INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Symbiose

Biological systems and models, bioinformatics and sequences

Rennes - Bretagne-Atlantique

Theme : Computational Biology and Bioinformatics

Table of contents

1. Team

Research Scientists

Dominique Lavenier [Team leader, Senior Researcher, Cnrs, HdR]
Jérémie Bourdon [Junior researcher, on leave from Univ. Nantes, until october 2010]
François Coste [Junior researcher, Inria]
Claire Lemaitre [Junior researcher, Inria, since oct. 2010]
Jacques Nicolas [Senior researcher, Inria, HdR]
Pierre Peterlongo [Junior researcher, Inria]
Anne Siegel [Senior researcher, since Oct. 2010, Cnrs, HdR]

Faculty Members

Rumen Andonov [Professor, Univ. Rennes 1, HdR]
Catherine Belleannée [Associate Professor, Univ. Rennes 1]
Michel Le Borgne [Associate Professor, Univ. Rennes 1]
Raoul Vorc'h [Associate Professor, Univ. Rennes 1]

External Collaborators

Ovidiu Radulescu [Professor, Univ. Montpellier 2, HdR]
Nathalie Theret [Research director, INSERM, Rennes, HdR]

Technical Staff

Olivier Collin [GENOUEST, permanent senior Research engineer, Cnrs]
Charles Deltel [GENOUEST, permanent senior Research engineer, Inria, 50% time]
Olivier Sallou [GENOUEST, permanent senior Research engineer, Univ. Rennes 1]
Fabrice Legeai [permanent Engineer, INRA, 20% time dedicated to the symbiose project]
François Moreews [permanent Engineer, INRA, 20% time dedicated to the symbiose project]
Anthony Bretaudeau [GENOUEST, non permament junior engineer, Ibisa grant]
Delphine Naquin [GENOUEST, non permament junior engineer, Région Bretagne grant]
Jonathan Piat [GENOUEST, non permament junior engineer, ANR grant (BioWIC)]
Aurélien Roult [GENOUEST, non permament junior engineer, Région Bretagne grant]
Romaric Sabas [GENOUEST, non permament junior engineer, INRIA grant (ADT BioMAJ)]
Ludmila Sarbu [GENOUEST, non permament junior engineer, Région Bretagne grant]
Andres Burgos [non permament junior engineer, ADT inria]
Alexandre Cornu [non permament junior engineer, ANR contract BioWic]
Nolwenn Le Meur [INSERM/Region, Action Biologie integrative]
Odile Rousselet [non permament junior engineer, ANR BioWic, until July 2010]
Pavel Senin [non permament junior engineer, ANR Lepidolf]

PhD Students

Oumarou Abdou-Arbi [MENRT, since oct. 2010]
Geoffroy Andrieux [MENRT, since oct. 2010]
Guillaume Chapuis [Région/ENS, since oct. 2010]
Rayan Chikhi [MENRT/ENS]
Guillaume Collet [MENRT, until july 2010]
Matthias Gallé [Inria/CORDI]
Nicolas Maillet [ANR Mappi, since nov. 2010]
Noel Malod-Dognin [MENRT]
Guillaume Rizk [MENRT]

Post-Doctoral Fellows

Pierre Blavy [INRA]
Pavlos Antoniou [ARC Alcovna, since october 2010]

Administrative Assistant

Marie-Noëlle Georgeault [Assistant, Inria]

2. Overall Objectives

2.1. A Bioinformatics Center

Symbiose is a bioinformatics research project. It focuses on methodological research at the interface between computer science and molecular biology. The Symbiose team gathers two entities: a research group and a technical platform, called GenOuest.

- The *research group* focuses on high performance computing for large-scale genomic data and modeling of large-scale biological systems. Research activities cover sequence comparison, Next Generation Sequence processing, comparative genomic, identification of genome structures, structural biology, dynamic systems, and gene regulation network.
- The *GenOuest platform* belongs to Biogenouest, the French west life science network. Since 2009, it also belongs to the IBiSA[1] network and is certified ISO 9001:2008. The platform coordinates the activities of the RENABI-GO[2] regional center [36], one of the six French bioinformatics resource centers, and offers different bioinformatics services: computing power, storage, databanks, development, training, etc.

Both entities tightly collaborate to offer a full technological, research and training support to the biological community. Research and technological development projects are conducted in collaboration with INRA, Inserm and CNRS biological teams from the full country.

This environment offers the opportunity to locally mix computer scientists with strong expertise in high performance computing and dynamical modeling, together with genomic research labs. In the competitive field of environment, we are concerned with the storage, analysis and interpretation of large-scale and multi-timescales datasets produced by other platforms and research teams, including –although not exclusively – the analysis of Next Generation Sequencing Data.

The Symbiose project addresses both the pragmatic needs of high throughput resource management and the longer-term needs of the development of original algorithms and applications through dedicated researches.

2.2. Highlights

- **Best score at the SHREC'10 competition** In 2010, we won the SHape REtrieval Contest (SHREC'10) in the Protein Model Classification Track organized within Eurographics. The A_purva software developed by the Symbiose team has been intensively used to successfully classify a set of unknown proteins based on their 3D shape.
- **GASSST software transfer to GenomeQuest company** GASSST is a NGS mapping software able to fastly map a huge number of reads over full genomes. The GenomeQuest company, from which a licence agreement has been signed, includes this original software to significantly speeding up NGS bioinformatics pipelines.
- **Identification of new miRNAs in the Pea Aphid genome** In collaboration with INRA, Bio3P lab, high computing treatments involving GPU hardware have been used to systematically identify secondary structures of miRNA, and to annotate the Pea Aphid genome. miRNA play a key role in the Pea Aphid reproduction process.
- **Systems biology approach on Ewing pediatric tumor** The Bioquali software analyzes and validates the consistency of regulatory networks with respect to a gene expression profile. In collaboration with Institut Curie, it has been successfully applied on cancer systems to investigate the regulation network underlying Ewing pediatric tumor.

[1] GIS IBiSA: Infrastructures en Biologie Sante et Agronomie
[2] RENABI: Réseau national des plates-formes de bioinformatique, GO: Grand Ouest

3. Scientific Foundations

3.1. Sequence and Structure Modeling

This track concerns the search for relevant (e.g. functional) spatial or logical structures in macromolecules, either with intent to model specific spatial structures (secondary and tertiary structures, disulfide bounds, ...) or general biological mechanisms (transposition, ...). In the framework of **language theory and combinatorial optimization**, we address various types of problems: design of grammatical models on biological sequences and machine learning of grammatical models from sequences; efficient filtering and model matching in data banks; protein structure prediction.

Corresponding disciplinary fields are language theory, algorithmic on words, machine learning, data analysis and combinatorial optimization.

