IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2012

Team GENSCALE

Scalable, Optimized and Parallel Algorithms for Genomics

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Biology and Bioinformatics

Table of contents

Team GENSCALE

Keywords: Computational Biology, Next Generation Sequencing, Genomics, Protein Structure, Big Data, Parallelism

Creation of the Team: January 01, 2012 *, Updated into Project-Team:* January 01, 2013 .

1. Members

Research Scientists

Dominique Lavenier [Team leader, Senior Researcher, Cnrs, HdR]
Claire Lemaitre [Junior researcher, Inria]
Pierre Peterlongo [Junior researcher, Inria]

Faculty Members

Rumen Andonov [Professor, Univ. Rennes 1, HdR]
Antonio Mucherino [Associate Professor, Univ. Rennes 1]

Engineers

Charles Deltel [Research engineer, Inria, 50% time dedicated to the genscale project]
Fabrice Legeai [Engineer, INRA, 20% time dedicated to the genscale project]
François Moreews [Engineer, INRA, 60% time dedicated to the genscale project]
Suzete Alves-Carvalho [non permanent engineer, INRA project Peapol, since December 2012]
Alexan Andrieux [non permanent engineer, ADT Mapsembler, since September 2012]
Erwan Drezen [non permanent engineer, KoriPlast]
Anaïs Gouin [non permanent engineer, INRA Speciaphid, since October 2012]

PhD Students

Guillaume Chapuis [Région Bretagne/ENS]
Rayan Chikhi [MENRT/ENS, until October 2012]
Mathilde Le Boudic-Jamin [MENRT]
Nicolas Maillet [ANR Mappi]
Erwann Scaon [MENRT, since October 2012]

Post-Doctoral Fellows

Raluca Uricaru [INRA project Peapol, until August 2012]
Thomas Derrien [INRA project Myzus, until July 2012]
Liviu Ciortuz [Brittany Region, project Mirage, since September 2012]

Administrative Assistant

Marie-Noëlle Georgeault

2. Overall Objectives

2.1. High throughput processing of genomic data

GenScale is a bioinformatics research team. It focuses on methodological research at the interface between computer science and genomic. The main objective of the group is the design of scalable, optimized and parallel algorithms for processing the huge amount of genomic data generated by the recent advances of biotechnologies.

GenScale research activities cover the following domains:

- Next Generation Sequencing (NGS)
  - Fast and low memory fingerprint assembly
  - Variant extraction on raw data (without assembly)
  - Mapping
- High throughput sequence analysis
  - Bank to bank comparison
  - Metagenomic sample processing
- 3D Protein structures
  - Alignment, comparison, classification
  - Conformation extraction from NMR data
- Bioinformatics workflow
  - Graphical capture
  - Parallel processing (cluster, cloud)

This pure computer science activity is maintained with strong collaboration with life science research groups on challenging genomic projects.

2.2. Highlights of the Year

- GenScale organized **JOBIM 2012**, the French conference on computational biology wich gathered 375 participants in Rennes. [web site: http://jobim2012.inria.fr/].
- GenScale and CWI proposed the first web server for comparison of protein structure alignments (**CSA**). [web site: http://csa.project.cwi.nl]
- KLAST software released by Korilog. KLAST is an improved version of the PLAST technology developed by GenScale for bank-to-bank sequence similarity search. [Korilog promotion]

3. Scientific Foundations

3.1. Introduction

To tackle challenges brought by the processing of huge amount of genomic data, the main strategy of GenScale is to merge the following computer science expertise:

- Data structure;
- Combinatorial optimization;
- Parallelism.

3.2. Data structure

To face the genomic data tsunami, the design of efficient algorithms involves the optimization of memory fingerprints. A key point is the design of innovative data structures to represent large genomic datasets into computer memories. Today's limitations come from their size, their construction time, or their centralized (sequential) access. Random accesses to large data structures poorly exploit the sophisticated processor cache memory system. New data structures including compression techniques, probabilistic filters, approximate string matching, or techniques to improve spatial/temporal memory access are developed [3].

