUEE, D. CAUSEUR, A. SIEGEL. *Data integration*, in "Meeting INRA-ISU", Ames, United States, March 2015, 11 p. , https://hal.archives-ouvertes.fr/hal-01210940

[75] F. GONDRET, A. VINCENT, M. HOUEE, S. LAGARRIGUE, A. SIEGEL, D. CAUSEUR, I. LOUVEAU. *Integrative responses of pig adipose tissues to high-fat high-fiber diet: towards key regulators of energy flexibility*, March 2015, ASAS/ADSA midwest meeting, Poster, https://hal.archives-ouvertes.fr/hal-01210925

[76] F. GONDRET, A. VINCENT, M. HOUÉE-BIGOT, A. SIEGEL, S. LAGARRIGUE, I. LOUVEAU, D. CAUSEUR. *Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location*, in "BMC Genomics", December 2016, vol. 17, n⁰ 1, 120 p. , article number 120 [*DOI :* 10.1186/S12864-016-2438-3], https://hal.archives-ouvertes.fr/hal-01286555

[77] C. GUZIOLOWSKI, A. BOURDÉ, F. MOREEWS, A. SIEGEL. *BioQuali Cytoscape plugin: analysing the global consistency of regulatory networks*, in "Bmc Genomics", 2009, vol. 26, n⁰ 10, 244 p. [*DOI :* 10.1186/1471-2164-10-244], http://hal.inria.fr/inria-00429804

[78] C. GUZIOLOWSKI, S. VIDELA, F. EDUATI, S. THIELE, T. COKELAER, A. S. SIEGEL, J. SAEZ-RODRIGUEZ. *Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming*, in "Bioinformatics", 2013, vol. 29, pp. 2320-2326 [*DOI :* 10.1093/BIOINFORMATICS/BTT393], https://hal.inria.fr/hal-00853704v1

[79] F. HERAULT, A. VINCENT, O. DAMERON, P. LE ROY, P. CHEREL, M. DAMON. *The longissimus and semimembranosus muscles display marked differences in their gene expression profiles in pig*, in "PLoS ONE", 2014, vol. 9, n⁰ 5, e96491 [*DOI :* 10.1371/JOURNAL.PONE.0096491], https://hal.inria.fr/hal-00989635

[80] R. KAMINSKI, A. SIEGEL, T. SCHAUB, S. VIDELA. *Minimal Intervention Strategies in Logical Signaling Networks with ASP*, in "Theory and Practice of Logic Programming", September 2013, vol. 13, n⁰ Special issue 4-5, pp. 675-690 [*DOI :* 10.1017/S1471068413000422], http://hal.inria.fr/hal-00853747

[81] H. KLARNER, A. STRECK, H. SIEBERT. *PyBoolNet: a python package for the generation, analysis and visualization of boolean networks*, in "Bioinformatics", 2017, vol. 33, n⁰ 5, pp. 770-772, http://dx.doi.org/10.1093/bioinformatics/btw682

[82] S. KURTZ. *The Vmatch large scale sequence analysis software*, in "Ref Type: Computer Program", 2003, vol. 412

[83] S. LAURENT, A. RICHARD, O. MULNER-LORILLON, J. MORALES, D. FLAMENT, V. GLIPPA, J. BOURDON, P. GOSSELIN, A. SIEGEL, P. CORMIER, R. BELLÉ. *Modelization of the regulation of protein synthesis following fertilization in sea urchin shows requirement of two processes: a destabilization of eIF4E:4E-BP complex and a great stimulation of the 4E-BP-degradation mechanism, both rapamycin-sensitive*, in "Frontiers in Genetics", 2014, vol. 5, 117 p. [*DOI :* 10.3389/FGENE.2014.00117], https://hal.inria.fr/hal-01079758

[84] F. LEGEAI, C. BETTEMBOURG, A. BRETAUDEAU, Y. CHAUSSIN, O. DAMERON, D. TAGU. *BIPAA/Askomics, a new and easy approach for querying genomics and epigenomics elements in interaction*, in "XXVth International Congress of Entomology 2016", Orlando, Florida, United States, September 2016, https://hal.inria.fr/hal-01391080