3.2. System Biology

We address the question of constructing accurate models of biological systems with respect to available data and knowledge. The availability of high-throughput methods in molecular biology has led to a tremendous increase of measurable data along with resulting knowledge repositories, gathered on the web (e.g. KEGG,MetaCyc, RegulonDB). However, both measurements as well as biological networks are prone to incompleteness, heterogeneity, and mutual inconsistency, making it highly non-trivial to draw biologically meaningful conclusions in an automated way. Based on this statement, we develop methods for the analysis of large-scale biological networks which formalize various reasoning modes in order to highlight incomplete regions in a regulatory model and to point at network products that need to be activated or inactivated to globally explain the experimental data. We also consider small-scale biological systems for a fine understanding of conclusions that can be drawn on active pathways from available data, working on deducible properties rather than simulation.

Corresponding disciplinary fields are model checking, constraint-based analysis and dynamical systems.

3.3. High Performance Computing

HPC for bioinformatics aims to bring efficient computing solutions in the two following challenging areas: processing of high throughput genomic data and processing of computational intensive algorithms. These two areas have in common to be highly time-consuming, but for different reasons. The first has to handle very huge amounts of data while the second has to solve very large optimization problems. More precisely, the first area required to organize, to structure, to index, and more generally, to manipulate very large data structures, making memory management issue a real challenge. The second area involves high complexity algorithms, but on a much more reduced data set.

In both cases, space and time limitations can be pushed away by the use of parallel machines as they can provide large aggregated memory space and/or high computational power. We believe that the design of parallel and optimized parallel algorithms is a key issue to face the avalanche of genomic data, and that all forms of parallelism must be exploited, from cloud computing to hardware accelerators such as GPGPU.

Corresponding disciplinary fields are optimization, parallelism, processing mass of data, hardware accelerator and advanced indexing structures.

4. Application Domains

4.1. Data and knowledge management

Multiple technologies are producing raw data that have to be cleared and assembled into meaningful observations. It is the realm of statistical studies, with sophisticated normalization procedures, most of them being

included in routine treatments. Information is produced in a highly distributed way, in each laboratory. Standardization, structuring of data banks, detection of redundancies and inconsistencies, integration of several sources of data and knowledge, extraction of knowledge from texts, all these are very crucial tasks for bioinformatics. High throughput techniques are also a source of algorithmic issues (assembling of fragments, design of probes).

4.2. Comparative genomics

Referring to a set of already known sequences is the most important method for studying new sequences, in the search for homologies. The basic issue is the alignment of a set of sequences, where one is looking for a global correspondence between positions of each sequence. A more complex issue consists in aligning structures. More macroscopic studies are also possible, involving more complex operations on genomes such as permutations. Genotyping studies consider Single Nucleotide Polymorphism data, which correspond to mutations observed at given positions in a sequence with respect to a population. Analyzing this type of data and relating them to phenotypic data leads to new research issues. Once sequences have been compared, phylogenies, that is, trees tracing back the evolution of genes, may be built from a set of induced distances.

4.3. From structural analysis to systems biology

This large domain aims at extracting biological knowledge from Xome studies, where X varies from genes to metabolites. Biological sequences and networks of components in the cell must verify a number of important constraints with respect to stable and accessible conformations, functional mechanisms and dynamics. These constraints result in the conservation during evolution of "patterns" and types of interactions to be deciphered. Many advanced researches consider now the study of life as a system, abstracted in a network of components governed by interaction laws, mostly qualitative or quantitative for reduced systems.

5. Software

5.1. Software tools for bioinformatics

Participants: Olivier Collin [**contact**], Olivier Sallou, Anthony Bretaudeau, Alexandre Cornu, Charles Deltel, Dominique Lavenier, François Moreews, Delphine Naquin, Aurélien Roult, Romaric Sabas, Ludmila Sarbu.

- **BioMAJ** BioMAJ is a workflow engine dedicated to data synchronization and processing. The Software automates the update cycle and the supervision of the locally mirrored databank repository. [Web site: http://biomaj.genouest.org]
- **GRISBI** The GRISBI project is aiming to set up a grid infrastructure devoted to the Bioinformatics community. This infrastructure is built upon the resources available on different bioinformatics facilities through gLite middleware. [Web site: http://www.grisbio.fr]
- **Mobylenet** In partnership with other bioinformatics platforms, GenOuest will set up a distributed network of bioinformatics resources built upon web portals based on the Mobyle platform. [Web site: http://mobylenet.rpbs.univ-paris-diderot.fr:8080/]
- **BioWIC** The BioWIC project aims to speed up both the design and the execution of bioinformatics workflows. The GenOuest platform is in charge of the development of an execution environment based on the scientific workflow environment Kepler. GenOuest is also installing different hardware resources (GPU, FPGA) in order to be able to host the different BioWIC software developed by all the partners. [Web site: http://biowic.inria.fr/]
- **DrMotifs** DrMotifs is a new software resources aiming at the integration of different software commonly used in pattern search and discovery. This resource will also integrate new software elaborated by the Symbiose team. [Web site: http://www.drmotifs.org] [Blog site: http://drmotifs.genouest.org]
- **NGS toolbox** In tight partnership with a sequencing facility, a complete NGS analysis environment has been set up. This environment includes all the essential softwares and many additional in-house tools have been developed around this set of softwares. [Web site: http://www.genouest.org]

5.2. Sequence alignement and comparison

Participants: Dominique Lavenier [**contact**], Claire Lemaitre, Pierre Peterlongo, Guillaume Rizk.

- **Cassis: rearrangement breakpoints**. Cassis is a software for precise detection of rearrangement breakpoints in whole (sequenced and assembled) genomes.
  web site: http://pbil.univ-lyon1.fr/software/Cassis/
- **GASSST: short reads mapper**. GASSST takes as input a huge number of reads comming from NGS machines and map them over full genomes.
  web site: http://www.irisa.fr/symbiose/projects/gassst/
- **PLAST: intensive bank sequence comparison**. PLAST is a parallel version of BLAST-like software targetting multipule parallel hardware such as FPGA accelerator or GPU boards.
  web site: http://www.irisa.fr/symbiose/projects/plast/
- **Tuiuiu: preliminary step before applying a multiple local aligner tool**. Tuiuiu removes from a sequence or from a set of sequences areas as large as possible that do not contain researched repeats.
  web site: http://mobyle.genouest.org/cgi-bin/Mobyle/portal.py?form=tuiuiu

5.3. Genome structure

Participants: Jacques Nicolas [**contact**], Dominique Lavenier [**contact**], Catherine Belleannée, Anthony Bretaudeau, Alexandre Cornu, Pierre Peterlongo, Guillaume Rizk, Olivier Sallou, Raoul Vorc'h.