3.3. Combinatorial optimization

For wide genome analysis, Next Generation Sequencing (NGS) data processing or protein structure applications, the main issue concerns the exploration of sets of data by time-consuming algorithms, with the aim of identifying solutions that are optimal in a predefined sense. In this context, speeding up such algorithms requires acting on many directions: (1) optimizing the search with efficient heuristics and advanced combinatorial optimization techniques [2], [5] or (2) targeting biological sub-problems to reduce the search space [7], [9]. Designing algorithms with adapted heuristics, and able to scale from protein (a few hundreds of amino acids) to full genome (millions to billions of nucleotides) is one of the competitive challenges addressed in the GenScale project.

3.4. Parallelism

The traditional parallelization approach, which consists in moving from a sequential to a parallel code, must be transformed into a direct design and implementation of high performance parallel software. All levels of parallelism (vector instructions, multi-cores, many-cores, clusters, grid, clouds) need to be exploited in order to extract the maximum computing power from current hardware resources [6], [8], [1]. An important specificity of GenScale is to systematically adopt a design approach where all levels of parallelism are potentially considered.

4. Application Domains

4.1. Sequence comparison

Historically, sequence comparison has been one of the most important topics in bioinformatics. BLAST is a famous software tool particularly designed for solving problems related to sequence comparisons. Initially conceived to perform searches in databases, it has mostly been used as a general-purpose sequence comparison tool. Nowadays, together with the inflation of genomic data, other software comparison tools that are able to provide better quality solutions (w.r.t the ones provided by BLAST) have been developed. They generally target specific comparison demands, such as read mapping, bank-to-bank comparison, meta-genomic sample analysis, etc. Today, sequence comparison algorithms must clearly be revisited to scale up with the very large number of sequence objects that new NGS problems have to handle.

4.2. Genome comparison

This application domain aims at providing a global relationship between genomes. The problem lies in the different structures that genomes can have: segments of genome can be rearranged, duplicated or deleted (the alignment can no longer be done in one piece). Therefore one major aim is the study of chromosomal rearrangements, breaking points, structural variation between individuals of the same species, etc. However, even analyses focused on smaller variations such as Single Nucleotide Polymorphisms (SNP) at the whole genome scale are different from the sequence comparison problem, since one needs first to identify common (orthologous) parts between whole genome sequences and thus obtain this global relationship (or map) between genomes. New challenges in genome comparison are emerging with the evolution of sequencing techniques. Nowadays, they allow for comparing genomes at intra-species level, and to deal simultaneously with hundreds or thousands of complete genomes. New methods are needed to find the sequence and structural variants between such a large number of non-assembled genomes. Even for the comparison of more distant species, classical methods must be revisited to deal with the increasing number of genomes but more importantly their decreasing quality: genomes are no longer fully assembled nor annotated.

4.3. Protein comparison

Comparing protein is important for understanding their evolutionary relationships and for predicting their structures and their functions. While annotating functions for new proteins, such as those solved in structural genomics projects, protein structural alignment methods may be able to identify functionally related proteins when the sequence identity between a given query protein and the related proteins are low (i.e. lower than 20%). Moreover, protein comparison allows for solving the so-called protein family identification problem. Given an unclassified protein structure (query), the comparison of protein structures can be used for assigning a score measuring the "similarity" between the query and the proteins belonging to a set of families. Based on this score, the query is assigned to one of the families of the set. The knowledge acquired by performing such analyses can then be exploited in methods for protein structure prediction that are based on a homology modeling approach.

5. Software

5.1. Next Generation Sequencing

Participants: Alexan Andrieux, Rayan Chikhi, Dominique Lavenier, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo, Raluca Uricaru.

- **Genome assembly** [contact: P. Peterlongo]
  - **Minia : ultra low memory assembly** Minia is a short-read assembler based on a de Bruijn graph, capable of assembling a human genome on a desktop computer in a day. The output of Minia is a set of contigs. Minia produces results of similar contiguity and accuracy to other de Bruijn assemblers (e.g. Velvet). http://minia.genouest.org/
  - **Mapsembler: targeted assembly software.** Mapsembler is a targeted assembly software. From sets of NGS raw reads and a set of input sequences (starters), it determines if each starter could be constructed from the reads. Then for each "read-coherent" starter, Mapsembler outputs its sequence neighborhood as a linear sequence or as a graph, depending on the user choice. http://alcovna.genouest.org/mapsembler/
- **Variant detection** [contact: C. Lemaitre]
  - **kisSnp and kisSplice : variant identification without the use of a reference genome**. kisSnp is a tool to find single nucleotide polymorphisms (SNP) by comparing two sets of raw NGS reads. http://alcovna.genouest.org/kissnp/ KisSplice finds alternative splicings but also short insertions, deletions and duplications, SNPs and sequencing errors in one or two RNA-seq sets, without assembly nor mapping on a reference genome. http://alcovna.genouest.org/kissplice/
  - **Kissreads: quantification of variants** Kissreads considers sets of NGS raw reads and a set of input sequences (starters). Mapping reads to each starter, it provides quantitative (coverage depth) and qualitative (mapped read quality) information about each starter.
- **Read mapping** [contact: D. Lavenier]
  - **GASSST: short reads mapper** The GASSST software (Global Alignment Short Sequence Search Tool) is a general purpose mapper. GASSST finds global alignments of short DNA sequences against large DNA banks. One main characteristic of GASSST is its ability to perform fast gapped alignments and to process long reads compared to other current similar tools. http://www.irisa.fr/symbiose/projects/gassst/