[85] F. LEGEAI, T. DERRIEN, V. WUCHER, D. AUDREY, G. LE TRIONNAIRE, D. TAGU. *Long non-coding RNA in the pea aphid; identification and comparative expression in sexual and asexual embryos*, in "Arthropod Genomics Symposium", Urbana, United States, June 2014, https://hal.inria.fr/hal-01091304

[86] S.-W. LEUNG, C. MELLISH, D. ROBERTSON. *Basic Gene Grammars and DNA-ChartParser for language processing of Escherichia coli promoter DNA sequences*, in "Bioinformatics", 2001, vol. 17, n⁰ 3, pp. 226-236 [*DOI :* 10.1093/BIOINFORMATICS/17.3.226], http://bioinformatics.oxfordjournals.org/content/17/3/226.abstract

[87] A. LIHONOSOVA, A. KAMINSKAYA. *Using Formal Concept Analysis for Finding the Closest Relatives among a Group of Organisms*, in "Procedia Computer Science", 2014, vol. 31, n[o] Complete, pp. 860-868, http://dx. doi.org/10.1016/j.procs.2014.05.337

[88] K. M. LIVINGSTON, M. BADA, W. A. BAUMGARTNER, L. E. HUNTER. *KaBOB: ontology-based semantic integration of biomedical databases*, in "BMC bioinformatics", 2015, vol. 16, 126 p.

[89] M. MARTIN. *Cutadapt removes adapter sequences from high-throughput sequencing reads*, in "EMB-net.journal", 2011, vol. 17, n[o] 1, 10 p. , http://dx.doi.org/10.14806/ej.17.1.200

[90] N. MESSAI, M.-D. DEVIGNES, A. NAPOLI, M. SMAIL-TABBONE. *Many-Valued Concept Lattices for Conceptual Clustering and Information Retrieval*, in "Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence", Amsterdam, The Netherlands, The Netherlands, IOS Press, 2008, pp. 127–131, http://dl.acm.org/citation.cfm?id=1567281.1567313

[91] P. T. MONTEIRO, C. CHAOUIYA. *Efficient Verification for Logical Models of Regulatory Networks*, M. P. ROCHA, N. LUSCOMBE, F. FDEZ-RIVEROLA, J. M. C. RODRÍGUEZ (editors), Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 259–267, https://doi.org/10.1007/978-3-642-28839-5_30

[92] J. NICOLAS, P. DURAND, G. RANCHY, S. TEMPEL, A.-S. VALIN. *Suffix-Tree Analyser (STAN): looking for nucleotidic and peptidic patterns in genomes*, in "Bioinformatics (Oxford, England)", 2005, vol. 21(24), pp. 4408-4410, https://hal.archives-ouvertes.fr/hal-00015234

[93] M. OBITKO, V. SNÁSEL, J. SMID. *Ontology Design with Formal Concept Analysis*, in "Proceedings of the CLA 2004 International Workshop on Concept Lattices and their Applications, Ostrava, Czech Republic, September 23-24, 2004", V. SNÁSEL, R. BELOHLÁVEK (editors), CEUR Workshop Proceedings, CEUR-WS.org, 2004, vol. 110, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-110/paper12. pdf

[94] M. OSTROWSKI, L. PAULEVÉ, T. SCHAUB, A. SIEGEL, C. GUZIOLOWSKI. *Boolean Network Identification from Perturbation Time Series Data combining Dynamics Abstraction and Logic Programming*, in "BioSystems", 2016 [*DOI :* 10.1016/J.BIOSYSTEMS.2016.07.009], https://hal.archives-ouvertes.fr/hal-01354075

[95] V. PICARD, O. MULNER-LORILLON, J. BOURDON, J. MORALES, P. CORMIER, A. SIEGEL, R. BELLÉ. *Model of the delayed translation of cyclin B maternal mRNA after sea urchin fertilization*, in "Molecular Reproduction and Development", 2016 [*DOI :* 10.1002/MRD.22746], http://hal.upmc.fr/hal-01390047

[96] P.-F. PLUCHON, T. FOUQUEAU, C. CREZE, S. LAURENT, J. BRIFFOTAUX, G. HOGREL, A. PALUD, G. HENNEKE, A. GODFROY, W. HAUSNER, M. THOMM, J. NICOLAS, D. FLAMENT. *An Extended Network of Genomic Maintenance in the Archaeon Pyrococcus abyssi Highlights Unexpected Associations between Eucaryotic Homologs*, in "PLoS ONE", 2013, vol. 8, n[o] 11, e79707 [*DOI :* 10.1371/JOURNAL.PONE.0079707], http://hal.inria.fr/hal-00911795