- **GPU-UnaFold**. a parallel version of the UnaFold package to compute 2D structure of DNA/RNA sequences.
- **kisSnp: SNP identification**. a tool to find SNPs by comparing two sets of raw NGS reads without assembly nor mapping on a reference genome.
  web site: http://alcovna.genouest.org/kissnp/
- **Module organizer: segmentation of DNA sequences**. ModuleOrganizer is a software package proposing a synthetic view of a set of DNA sequences by providing both a segmentation of them into domains and a classification on the basis of these domains.
  web site: http://moduleorganizer.genouest.org
- **CRISPI: CRISPR identification**. CRISPI is a user-friendly web interface with many graphical tools and facilities allows extracting CRISPR, finding out CRISPR in personal sequences or calculating sequence similarity with spacers.
  web site: http://crispi.genouest.org
- **Logol**. Logol is a language and a tool to define biological patterns to look for in one or more sequences (dna/rna/proteins). Patterns can be complex: the tool allows the use of variables to look for repetitions for example, the use of gaps and morphisms (reverse word complement for example), etc.
  web site: http://www.genouest.org/spip.php?article758

5.4. Protein sequence and structure

Participants: Rumen Andonov [**contact**], Andres Burgos, François Coste, Guillaume Collet, Noel Malod-Dognin, Pavel Senin.

- **A_purva: Scoring similarities between proteins**. A_purva is a Contact Map Overlap maximization (CMO) solver. Given two protein structures represented by two contact maps, A_purva computes the amino-acid alignment which maximize the number of common contacts.
  web site: http://apurva.genouest.org
- **Protomata-Learner: fine characterization of protein families**. The tool infer graphical models (automata) to characterize a sample of (unaligned) sequences belonging to a structural or functional family of proteins.

web site: http://protomata-learner.genouest.org/

5.5. Systems biology

Participants: Anne Siegel [**contact**], Anthony Bretaudeau, Michel Le Borgne, François Moreews.

- **Bioquali: confront knowledge-based regulatory models with data**. Bioquali tests the consistency between an interaction graph and transcriptomic data. It outputs nodes in the network whose variation cannot be globally explained by the other available observations.
  web site:http://bioquali.genouest.org Cytoscape java web start

6. New Results

6.1. Annotation

Participants: Olivier Collin [**contact**], Fabrice Legeai, Olivier Sallou, Dominique Lavenier, Jacques Nicolas.

- **AphidBase**: This database is a comprehensive information system dedicated to aphids. It has been set up to safely centralize, manage, mine, disseminate and promulgate data generated by International Aphid Genomics Consortium (IAGC). [19]. [Online publication: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2009.00930.x/abstract]
- **Pea Aphid genome annotation**: Annotation and analysis of the pea aphid genome by a large international community. We were strongly involved into the database management, annotation protocols and gene curation [22]

6.2. Sequences alignement and comparison

Participants: Dominique Lavenier [**contact**], Jérémie Bourdon, Claire Lemaitre, Pierre Peterlongo, Raoul Vorc'h, Fabrice Legeai, Rayan Chikhi, Alexandre Cornu, Nicolas Maillet, Guillaume Rizk.

- **Genomic sequence comparison**: Comparing DNA or protein sequences remains a basic bioinformatics task. PLAST is a software parallelizing this treatment on various parallel supports: manycores, GPU, and FPGA accelerators. [41], [42] [Online publication: http://www.intechopen.com/books/show/title/parallel-and-distributed-computing] [Online publication: http://www.crcnetbase.com/isbn/978-1-4398-1488-8]
- **NGS mapping**: With the last progress of high performance sequencing machines, mapping short sequences (reads) over full genomes is becoming a challenging bioinformatics treatment. We have developed GASSST (Global Alignment Short Sequence Search Tool) a software able to map millions of short reads very rapidely [23]. [Online publication: http://bioinformatics.oxfordjournals.org/content/early/2010/08/24/bioinformatics.btq485.abstract]
- **Homology searches**: Book chapter dedicated to the use of advanced algorithmic techniques based on filtration and on the use of seeds for retrieving homologies with high specificity and sensitivity in large datasets [44].
- **Breakpoints in genomes**: We propose a new method to detect rearrangement breakpoints in a genome by comparison with the genome of a related species. The originality of the method lies in a second step where each detected breakpoint is refined by sequence alignments and statistical segmentation. This method improves the precision of breakpoint locations on genomes and enables to better characterize their sequences. [7]. [Online publication: http://bioinformatics.oxfordjournals.org/content/26/15/1897.long]
- **Indexing**: Factor and suffix oracles provide an economic and efficient solution for storing all the factors and suffixes respectively of a given text. We give an estimation of the average size for the dedicated examples of factor/suffix oracles [10]. [Online publication: http://www.sciencedirect.com/science/article/B758J-51962WF-1/2/cd6548b7641052c921582fa3ccbb99cd]

6.3. Genome Structure

Participants: Jacques Nicolas [**contact**], Catherine Belleannée, Jérémie Bourdon, François Coste, Pierre Peterlongo, Fabrice Legeai, Matthias Gallé.

- **Multiple repeats**: Tuiuiu is an algorithm designed for the fast filtration of full genomes in order to detect multiple repeats. It applies to a set of sequences as big as genomes a necessary condition that quickly remove large parts of sequences for which one is sure that they do not belong to the searched repeats. The searched repeats may then be found in the remaining sequences, often limited to a few percents of the initial input data. [29]. [Online publication: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5587026]
- **Inference of genomic sequences structure**: We address the problem of searching for the smallest grammar problem on large sequences - that is, finding a smallest context-free grammar that generates exactly one sequence. We use the concept of maximal repeats and propose a new algorithm which can be applied on whole genomes of model organisms and able to find up to 10% smaller grammars than state-of-the-art [28], [12], [26], [30]. [Online publication: http://www.springerlink.com/content/4220ww47656j4n25/]
- **SNP Identification**: Identification of SNP without a reference genome. The approach for calling SNPs compares two sets of raw NGS reads without assembly without the need of a mapping step on a reference genome [35]. [Online publication: http://dx.doi.org/10.1007/978-3-642-16321-0_14]
- bf Syntheny conservation: Genomics region being anchored by known orthologous gene(s) were compared to analyze syntenic relationships and genome rearrangements among the three lepidopteran species. The Lepido-DB Information System (http://www.inra.fr/lepidodb) has been created to facilitate the analyses and the exploration of the data generated during this project. [14] [Online publication: http://www.pnas.org/content/107/17/7680.long],
- **Transcription factors**: Transcription factors. We introduced a statistical criterium enabling to compare transcription factor binding sites matrices. We used it in a complete pipeline that extracts transcription factor binding sites matrices by using several pattern extraction algorithms, eliminates the redundancies between matrices and compares the resulting matrices with classical databases. [11] [Online publication: http://www.worldscinet.com/jbcb/08/0803/S0219720010004689.html
- **Transposable elements**: We introduce the concept of a transposable element module. We propose a new assembly method that does not require multiple sequence alignment. We show its sensitivity in several examples [24]. [Online publication: http://www.biomedcentral.com/1471-2105/11/474]

6.4. Protein Structures

Participants: Rumen Andonov [**contact**], Jacques Nicolas, Guillaume Chapuis, Rayan Chikhi, Guillaume Collet, Alexandre Cornu, Noel Malod-Dognin.