5.2. High throughput sequence analysis

Participants: Rayan Chikhi, Erwan Drezen, Dominique Lavenier, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo.

- **PLAST : efficient bank-to-bank alignments** PLAST (Parallel Local Alignment Search Tool) is a parallel alignment search tool for comparing large protein banks. PLAST runs 3 to 5 times faster than the NCBI-BLAST software. An improved version is commercialized by the Korilog Company, including the DNA bank-to-bank option. [contact: D. Lavenier] http://www.irisa.fr/symbiose/projects/plast/
- **Compareads : efficient comparison of large metagenomics NGS datasets** This software extracts similar DNA sequences (reads) between two metagenomic datasets. It requires a small and fixed amount of memory and can thus be used on huge datasets. [contact: P. Peterlongo] http://alcovna.genouest.org/compareads/

5.3. 3D Protein structures

Participants: Rumen Andonov, Guillaume Chapuis, Mathilde Le Boudic-Jamin, Antonio Mucherino.

- **CSA and DALIX** CSA (Comparative Structural Alignment) is a webserver for computing and comparing protein structure alignments. CSA is able to compute score-optimal alignments with respect to various inter-residue distance-based scoring schemes. [contact: R. Andonov] http://csa.project.cwi.nl/
- **A_purva** 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. [contact: R. Andonov] http://mobyle.genouest.org/cgi-bin/Mobyle/portal.py?forms::A_Purva
- **MD-Jeep** MD-jeep is a software tool for solving distance geometry problems. It is able to solve a subclass of instances of the problem for which a discrete reformulation can be supplied. We refer to this subclass of instances as the Discretizable Molecular Distance Geometry Problem (DMDGP). We employ a Branch & Prune (BP) algorithm for the solution of DMDGPs. [contact: A. Mucherino] http://www.antoniomucherino.it/en/mdjeep.php

5.4. HPC and Parallelism

Participants: Guillaume Chapuis, Dominique Lavenier, François Moreews.

- **QTLmap** QTLMap is a tool dedicated to the detection of Quantitative Trait Loci (QTL) from experimental designs in outbred population. QTLMap was recently ported to GPU and offers reduced run times. [contact: D. Lavenier] http://www.inra.fr/qtlmap/
- **SLICEE** (Service Layer for Intensive Computation Execution Environment) is part of the BioWIC project. This software proposes (1) to abstract the calls to the cluster scheduler by handling command submission; (2) to take care of exploiting the data parallelism with data specific methods; (3) to manage data using a cache references mecanism and route data between tasks. [contact: F. Moreews] http://vapor.gforge.inria.fr/

6. New Results

6.1. Next Generation Sequencing

Participants: Alexan Andrieux, Rayan Chikhi, Liviu Ciortuz, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo, Erwann Scaon, Raluca Uricaru.

- **Ultra-low memory data structure for de novo genome assembly** : We propose a new encoding of the de Bruijn graph, which occupies an order of magnitude less space than current representations. The encoding is based on a Bloom filter, with an additional structure to remove critical false positives. [24]
- **Transcriptomic variant detection** : We developped a new method, called kissplice, that calls splicing variant events from sets of RNA-seq NGS reads. It constructs the de-Bruijn graph from the reads and then detects in this graph all patterns corresponding to alternative splicing events. [21]
- **Targeted assembly of NGS data**: The method is based on an iterative targeted assembler which processes large datasets of reads on commodity hardware. Basically, it checks for the presence of given regions of interest in the reads and reconstructs their neighborhood, either as a plain sequence (consensus) or as a graph (full sequence structure). [20]
- **Mapping reads on a graph:** We developped a strategy for directly mapping sequences on bi-directed de-Bruijn graphs. Based on a seed-and-extend algorithm it can be applied on large datasets.[31]
- **Pea aphid genomics and evolution.** Using some of the softwares developped by Genscale, genomic variants and expression data of the pea aphid were analysed, revealing candidate regions involved in the adaptation to host plant, and genes involved in the reproduction mode, either with differential expression patterns or particular patterns of evolutionary rates in other aphid species. [11], [12], [19]