[97] S. PRIGENT, G. COLLET, S. M. DITTAMI, L. DELAGE, F. ETHIS DE CORNY, O. DAMERON, D. EVEIL-LARD, 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

[98] Y. RIVAULT, N. LE MEUR, O. DAMERON. *Complications post-opératoires et mode de prise en charge en angioplastie : apport du Programme de Médicalisation des Systèmes d'Information (PMSI)*, in "Congrès Adelf-Epiter 2016", Rennes, France, Congrès Adelf-Epiter 2016, September 2016, https://hal.inria.fr/hal-01389688

[99] Y. RIVAULT, N. LE MEUR, O. DAMERON. *La gestion de données médico-administratives grâce aux outils du Web sémantique*, in "Journées ADELF-EMOIS " Système d'information hospitalier et Epidémiologie "", Dijon, France, Revue d'Épidémiologie et de Santé Publique, March 2016, vol. 64, n⁰ Supplement 1, S15 p. [*DOI :* 10.1016/J.RESPE.2016.01.051], https://hal.archives-ouvertes.fr/hal-01297389

[100] A. ROSIER, P. MABO, L. TEMAL, P. VAN HILLE, O. DAMERON, L. DELÉGER, C. GROUIN, P. ZWEIGENBAUM, J. JACQUES, E. CHAZARD, L. LAPORTE, C. HENRY, A. BURGUN. *Personalized and automated remote monitoring of atrial fibrillation*, in "EP-Europace", 2016, vol. 18, n⁰ 3, pp. 347-352 [*DOI :* 10.1093/EUROPACE/EUV234], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01331019

[101] A. ROSIER, P. MABO, L. TEMAL, P. VAN HILLE, O. DAMERON, L. DELÉGER, C. GROUIN, P. ZWEIGENBAUM, J. JACQUES, E. CHAZARD, L. LAPORTE, C. HENRY, A. BURGUN. *Remote Monitoring of Cardiac Implantable Devices: Ontology Driven Classification of the Alerts*, in "Studies in Health Technology and Informatics", 2016, vol. 221, pp. 59–63 [*DOI :* 10.3233/978-1-61499-633-0-59], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01319949

[102] M. ROUANE-HACENE, M. HUCHARD, A. NAPOLI, P. VALTCHEV. *Using Formal Concept Analysis for Discovering Knowledge Patterns*, in "CLA'10: 7th International Conference on Concept Lattices and Their Applications", Sevilla, Spain, M. KRYSZKIEWICZ, S. OBIEDKOV (editors), CEUR, University of Sevilla, October 2010, n⁰ 672, pp. 223-234, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00531802

[103] C. ROUSSEAU, M. GONNET, M. LE ROMANCER, N. NICOLAS. *CRISPI: a CRISPR Interactive database*, in "Bioinformatics", 2009, vol. 25, pp. 3317-3318 [*DOI :* 10.1093/BIOINFORMATICS/BTP586], https://hal.inria.fr/inria-00438512

[104] L. ROYER, M. REIMANN, B. ANDREOPOULOS, M. SCHROEDER. *Unraveling Protein Networks with Power Graph Analysis*, in "PLoS Comput Biol", 07 2008, vol. 4, n⁰ 7, e1000108, http://dx.plos.org/10.1371%2Fjournal.pcbi.1000108

[105] M. RUFFINI. *Better scoring schemes for the recognition of functional proteins by protomata*, Rennes 1, June 2017, https://hal.inria.fr/hal-01557941

[106] T. SCHAUB, A. SIEGEL, S. VIDELA. *Reasoning on the response of logical signaling networks with Answer Set Programming*, in "Logical Modeling of Biological Systems", Wiley Online Librairy, 2014, pp. 49-92 [*DOI :* 10.1002/9781119005223.CH2], https://hal.inria.fr/hal-01079762

[107] T. SCHAUB, S. THIELE. *Metabolic Network Expansion with Answer Set Programming*, in "ICLP 2009", LNCS, Springer, 2009, vol. 5649, pp. 312-326