- **Local Protein Threading, sequence-structure alignment**: A novel approach to PTP has been investigated. It aligns a part of a protein structure onto a protein sequence in order to detect local similarities [13], [2]. [Online publication: http://www.sciencedirect.com/science/article/B6TYW-50G78H4-1/2/947312da7a7bbf175cab7b3288ba4f03]
- **Protein structure comparison**: A new integer programming model for Contact Map Overlap Revisited (CMO) has been proposed, together with a scoring scheme for similarities between protein structures. We propose an exact branch-and-bound algorithm with bounds obtained by a novel Lagrangian relaxation [6].

A new protein structure comparison method based on internal distances (DAST) has also been investigated. Its main characteristic is that it generates alignments having RMSD smaller than any previously given threshold [32]. [Online publication: http://www.springerlink.com/content/c1033x5842087418/]

- **Protein classification**: we participated to the SHREC'10 Protein Models Classification Track (Eurographics Workshop on 3D Object Retrieval competition). The aim of was to evaluate how well 3D shape recognition algorithms can classify protein structures according to the CATH superfamily classification. The tools developed in the Symbiose team won the competition [33]. [Online publication: http://www.loria.fr/ ritchied/papers/mavridis_shrec10.pdf

6.5. Confronting (omic) data with knowledge-based regulatory models

Participants: Anne Siegel [**contact**], Jérémie Bourdon, Michel Le Borgne, Ovidiu Radulescu, Pierre Blavy.

- **Model construction**: In collaboration with the Roscoff laboratory "Mer et Santé", we build a graph-based model of cap-dependent translation initiation in sea urchin and checked that its discrete behavior is consistent with available knowledge on the behavior of the system [9] [Online publication: http://onlinelibrary.wiley.com/doi/10.1002/mrd.21142/abstract]
- **Extract relevant information with respect to a cancer phenotype**: We use constraint-based approaches (Bioquali tool, see software section) to localize potentially active post-transcriptional regulations in the Ewing's sarcoma gene regulatory network [8]. We also designing dedicated logical rules to model the static response of biomolecular interactions implied in the cancer network. This allowed us to trace back genes implied in the cancer phenotype [16]. [Online publication: http://www.biomedsearch.com/nih/Localizing-potentially-active-post-transcriptional/21044309.html]

  [Second online publication: http://www.computer.org/portal/web/csdl/doi/10.1109/TCBB.2010.71]
- **Model correction tools**: We used Answer Set Programming, a combination of logic programming with SAT algorithmics, to prove that it is feasible to automatically perform predictions over incomplete and incorrect large scale biological network (E. Coli network) [31]. [Online publication: http://aaai.org/ocs/index.php/KR/KR2010/paper/view/1334]

6.6. Extract relevant and robust information from dynamical models

Participants: Ovidiu Radulescu [**contact**], Michel Le Borgne [**contact**], Jérémie Bourdon, Anne Siegel, Nathalie Theret, Oumarou Abdou-Arbi, Geoffroy Andrieux, Pierre Blavy, Nolwenn Le Meur.

- **Signaling pathways**: A differential continuous model was used to investigate the role of the ubiquitine ligase TIF1-g in the dynamic of SMAD proteins and a new discrete modeling approach based on statechart formalism was used to study TGF-beta signaling [45].
- **Metabolisms flexibility**: Using elementary mode based methods over metabolic networks to illustrate the flexibility of mammary gland in lactating dairy cows [43] [Online publication: www.wageningenacademic.com/default.asp?pageid=8&docid=16&artdetail=nutrientdigestion&webgroupfilter=1&]
- **Average behavior of models**: Several probabilistic approaches may be used to investigating biological networks [40]. There exists two classical opposite assumptions when modeling the dynamics of biological system: synchronous versus asynchronous depending on whether different genes may evolve at the same time or not. We present two flexible extensions of these strategies allowing us to synchronize the evolution of genes or functions. [34]. [Pub. link 2] [Online publication: http://www.cs.tut.fi/wcsb10/proc_of_wcsb10.pdf]
- **Combining genetic and metabolic regulations**: We mix Gale-Nikaido reduction steps and differential inequalities to understand how genetic regulation modify the behavior of a very abstracted model of lipid metabolism [21].

- **Time-scale reduction**: We investigate the concept of limiting steps in multiscale reaction networks. We develop a theory for linear networks with well separated reaction rate constants is developed. Performance of the algorithms is demonstrated on simple examples. [15] [Online publication: http://www.sciencedirect.com/science/article/B6TFK-4X66S69-3/2/5b4d1229475a284e659aa243813bb21a]

7. Contracts and Grants with Industry

7.1. GASSST

GASSST-GQ is an industrial contract with the GenomeQuest Company for tuning the GASSST software with industrial requirements. It is coordinated by D. Lavenier (EPI Symbiose) and JJ. Codani (GenomeQuest). http://www.genomequest.com/

[Web annoucment]

8. Other Grants and Activities

8.1. Regional Initiatives

We benefit from the strong implication of the GenOuest Ressource center in the regional Genopole to have long-term research and development relationships with most of laboratories in Brittany involved in molecular biology.

8.2. National Initiatives

8.2.1. ANR contracts

8.2.1.1. BIOWIC

Participants: Dominique Lavenier, Olivier Collin, Alexandre Cornu, François Moreews, Jonathan Piat, Guillaume Rizk, Odile Rousselet.

The BioWIC project aims to speed up both the design and the execution of bioinformatics workflows. It is funded by ANR call ARPEGE and coordinated by D. Lavenier from Jan. 2009 to Dec. 2011. http://biowic.inria.fr/

8.2.1.2. LEPIDOLF

Participants: Andres Burgos, François Coste, Fabrice Legeai, Jacques Nicolas, Pavel Senin.

The LEPIDOLF project aims at better understanding olfactory mechanisms in insects. The goal is to establish the antennal transcriptome of the cotton leafworm Spodoptera littoralis, a noctuid representative of crop pest insects. It is funded by ANR call Blanc and coordinated by E. Jacquin-Joly from UMR PISC (INRA) from 2009 to 2012.

8.2.1.3. MAPPI

Participants: Dominique Lavenier, Guillaume Chapuis, Rayan Chikhi, Nicolas Maillet, Pierre Peterlongo.