6.2. Protein structures

Participants: Rumen Andonov, Guillaume Chapuis, Dominique Lavenier, Mathilde Le Boudic-Jamin, Antonio Mucherino.

- **Comparison of pairwise protein structure alignments.** The method provides either optimal, top-scoring alignments or heuristic alignments with quality guarantee for some inter-residue distance-based measures. Alignments are compared using a number of quality measures and intuitive visualizations. The methodology brings new insight into the structural relationship of the protein pairs and is a valuable tool for studying structural similarities. [23]
- **Alignment graph.** This object is the main input to find similarities between biomolecules (ARN, proteins). This kind of graph has to model physical and/or chemical properties of the biomolecules and need to take into account constraints dictated by the type of applications (3D comparison, docking, etc.). Our research aims to provide a strategy to automate the building of alignment graphs. A prototype software, called MAGE, is currently under test to validate our approach.
- **Mathematical model and exact algorithm for optimally aligning protein structures.** The algorithm proposes for the first time, to evaluate the popular DALI heuristic in sound mathematical terms. The results indicate that DALI usually computes optimal or close to optimal alignments. However, we detect a subset of small proteins for which DALI fails to generate any significant alignment, although such alignments do exist [22].
- **Modeling the protein flexibility by distance geometry.** We suggest a strategy for modeling protein flexibility that is based on the discretization of the space of possible molecular conformations for a protein. The same discretization process was previously employed for discretizing Molecular Distance Geometry Problems (MDGPs) [30].
- **NMR problems.** We introduce formally the Discretizable Molecular Distance Geometry Problem (DMDGP) for solving the 3D structure of a protein based on Nuclear Magnetic Resonance data together with an algorithm, which we named the "Branch & Prune" (BP), for the solution of DMDGPs [16].We also provide surveys on these recent works about DMDGPs [15], [27].
- **Improvements and variants of the DMDGP**. We exploit symmetries in DMGP trees. We consider similar or related problems (re-ordering of the vertices, relaxing vertices consecutivity assumtion, including side chains and finding low energy homopolymer conformations). Parallelism has also been investigated. [17], [14], [18], [28], [26], [29]

6.3. High Throughput Sequence Analysis

Participants: Rayan Chikhi, Erwan Drezen, Dominique Lavenier, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo.

- **Comparing metagenomes.** This research aims to define new ways of comparing billions of sequences generated by NGS sequencers. Standard techniques don't scale with such volume of data, both in terms of memory fingerprint and execution time. We have successfully tested a new method based on probabilistic data structures (Bloom filter) allowing large sets of sequences to be indexed in a short time on standard computers. [25]
- **Bank-to-bank comparison.** In cooperation with the Korilog company we improve the PLAST technology developed for bank-to-bank sequence similarity search. Structuration of the index has been revisited to reduce the memory fingerprint and the execution time. The Korilog company has successfully integrated this improvments sofware component in its own software and has just began its promotion with promising responses from several potential clients.[Korilog promotion]

6.4. HPC and Parallelism

Participants: Rumen Andonov, Guillaume Chapuis, Charles Deltel, Dominique Lavenier, Fabrice Legeai, François Moreews.

- **High performance pipelines for annotation**. We participed to effort of URGI (INRA Versailles) to set up TriAnnot, a modular architecture allowing for the annotation of genomes. The TriAnnot pipeline is parallelized on a 712 CPU computing cluster that can run a 1-Gb sequence annotation in less than 5 days. [13]
- **Bioinformatics Workflows.** SLICEE is an environment to capture and parallelize time-consuming bioinformatics applications on grid or cloud platforms. In 2012, a web interface has been designed to interactively draw and run workflows from standard browsers ([workflow portal]). Several workflows used in the BioWIC ANR project have been successfully tested on this platform (http://biowic.inria.fr/)
- **Parallelization of a pseudo-clique solver.** Following such solvers as DAST and A_purva, we develop a pseudo-clique solver for alignment graphs. Looking for pseudo-cliques allows us to relax some of the constraints that are inherent to clique finding and thus maintain polynomial run times. We focus on defining a parallel algorithm and developping an implementation that benefits from multiple levels of parallelism: fine grain parallelism (bit-level parallelism, SSE instructions) and coarse grain parallelism (mutli-core parallelism). Intended applications range from protein local similarity search to protein surface similarity search or even docking.