[108] S. SCHULZ, L. BALKANYI, R. CORNET, O. BODENREIDER. *From Concept Representations to Ontologies: A Paradigm Shift in Health Informatics?*, in "Healthcare informatics research", 2013, vol. 19, n⁰ 4, pp. 235-242

[109] D. SEARLS. *The language of genes*, in "Nature", 2002, vol. 420, pp. 211-217

[110] D. SEARLS. *String Variable Grammar: A Logic Grammar Formalism for the Biological Language of DNA*, in "Journal of Logic Programming", 1995, vol. 24, n⁰ 1&2, pp. 73-102

[111] C. J. A. SIGRIST, E. DE CASTRO, L. CERUTTI, B. A. CUCHE, N. HULO, A. BRIDGE, L. BOUGUELERET, I. XENARIOS. *New and continuing developments at PROSITE*, in "Nucleic Acids Research", 2013, vol. 41, n⁰ D1, D344 p. , http://dx.doi.org/10.1093/nar/gks1067

[112] 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, n⁰ 7, e1002195 p.

[113] S. TEMPEL, C. ROUSSEAU, F. TAHI, J. NICOLAS. *ModuleOrganizer: detecting modules in families of transposable elements*, in "BMC Bioinformatics", 2010, vol. 11, 474 p. [*DOI :* 10.1186/1471-2105-11-474], http://hal.inria.fr/inria-00536742

[114] S. THIELE, J. SAEZ-RODRIGUEZ, L. CERONE, A. SIEGEL, C. GUZIOLOWSKI, S. KLAMT. *Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies*, in "BMC Bioinformatics", 2015, vol. 16, 345 p. [*DOI :* 10.1186/s12859-015-0733-7], https://hal.inria.fr/hal-01225228

[115] S. VIDELA, C. GUZIOLOWSKI, F. EDUATI, S. THIELE, M. GEBSER, J. NICOLAS, J. SAEZ-RODRIGUEZ, T. SCHAUB, A. SIEGEL. *Learning Boolean logic models of signaling networks with ASP*, in "Journal of Theoretical Computer Science (TCS)", September 2015, n⁰ 599, pp. 79-101 [*DOI :* 10.1016/J.TCS.2014.06.022], https://hal.inria.fr/hal-01058610

[116] S. VIDELA, I. KONOKOTINA, L. ALEXOPOULOS, J. SAEZ-RODRIGUEZ, T. SCHAUB, A. SIEGEL, C. GUZIOLOWSKI. *Designing experiments to discriminate families of logic models*, in "Frontiers in Bioengineering and Biotechnology", September 2015, 9 p. [*DOI :* 10.3389/FBIOE.2015.00131], https://hal.inria.fr/hal-01196178

[117] R. WILLE. *restructuring lattice theory: an approach based on hierarchies of concepts*, in "Proceedings of the 7th International Conference on Formal Concept Analysis", Berlin, Heidelberg, ICFCA '09, Springer-Verlag, 2009, pp. 314–339, http://dx.doi.org/10.1007/978-3-642-01815-2_23

[118] V. WUCHER, D. TAGU, J. NICOLAS. , B. LAUSEN, S. KROLAK-SCHWERDT, M. BÖHMER (editors) *Edge Selection in a Noisy Graph by Concept Analysis – Application to a Genomic Network*, Data Science, Learning by Latent Structures, and Knowledge Discovery, Springer, 2014, 550 p. , https://hal.inria.fr/hal-01093337

[119] V. WUCHER. *Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid*, Université Rennes 1, November 2014, https://tel.archives-ouvertes.fr/tel-01135870

[120] V. WUCHER. *Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid*, Universite Rennes 1, November 2014, https://hal.archives-ouvertes.fr/tel-01095967

[121] T. YAN, D. YOO, T. Z. BERARDINI, L. A. MUELLER, D. C. WEEMS, S. WENG, J. M. CHERRY, S. Y. RHEE. *PatMatch: a program for finding patterns in peptide and nucleotide sequences*, in "Nucleic Acids Research", 2005, vol. 33, n⁰ Web-Server-Issue, pp. 262–266, https://doi.org/10.1093/nar/gki368