The MAPPI project aims to develop new algorithms and Bioinformatics methods for processing high trougthput genomic data. It is funded by ANR call COSINUS and coordinated by M. Raffinot (LIAFA, Paris VII) from Oct 2010 to Dec. 2013. http://mappi.arthy.org/

8.2.1.4. PELICAN

Participants: Andres Burgos, Olivier Collin, François Coste.

The PELICAN project addresses competition for light in the ocean: An integrative genomic approach of the ecology, diversity and evolution of cyanobacterial pigment types in the marine environment. It is coordinated by F. Partensky (CRNS Roscoff) from 2010 to 2013. http://www.sb-roscoff.fr/anr-pelican/

8.2.1.5. PROTEUS

Participants: Rumen Andonov, Guillaume Collet, Noel Malod-Dognin.

The PROTEUS project addresses fold recognition and inverse folding problem towards large scale protein structures prediction. It is funded by ANR call Calcul intensif and coordinated by CEA from 2006 to 2010. http://migale.jouy.inra.fr/proteus

8.2.2. Programs from research institutions

Participants: Pavlos Antoniou, Jérémie Bourdon, Guillaume Chapuis, Olivier Collin, Charles Deltel, Dominique Lavenier, Michel Le Borgne, Claire Lemaitre, Nolwenn Le Meur, Pierre Peterlongo, Anne Siegel.

- **Alcovna** The Alcovna project aims to explore possibilities of extracting information among possibly huge sets of reads without reference genome and avoiding to assemble the data. It is funded by INRIA ARC call and coordinated by P. Peterlongo from oct. 2009 to sept. 2011. http://alcovna.genouest.org
- **BioManyCores** The BioManyCores project aims to develop a librairy of bioinformatics software implemented on manycore structures such as GPU. It is funded by INRIA ADT call and supervised by J.S. Varré in Sequoia Team in Lille. http://www.biomanycores.org/
- **ParaQtlMap** The ParaQtlMap project is a join initiative from EPI Symbiose and Genetique Animale. to design high performance software for detecting quantitative trait locus. It is funded by INRIA/INRA call and coordinated by D. Lavenier (EPI Symbiose) and P. Leroy (GA INRA) from oct. 2010 to sept. 2012. https://qgp.jouy.inra.fr/index.php?option=com_content&task=view&id=17&Itemid=28
- **QuantOursin** The QuantOursin project aims at developping modeling tools based on probabilistic framework and average analysis, and apply then to the initiation of urchin translation. It is funded by a PEPS program at CNRS and coordinated by A. Siegel from april. 2010 to december. 2012. http://quantoursin.genouest.org/wiki.php/Accueil

8.2.3. Transfert and service ressources - GenOuest ressource center

Participants: Olivier Collin, Olivier Sallou, Charles Deltel, Anthony Bretaudeau, Delphine Naquin, Aurélien Roult, Romaric Sabas, Ludmila Sarbu.

- **GRISBI** The project intends at developping a production grid dedicated to bioinformatics, by gathering computational ressources of six french ressource centers. It is funded by IBISA and coordinated by C. Blanchet (IPCP Lyon) from 2009 to 2011. http://www.grisbio.fr
- **DrMotifs** is a project dedicated to develop tools for pattern discovery and research. Eventually the resource will integrate the tools in a workflow plugged on different databases in order to provide a user friendly tool geared toward motif discovery. It is funded by IBISA and coordinated by O. Collin (Symbiose) from 20010 to 2011. http://drmotifs.genouest.org
- **BioMaj** The project aims at developing a workflow engine dedicated to data synchronization and processing.The Software automates the update cycle and the supervision of the locally mirrored databank repository. It is funded by INRIA ADT program from 2009 to 2011 and coordinated by O. Collin. http://biomaj.genouest.org.
- **Mobylenet** The MobyleNet project is funded by IBiSA (2009-2010) and gathers 9 bioinformatics facilities in France. Its goal is to integrate bioinformatic services over distributed sites. The resulting network of bioinformatics web portals, hosted on the different facilities will allow a seamless integration of different softwares.

8.3. European and International Initiatives

8.3.1. Integrated program ACGT

The project (http://eu-acgt.org/home.html) aims at delivering the cancer research community an integrated CIT environment enabled by a powerful GRID infrastructure. Our contribution concerns parallelism (Grid development, tumor growth simulation) and data mining (integration of CHAVL in a R environment). It lasted from 2006 to 2010.

8.4. International Initiatives and visitors

- **Argentina**. Visit from G. Infante-Lopez ("Grupo de Procesamiento de Lenguaje Natural ") on grammar learning of sequence structure. Granted by a MinCyT-INRIA program.
- **Bulgaria**. Visit from N. Yanev (Sofia university) and Petar Milanov (SWU South-West University) on algorithmics for structural biology.
- **Chile**. Visiting from A. Maass on systems biology. The collaboration will be follow up by an Inria-Conycit project 2011-2013. The CMM team in Chile will also become an associated team of the symbiose team in 2011.
- **Germany**. Visiting in Postdam university and forward on logic programming and boolean constraint solving. The collaboration will be formalized with a procope Egide project from 2011 to 2013.
- **USA**. Guillaume Rizk has spent 4 months at the Collorado State University (S. Rajopadhye) working on the parallelization of the UnaFold package on GPU.
- **Netherlands** (CWI Life Sciences and NISB Algorithmic computational biology). Visit from G. Klau and ph-D students (two months) on algorithmics for structural biology.

9. Dissemination

9.1. Animation of the scientific community

9.1.1. Administrative functions: scientific committees, journal boards, jury

- Scientific Advisory Board of ITMO Genetics Genomics and Bioinformatics [J. Nicolas].
- Scientific Advisory Board of GDR BIM " Molecular Bioinformatics"[J. Nicolas].
- Member of the Evaluation Committee of Inria [A. Siegel]
- Member of the IRISA laboratory council [F. Coste]
- ANR committees [J. Nicolas / ANR Genomics; D. Lavenier / ANR Cosinus]
- Member of ReNaBi steering committee and coordinator for ReNaBi-GO (Grand Ouest). This regional centre includes the platforms of Nantes, Rennes and Roscoff [O. Collin]
- Scientific Advisory Board of Biogenouest [J. Bourdon, O. Collin, J. Nicolas].
- Steering committee of the International Inference community (ICGI) [F. Coste]
- Recruitment committees: junior research, assistant professor, scientific study officer [O. Collin, A. Siegel]

9.1.2. Jury of PhD Theses

- *President of Ph-D thesis juries*. S. Collange, université de Perpignan [D. Lavenier]
- *Member of Habilitation thesis jury*. A. Tisserand, université de Rennes 1 [D. Lavenier]. S. Pillement, université de Rennes 1 [D. Lavenier]. S. Verlan Université Paris Est [J. Nicolas].