7. Bilateral Contracts and Grants with Industry

7.1. Sequence Comparison, Korilog

Intensive bank-to-bank comparison with Korilog : this collaborative project between the Korilog company and the GenScale team aims to investigate new research directions in the bank-to-bank sequence comparison problem. Two research axes are followed : constrained exploration of the search space and adaptation of the ORIS algorithm, developed by D. Lavenier for fast DNA comparison, to the protein sequences. It is funded for 3 months (Nov. 2012 - Feb. 2013), including the visit of assistant professor Van-Hoa Nguyen from Vietnam.

KoriPlast: this project is a cooperation between GenScale and the Korilog company, it is funded by Région Bretagne from June 2011 to Nov. 2012. It aims to industrialize the PLAST software prototype, previously developed in GenScale, that performs intensive genomic bank-to-bank comparisons. The commercial version is now called KLAST http://www.korilog.com/index.php/KLAST-high-performance-sequence-similarity-search-tool.html

7.2. Peapol

The Peapol project is funded by Sofiproteol company whose mission is to develop the French vegetable oil and protein industry, open up new markets, and ensure an equal distribution of value among its members. The Peapol project counts two collaborators, Biogemma, and INRA, the latter working in collaboration with the Genscale team, in charge of algorithmic research in the context of the project. This collaboration enabled to hire in the Genscale team Raluca Uricaru for 18 months on an INRA post doctoral position, followed by Suzete Alves-Carvalho (engineer).

7.3. Rapsodyn

RAPSODYN is a long term project funded by the IA French program (Investissement d'Avenir) and several field seed companies, such as Biogemma, Limagrain and Euralis. The objective is the optimisation of the rapeseed oil content and yield under low nitrogen input. GenScale is involved in the bioinformatics workpackage, in collaboration with Biogemma's bioinformatics team, to elaborate advanced tools dedicated to polymorphism.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Program from Région Bretagne : MIRAGE

Participants: Liviu Ciortuz, Claire Lemaitre, Pierre Peterlongo.

The MIRAGE project is funded by Région Bretagne in the framework of the SAD call (Stratégie Attractivité Durable) which aims at attracting international post-doctorant for one year. The MIRAGE project aims at developping new methods to detect complex variation (structural variations) in non-assembled NGS data. It is funded from Sept. 2012 until August 2013 and coordinated by C. Lemaitre.

8.1.2. Partnership with INRA

Participants: Thomas Derrien, Anaïs Gouin, Fabrice Legeai, François Moreews, Raluca Uricaru.

We have a strong and long term collaboration with biologists of INRA in Rennes : IGEPP and SENAH units. This partnership concerns both service and research activities and is acted by the hosting of two ingineers (F. Legeai, F. Moreews) and by the co-supervision of two post-doctorants and one non permanent engineer. In particular, the collaboration with the IGEPP team includes several research projects in which Genscale is formally a partner : an INRA project PEAPOL including an industrial partner, Biogemma, and an ANR project SPECIAPHID. These projects fund the non-permanent INRA members.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. BIOWIC

Participants: Rumen Andonov, Dominique Lavenier, François Moreews.

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 June 2012. http://biowic.inria.fr/

8.2.1.2. MAPPI

Participants: Rayan Chikhi, Dominique Lavenier, Claire Lemaitre, 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.

8.2.1.3. FATINTEGER

Participants: Dominique Lavenier, François Moreews.

The FatInteger project aims to identify some of the transcriptional key players of animal lipid metabolism plasticity, combining high throughput data with statistical approaches, bioinformatics and phylogenetic. It is funded by ANR call BLANC and coordinated by F. Gondret from 2012 to 2015.

8.2.1.4. SPECIAPHID

Participants: Thomas Derrien, Anaïs Gouin, Fabrice Legeai, Claire Lemaitre.

The SPECIAPHID project aims to understand the adaptation and speciation of pea aphids by re-sequencing and comparing the genomes of numerous aphid individuals. Genscale's task, as associate partner, is to apply and develop new methods to detect variation between re-sequenced genomes, and in particular complex variants such as structural ones. It is funded by ANR call BLANC and coordinated by J-C Simon (Inra, Rennes) from January 2012 to Dec. 2014.

8.2.1.5. ADA-SPODO

Participants: Rumen Andonov, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, François Moreews, Pierre Peterlongo.

The ADA-SPODO project aims at identifying all sources of genetic variation between two strains of an insect pest : Lepidoptera Spodoptera frugiperda in order to correlate them with host-plant adaptation and speciation. Genscale's task is to develop new efficient methods to compare complete genomes along with their post-genomic and regulatory data. It is funded by ANR call BLANC and coordinated by E. d'Alençon (Inra, Montpellier) from October 2012 to Dec. 2015.

8.2.1.6. RAPSODYN

Participants: Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo, Erwann Scaon.

RAPSODYN is a long term project funded by the IA French program (Investissement d'Avenir) for 7.5 years (07/2012-12/2019). The objective is the optimisation of the rapeseed oil content and yield under low nitrogen input. GenScale is involved in the bioinformatics workpackage to elaborate advanced tools dedicated to polymorphism.

8.2.1.7. LEPIDOLF

Participants: Dominique Lavenier, Fabrice Legeai.

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. As part of this project, a post-doctoral student, Aurore Gallot, visited Genscale for 5 months.

8.2.2. Programs from research institutions

8.2.2.1. Mapsembler

Participants: Alexan Andrieux, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

The Mapsembler project aims at finalizing and to distributing the Mapsembler tool. It is funded by Inria ADT call (2012) and coordinated by P. Peterlongo from oct. 2012 to sept. 2014. http://alcovna.genouest.org/mapsembler/

8.2.2.2. Mastodons

Participants: Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

This project, funded by the CNRS Big Data program in 2012, aims do investigate the challenge brought by the processing of high throughput sequencing genomic data. It is coordinated by D. Lavenier from june 2012 to december 2012.

8.2.2.3. BioManyCores

Participants: Guillaume Chapuis, Charles Deltel, Dominique Lavenier.

The BioManyCores project aims to develop a library of bioinformatics softwares 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/

8.2.2.4. ParaQtlMap

Participants: Guillaume Chapuis, Charles Deltel, Dominique Lavenier.

The ParaQtlMap project is a join initiative from Genscale team 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 (Genscale) 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

8.2.2.5. Barcoding de nouvelle génération

Participants: Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

This project is a join initiative between Genscale and LECA (Laboratoire d'Ecologie Alpine in Grenoble). It aims at developping new algorithmic approaches for the species identification from low coverage NGS data. It is funded by a PEPS program at CNRS/Inria and coordinated by C. Lemaitre from sept. 2012 to december. 2013.

8.2.2.6. Poly-BNF

Participants: Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo, Erwann Scaon.

This projects aims to develop bioinformatics strategies for studying polyploid genomes. It is a one year project (09/2012 – 09/2013) funded by the University of Rennes 1. It is a joined project with CNRS/EcoBio lab and INRA/IGEPP lab.

8.2.3. Cooperations

8.2.3.1. Inria Bamboo Team

Participants: Claire Lemaitre, Pierre Peterlongo.

We maintain a long term collaboration with Inria Bamboo Team on the problems of finding biological information, such as variants, in NGS raw data.

8.2.3.2. LIGM, Paris

Participant: Pierre Peterlongo.

P. Peterlongo collaborates with the LIGM lab in Paris (UMR 8049), on problems of large NGS raw data indexation.

8.2.3.3. LIX

Participant: Antonio Mucherino.

A. Mucherino collaborates since 5 years with LIX, Ecole Polytechnique, in Palaiseau on the distance geometry problem. We reformulated the problem as a combinatorial optimization problem and we conceived an ad-hoc algorithm for the solution of this class of problems.

8.3. European Initiatives

8.3.1. Collaborations with Major European Organizations

Partner: CWI, University of Amsterdam, (Netherland)

Subject of cooperation: Optimization algorithms for protein structures alignments.