- *Member of Ph-D thesis jury*. N. Lebreton, université de Rennes 1 [D. Lavenier]. C. Friguet, Agrocampus Rennes [A. Siegel].
- *Referee of Ph-D thesis*. A. Wirawan, Nanyang Technological University, Singapore [D. Lavenier]. A. Guerre, université Paris-Sud [D. Lavenier]. N. Terrapon, Université Montpellier II [J. Nicolas]. I. N'Diaye, université de Nice [A. Siegel].

9.2. Teaching

9.2.1. Masters

- *Master in computer science, ENS Rennes* Architecture of microprocessors (D. Lavenier, M1)
- *Master in computer science, Univ. Rennes 1* Symbolic sequential data (F. Coste, M2)
- *Master in computer science, ESEO, Angers* Bioinformatics (D. Lavenier, M2)
- *Master in bioinformatics, Université Rennes 1* Structural biology (R. Andonov, M2), Algorithmics of biological sequences (P. Peterlongo, M2) Systems Biology (A. Siegel, M2)
- *CCI Master, Université Rennes 1* Bioinformatics (J. Nicolas, M2)

9.2.2. Tutorials

- The challenges of bioinformatics, Thematic school for scientific computing on grid for bioinformatics: [O. Collin]
- Tutorial Evry school May 2010 "Modelling Complex Biological Systems in the Context of Genomics". [O. Radulescu]
- ICGI 2010, "Modelling Biological Sequences by Grammatical Inference" Spain [26] [F. Coste]

9.2.3. Training sessions

The platform GENOUEST proposes training sessions attended by a broad range of persons ranging from beginners to more accustomed users of bioinformatics tools. The subjects of the training sessions cover different aspects and methods of bioinformatics and computer science: sequence analysis, motif discovery, phylogenetics, programming languages and operating systems. Each year approx 60-70 persons attend to this training sessions whose detailed outline is available at http://sfc.univ-rennes1.fr/informatique/.

9.3. Conference and workshop committees, invited conferences

9.3.1. Conference program committees

- EEE International Conference on Application-specific Systems, Architectures and Processors (ASAP)
- International Conference on Field Programmable Logic and Applications (FPL)
- International Conference on Engineering of Reconfigurable Systems and Algorithms (ERSA)
- Southern Programmable Logic Conference (SPL)
- EuroPar: Workshop on Highly Parallel Processing on a Chip (HPPC)
- Workshop on Using Emerging Parallel Architectures for Computational Science (ICCS)
- ACM International Conference on Computing Frontiers (UCHPC Workshop) (CF)
- International Conference on ReConFigurable Computing and FPGAs (ReConFig)
- JOBIM'2010.
- International conference of Grammatical Inference (ICGI'10).
- Conférence d'apprentissage (CAP'10).

9.3.2. Meeting organization and scientific animation

- **Seminar** A weekly seminar of bioinformatics is organized within the laboratory. Attendees are member of the symbiose team, biologists from Brittany and computer scientists from the laboratory. A thematic day meeting on modeling issues was also organized by the team [web site: http://www.irisa.fr/symbiose/seminaires/].
- **GenOuest annual meeting** The 8th annual meeting of GenOuest computing center focused on NGS technologies. There were 120 attendees [web site: http://www.genouest.org/spip.php?article829].
- **Thematic school** A thematic school entitled "Calcul Scientifique sur Grille pour la Bioinformatique" took place in Roscoff (Station Biologique) from september 2010 the 27th to october the 1st. This school was organized by the GRISBI project and funded by the CNRS. 22 people coming from all France attended to this school. The GenOuest engineers were strongly involved in this event as organizers and lecturers.
- **Grammatical inference competition** Zulu is a competition about automata learning from membership queries. [web site: http://labh-curien.univ-st-etienne.fr/zulu/].
- **Integrative Biology project** The integrative biology project supported by the Genopole aims at integrating and exploiting data produced by the different technological platforms hosted by the regional genopole BioGenOuest. The final goal is to promote a system approach in biology. N. Le Meur was in charge of evaluation needs from technical platforms and propose relevant tools and methods.

9.3.3. International invited conferences

- 5th EGEE User Forum (Uppsala 12-15 April 2010) [O. Collin]
- workshop Perspectives in Systems biology, Instituto Butantan, sao Paolo [O. Radulescu]
- Arthropods genomics 2010, USA [27] [F. Legeai]

9.4. Theses defenses

9.4.1. Structural biology

- G. Collet. Supervised by R. Andonov [2].
- N. Malod-Dognin. Supervised by R. Andonov [4].

9.4.2. System biology

- P. Blavy. Co-supervised by S. Lagarrigue and A. Siegel. [1]
- C. Guziolowski. Supervised by A. Siegel [3].

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] P. BLAVY. *Identification des éléments clefs du métabolisme des lipides et de leurs régulateurs*, Université Rennes 1, 03 2010, http://tel.archives-ouvertes.fr/tel-00541207/PDF/These.pdf.

[2] G. COLLET. *Alignements locaux pour la reconnaissace de repliements des protéines par programmation linéaire en nombres entiers*, Université Rennes 1, Jul 2010, http://hal.inria.fr/tel-00512725/en.

[3] C. GUZIOLOWSKI. *Analysis of Large-Scale Biological Networks with Constraint-Based Approaches over Static Models*, Université Rennes 1, Jan 2010, http://hal.inria.fr/tel-00541903/en.

[4] N. MALOD-DOGNIN. *La comparaison structurale des protéines : de la maximisation du recouvrement de cartes de contacts à l'alignement basé sur les distances*, Université Rennes 1, Jan 2010, http://hal.inria.fr/tel-00509142/en.

Articles in International Peer-Reviewed Journal

[5] B. ADAMCZEWSKI, C. FROUGNY, A. SIEGEL, W. STEINER. *Rational numbers with purely periodic β-expansion*, in "Bulletin of the London Mathematical Society / The Bulletin of the London Mathematical Society", 2010, vol. 42, n^{o} 3, p. 538-552 [*DOI :* 10.1112/BLMS/BDQ019], http://hal.inria.fr/hal-00400799/en.

[6] R. ANDONOV, N. MALOD-DOGNIN, N. YANEV. *Maximum Contact Map Overlap Revisited*, in "Journal of Computational Biology", Jan 2011, vol. 18, n^{o} 1, p. 1-15, to appear [*DOI :* 10.1089/CMB.2009.0196], http://hal.inria.fr/inria-00536624/en.

[7] C. BAUDET, C. LEMAITRE, Z. DIAS, C. GAUTIER, E. TANNIER, M.-F. SAGOT. *Cassis: detection of genomic rearrangement breakpoints*, in "Bioinformatics", Aug 2010, vol. 26, n^{o} 15, p. 1897–1898 [*DOI :* 10.1093/BIOINFORMATICS/BTQ301], http://bioinformatics.oxfordjournals.org/content/26/15/1897, http://hal.inria.fr/inria-00532936/en.