8.4. International Initiatives

8.4.1. Participation In International Programs

8.4.1.1. CONICYT (Chile)

Program: Coopération bilatérale CNRS

Title: Wine fermentation analysis by biclustering

Inria principal investigator: Antonio MUCHERINO

International Partner (Institution - Laboratory - Researcher):

Technical University Federico Santa Maria (Chile)

Duration: Jan 2012 - Dec 2012

This project aims at using data mining techniques for predicting problematic wine fermentations from the first stages of the fermentation process.

8.4.2. Collaborations

Partner: IMECC, UNICAMP, Campinas-SP (Brazil)

Subject: distance geometry, bioinformatics.

Partner:COPPE, Federal University of Rio de Janeiro (Brazil)

Subject: distance geometry, bioinformatics.

Partner: Los Alamos National Laboratory (lanl), Los Alamos (USA)

Subjects: Combinatorial algorithms (shortest paths, graph partitioning, combinatorial optimization) and algorithm engineering (efficient implementation of combinatorial algorithms)

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- Carlile Lavor, from IMECC-UNICAMP, Campinas-SP, Brazil, visited Genscale 3 times (2 times, for 1 week, funded by his own projects and 1 time, for 1 month, funded by "mois ISTIC").
- Alejandra Urtubia, from Universidad Tecnica Federico Santa Maria, Valparaiso, Chile, visited genscale for 2 weeks. This visit was funded by CNRS-CONICYT project on wine fermentation (A. Mucherino).
- Hristo Djidjev from Los Alamos, USA, visited Genscale for a month in the framework of University of Rennes 1 visiting positions "professeur invité".
- Van-Hoa Nguyen from University of Angiang, Viet Nam, visited GenScale for 3 months (nov. 2012 - jan. 2013). The visit was funded by the French Mastodons program from CNRS to research focusing on bioinformatics big data problem.
- Rafael Santos, from UNICAM, Bresil, visited GensCale for 3 months (oct. 2012 - dec. 2012). The visit was funded by CNPq (collaboration with A. Mucherino on protein structure).
- Virginia Silva da Costa, from the Federal University of Rio, Bresil, visited Genscale for 4 months (mar. 2012 - june 2012), funded by CAPES.
- Mariade Cola, from the University of Rome, Italia, visited Genscale for 3 months (apr. 2012 - june 2012), funded by IASI-CNR.

- Sharat Bogaraju, from IIT Delhi, India, visited GenScale for 6 months (dec. 2011 - may 2012). The visit was funded by Rennes Metropole (International exchange of PhD Students). Collaboration with D. Lavenier on parallel bioinformatics algorithms .

8.5.2. Visits to International Teams

- Antonio Mucherino visited IMECC-UNICAMP, Campinas-SP, Brazil, for 2 months, under the program "chaires française à São Paulo"
- Claire Lemaitre and Pierre Peterlongo visited for 1 week the "Laboratory of Bioinformatics and Mathematics of the Genome" hosted at CMM at University of Chile. The visit was funded by CIRIC-omics research line of the Inria center in Chile.
- Nicolas Maillet (PhD) wisited during three months the LNCC (Laboratório Nacional de Computação Científica) in Petropolis (state of Rio de Janeiro, Brazil) from March to June 2012.
- Mathilde Le Boudic-Jamin (PhD) visited the CWI in Amsterdam, Neetherlands (June 2012, one month) and collaborated with Gunnar KLAU and Inken WOHLERS on the family identification problem.