[8] 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 Syst Biol", 2010, vol. 4, n^{o} 1, 146 [*DOI :* 10.1186/1752-0509-4-146], http://hal.inria.fr/inria-00538133/en.

[9] 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", Mar 2010, vol. 77, n^{o} 3, p. 257-64 [*DOI :* 10.1002/MRD.21142], http://hal.inria.fr/hal-00486418/en.

[10] J. BOURDON, I. RUSU. *Statistical Properties of Factor Oracles*, in "Journal of Discrete Algorithms", Oct 2010, vol. 5577, p. 326-338 [*DOI :* 10.1007/978-3-642-02441-2], http://hal.inria.fr/hal-00415952/en.

[11] S. CARAT, R. HOULGATTE, J. BOURDON. *A parallel scheme for comparing transcription factor binding sites matrices*, in "Journal of Bioinformatics and Computational Biology", Mar 2010, vol. 8, n^{o} 3, 18, http://hal.inria.fr/hal-00461973/en.

[12] R. CARRASCOSA, F. COSTE, M. GALLÉ, G. INFANTE-LOPEZ. *Searching for Smallest Grammars on Large Sequences and Application to DNA*, in "Journal of Discrete Algorithms", 2011, to appear, http://hal.inria.fr/inria-00536633/en.

[13] G. COLLET, R. ANDONOV, J.-F. GIBRAT, N. YANEV. *Local protein threading by Mixed Integer Programming*, in "Discrete Applied Mathematics", Jul 2010 [*DOI :* 10.1016/J.DAM.2010.05.024], http://hal.inria.fr/inria-00536537/en.

[14] E. D'ALENÇON, H. SEZUTSU, F. LEGEAI, E. PERMAL, S. BERNARD-SAMAIN, S. GIMENEZ, C. GAGNEUR, F. COUSSERANS, M. SHIMOMURA, A. BRUN-BARALE, T. FLUTRE, A. COULOUX, P. EAST, K.

GORDON, K. MITA, H. QUESNEVILLE, P. FOURNIER, R. FEYEREISEN. *Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements.*, in "Proceedings of the National Academy of Sciences of the United States of America", Apr 2010, vol. 107, n^{o} 17, p. 7680-5 [*DOI :* 10.1073/PNAS.0910413107], http://hal.inria.fr/inria-00537908/en.

[15] A. GORBAN, O. RADULESCU, A. ZINOVYEV. *Asymptotology of chemical reaction networks*, in "Chemical Engineering Science", 2010, vol. 65, n^{o} 7, p. 2310 - 2324, International Symposium on Mathematics in Chemical Kinetics and Engineering.

[16] C. GUZIOLOWSKI, S. BLACHON, T. BAUMURATOVA, G. STOLL, O. RADULESCU, A. SIEGEL. *Designing Logical Rules to Model the Response of Biomolecular Networks with Complex Interactions: An Application to Cancer Modeling.*, in "IEEE/ACM Trans Comput Biol Bioinform", Aug 2010 [*DOI :* 10.1109/TCBB.2010.71], http://hal.inria.fr/inria-00538134/en.

[17] F. LEGEAI, G. RIZK, T. WALSH, O. EDWARDS, K. GORDON, D. LAVENIER, N. LETERME, A. MÉREAU, J. NICOLAS, D. TAGU, S. JAUBERT-POSSAMAI. *Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, Acyrthosiphon pisum.*, in "BMC Genomics", 2010, vol. 11, 281 [*DOI :* 10.1186/1471-2164-11-281], http://www.biomedcentral.com/1471-2164/11/281/abstract.

[18] F. LEGEAI, G. RIZK, T. WALSH, O. EDWARDS, K. GORDON, D. LAVENIER, N. LETERME, A. MÉREAU, J. NICOLAS, D. TAGU, S. JAUBERT-POSSAMAI. *Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, Acyrthosiphon pisum.*, in "BMC Genomics", May 2010, vol. 11, n^{o} 1, 281 [*DOI :* 10.1186/1471-2164-11-281], http://hal.inria.fr/inserm-00482283/en.

[19] F. LEGEAI, S. SHIGENOBU, J.-P. GAUTHIER, J. COLBOURNE, C. RISPE, O. COLLIN, S. RICHARDS, A. WILSON, T. MURPHY, D. TAGU. *AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome*, in "Insect Molecular Biology", Feb 2010, http://hal.inria.fr/inria-00531562/en.

[20] M. OLLIVIER, F. LEGEAI, C. RISPE. *Comparative analysis of the Acyrthosiphon pisum genome and expressed sequence tag-based gene sets from other aphid species.*, in "Insect Molecular Biology", Mar 2010, vol. 19 Suppl 2, p. 33-45 [*DOI :* 10.1111/J.1365-2583.2009.00976.X], http://hal.inria.fr/inria-00537902/en.

[21] O. RADULESCU, A. SIEGEL, E. PECOU, C. CHATELAIN, S. LAGARRIGUE. *Genetically regulated metabolic networks: Gale-Nikaido modules and differential inequalities*, in "Transactions on Computational Systems Biology", 2010, http://hal.inria.fr/inria-00538136/en.

[22] S. RICHARDS, R. A. GIBBS, F. LEGEAI, D. LAVENIER, J. NICOLAS, W. HUNTER. *Genome sequence of the pea aphid Acyrthosiphon pisum.*, in "Plos Biology", Feb 2010, vol. 8, n^{o} 2, e1000313 [*DOI :* 10.1371/JOURNAL.PBIO.1000313], http://hal.inria.fr/inria-00537910/en.

[23] G. RIZK, D. LAVENIER. *GASSST: Global Alignment Short Sequence Search Tool*, in "Bioinformatics", Aug 2010, vol. 26, n^{o} 20, p. pp 2534-2540, http://hal.inria.fr/hal-00531499/en.

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

[25] A.-N. VOLKOFF, V. JOUAN, S. URBACH, S. SAMAIN, M. BERGOIN, P. WINCKER, E. DEMETTRE, F. COUSSERANS, B. PROVOST, F. COULIBALY, F. LEGEAI, C. BÉLIVEAU, M. CUSSON, G. GYAPAY, J.-M. DREZEN. *Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome.*, in "Plos Pathogens", May 2010, vol. 6, n° 5, e1000923 [*DOI :* 10.1371/JOURNAL.PPAT.1000923], http://hal.inria.fr/inria-00537901/en.

Invited Conferences

[26] F. COSTE. *Modelling Biological Sequences by Grammatical Inference*, in "ICGI 2010 Tutorial Day", Spain, José M. Sempere and Pedro Garcias, Sep 2010, http://hal.inria.fr/inria-00531552/en.

[27] F. LEGEAI, G. RIZK, T. WALSH, O. EDWARDS, J. NICOLAS, D. TAGU, S. JAUBERT-POSSAMAI. *Prediction and analyzes of non coding RNA sequences in the pea aphid genome*, in "Arthropods Genomics", USA, Jun 2010, http://hal.inria.fr/inria-00537927/en.

International Peer-Reviewed Conference/Proceedings

[28] R. CARRASCOSA, F. COSTE, M. GALLÉ, G. INFANTE-LOPEZ. *Choosing Word Occurrences for the Smallest Grammar Problem*, in "LATA", Germany, 2010, http://hal.inria.fr/inria-00476840/en.

[29] M. FEDERICO, P. PETERLONGO, N. PISANTI. *An optimized filter for finding multiple repeats in DNA sequences*, in "2010 ACS/IEEE International Conference on Computer Systems and Applications", Tunisia, 2010, http://hal.inria.fr/inria-00480001/en.

[30] M. GALLÉ. *A New Tree Distance Metric for Structural Comparison of Sequences*, in "Structure Discovery in Biology: Motifs, Networks & Phylogenies", Germany, A. APOSTOLICO, A. DRESS, L. PARIDA (editors), Dagstuhl Seminar Proceedings, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2010, n° 10231, http://drops.dagstuhl.de/opus/volltexte/2010/2737.

[31] 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", Canada, AAAI Press, 2010, http://hal.inria.fr/inria-00538137/en.

[32] N. MALOD-DOGNIN, R. ANDONOV, N. YANEV. *Maximum Cliques in Protein Structure Comparison*, in "SEA 2010 9th International Symposium on Experimental Algorithms", Italy, P. FESTA (editor), LNCS, Springer, May 2010, vol. 6049, p. 106-117, http://hal.inria.fr/inria-00536700/en.

[33] L. MAVRIDIS, V. VENKATRAMAN, D. RITCHIE, H. MORIKAWA, R. ANDONOV, A. CORNU, N. MALOD-DOGNIN, J. NICOLAS, M. TEMERINAC-OTT, M. REISERT, H. BURKHARDT, A. AXENOPOULOS, P. DARAS. *SHREC'10 Track: Protein Models*, in "Eurographics Workshop on 3D Object Retrieval - 3DOR 2010", Sweden, 2010, http://hal.inria.fr/inria-00536680/en.

[34] T. MERLE, J. BOURDON. *Complex update strategies for Probabilistic Boolean Networks*, in "7th Workshop on Computational Systems Biology", Luxembourg, TISCP series, Jun 2010, 4, http://hal.inria.fr/hal-00480366/en.

[35] P. PETERLONGO, N. SCHNEL, N. PISANTI, M.-F. SAGOT, V. LACROIX. *Identifying SNPs without a reference genome by comparing raw reads*, in "String Processing and Information Retrieval", Mexico, Oct 2010, http://hal.inria.fr/inria-00514887/en.

National Peer-Reviewed Conference/Proceedings

[36] C. BLANCHET, C. GAUTHEY, O. COLLIN, T. MARTIN, N. MELAB, F. PLEWNIAK, F. SAMSON, B. SPATARO, C. CARON. *ReNaBi-GRISBI: grande infrastructure pour la bioinformatique*, in "JOBIM 2010 - Journées Ouvertes en Biologie, Informatique et Mathématiques", France, Sep 2010, http://hal.inria.fr/inria-00540940/en.

[37] P. PICOUET, Z. DOUGHI, L. BRILLET, E. CORRE, C. CARON, O. COLLIN, F. MOREEWS, X. BAILLY. *BioDesc : gestion d'un entrepôt multiformat de descriptions de ressources bioinformatiques*, in "JOBIM", France, Sep 2010, http://hal.inria.fr/inria-00540815/en.

Workshops without Proceedings

[38] C. CHARLIER, J. MONTFORT, O. CHABROL, F. MOREEWS, M. SALHAB, L. RICHARD-PARPAILLON, R. DALBIES-TRAN, C. HENNEQUET-ANTIER, F. CHESNEL, S. UZBEKOVA, P. PONTAROTTI, J. BOBE. *Oocyte-Somatic Cells Interactions, Lessons from Evolution*, in "VIIIe Colloque Agenae / Genanimal", France, Nov 2010, http://hal.inria.fr/inria-00540830/en.

Scientific Books (or Scientific Book chapters)

[39] V. BERTHÉ, A. SIEGEL, J. THUSWALDNER. *Substitutions, Rauzy fractals and tilings*, in "Combinatorics, Automata and Number Theory", C. U. PRESS (editor), Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2010, vol. 135, http://hal.inria.fr/inria-00538135/en.

[40] J. BOURDON, D. EVEILLARD. *Probabilistic Approaches for Investigating Biological Networks*, in "Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications", M. ELLOUMI, A. Y. ZOMAYA (editors), Wiley, Jan 2011, 1066, to appear, http://hal.inria.fr/hal-00531568/en.

[41] D. LAVENIER. *Parallel Genomic Sequence Comparison*, in "Parallel and Distributed Computing", INTECH, Jan 2010, p. 273-290, http://hal.inria.fr/hal-00531505/en.

[42] D. LAVENIER, V. H. NGUYEN. *Seed-Based Parallel Protein SequenceComparison Combining Multithreading, GPU, and FPGA technologies*, in "Bioinformatics: High Performance Parallel Computer Architectures", Taylor & Francis Group / CRC Press, Jul 2010, p. 181-202, http://hal.inria.fr/hal-00531507/en.

[43] S. LEMOSQUET, O. ABDOU-ARBI, A. SIEGEL, J. GUINARD-FLAMENT, J. VAN MILGEN, J. BOURDON. *A generic stoichiometric model to analyse the metabolic flexibility of the mammary gland in lactating dairy cows*, in "Modelling nutrient digestion and utilization in farm animals", D. SAUVANT, J. V. MILGEN, P. FAVERDIN, N. FRIGGENS (editors), Wageningen Academic Publishers, 2010, http://hal.inria.fr/inria-00538138/en.

[44] N. PISANTI, M. GIRAUD, P. PETERLONGO. *Filters and seeds approaches for fast homology searches in large datasets*, in "Algorithms in computational molecular biology", M. ELLOUMI, A. Y. ZOMAYA (editors), John Wiley & sons, 2010, http://hal.inria.fr/inria-00425370/en.

Research Reports

[45] G. ANDRIEUX, N. LE MEUR, M. LE BORGNE, N. THERET. *Modélisation du Signal TGF-b : De lObtention des données à la Simulation*, Université Rennes 1, 06 2010, http://hal.inria.fr/inria-00544414/en/.

[46] N. MALOD-DOGNIN, N. YANEV, R. ANDONOV. *Comparing Protein 3D Structures Using A_purva*, INRIA, Nov 2010, n^{o} RR-7464, http://hal.inria.fr/inria-00539939/en.