9. Dissemination

9.1. Scientific Animation

9.1.1. Meeting organization and scientific animation

- **JOBIM** the annual french conference on computational biology was organised this year by Genscale. C. Lemaitre and P. Peterlongo coordinated the organisation comittee. JOBIM 2012 took place from the 3rd to the 6th of June at Université Rennes 1 and gathered 375 participants mainly from France and french-speaking countries, bringing together researchers from the mathematical, computational and life sciences. The program was of high quality, including 7 international key-note speakers and 33 selected communications, along with 112 exposed posters and 6 industrial partners, participating through communications and exhibitions. This conference is a cornerstone in the French (speaking) computational biology research. It enables the whole community to meet, to share scientific results and to organize the discipline. Moreover, it provided a large visibility to the host city and to the organizing team. [web site: http://jobim2012.inria.fr/].
- **Seminar** A weekly seminar of bioinformatics is organized within the laboratory. Attendees are member of the ex-symbiose team (now teams Genscale, Dyliss and Genouest), biologists from Brittany and computer scientists from the laboratory. [web site: http://symbiose.irisa.fr/symbiose-seminars]
- **Discussion group on NGS** This working group brings together biologists and computational scientists, mainly from Rennes. Approximatively every two months, it enables to share experiences, raise new questions or propose news solutions about NGS issues. This groups is highly appreciated, between 40 and 50 persons attend this events and some from distant sites (Roscoff, Nantes) follow the meeting through visioconference. [web site: http://ngs.genouest.org/]
- **Sessions organised at ISMP 2012** Sessions “Bioinformatics and Combinatorial Optimization I” and “Combinatorial Optimization: Distance geometry applications” were organised at the 21st International Symposium on Mathematical Programming (ISMP 2012) in Berlin. http://ismp2012.mathopt.org/
- **Winter School : Algorithms in Structural Bio-informatics**, was co-organised by R. Andonov in december 2012 at Inria Sophia Antipolis.
- **Bioinformatics tutorials**, at CARI (Conference Africaine de Recherche en Informatique) in Alger (Algérie), organized by D. Lavenier [web site: http://cari-info.org/]

9.1.2. Conference program committees

- International Conference on Field Programmable Logic and Applications (FPL) [D. Lavenier]
- International Conference on Engineering of Reconfigurable Systems and Algorithms (ERSA) [D. Lavenier]
- Southern Programmable Logic Conference (SPL) [D. Lavenier]
- International Conference on ReConFigurable Computing and FPGAs (ReConFig) [D. Lavenier]
- Workshop on Emerging Parallel Architectures (WEPA 2012) [D. Lavenier]
- ACM International Conference on Computing Frontiers (UCHPC Workshop) [D. Lavenier]
- IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP) [D. Lavenier]
- IEEE International Conference on Parallel and Distributed Systems (ICPADS 2012) [D. Lavenier]
- Workshop on Computational Optimization (WCO12), Wroclaw, Poland, September 9–12, 2012 [A. Mucherino, R. Andonov]
- Data Mining in Agriculture (DMA12), Berlin, Germany, July 20, 2012. [A. Mucherino]
- 21st International Symposium on Mathematical Programming (ISMP12), Berlin, Germany, August 19–24, 2012 [A. Mucherino, R. Andonov]
- JOBIM'2012 [D. Lavenier, C. Lemaitre, P. Peterlongo]
- SeqBio 2012 [P. Peterlongo]

9.1.3. Administrative functions: scientific committees, journal boards

- Member of the administrative council of ISTIC [R. Andonov]
- ANR Evaluation Committee (Numerical Models Program) [D. Lavenier]
- Reviewer for NSF projects [D. Lavenier]
- Recruitment committees: 2 assistant professors [P. Peterlongo, D. Lavenier], 5 Inra engineers [F. Legeai]
- Member of the Editorial Board of The Scientific World JOURNAL, bioinformatics domain [D. Lavenier]
- Member of Inria CDT [D. Lavenier]
- Member of the RAPSODYN Governing Council [D. Lavenier]
- MESR/DGRI - MEI Expert (International Cooperation Program) [D. Lavenier]
- Inria center referee of Scientific mediation [P. Peterlongo]
- Member of the redaction committee Ouest Inria [P. Peterlongo]
- publication reviewing for NAR, BMC Genomics, Bioinformatics, briefing in Bioinformatics, BMC Bioinformatics, PLoS One, European Journal of Entomology, Bulletin of entomological research, Symmetry, Mathematical Programming, RECOMB, International Journal of Reconfigurable Computing, journal of experimental algorithms, Recent Patents on DNA and Gene Sequence [D. Lavenier, F. Legeai, A. Mucherino, C. Lemaitre, P. Peterlongo].

9.1.4. Invited talks

- D. Lavenier gave an invited talk at CARI (Conference Africaine de Recherche en Informatique) in Alger (Algerie)
- A. Mucherino gave an invited talk at IMECC, UNICAMP, Campinas (Brazil)
- A. Mucherino gave an invited talk at BIA, INRA, Toulouse
- A. Mucherino gave an invited talk at the Department of Informatics, University of Florence (UNIFI), Florence (Italy)

- P. Peterlongo gave an invited talk at “Journee axe genomique biogenouest” (Nantes)
- C. Lemaitre gave an invited talk at Inria Bonsai team in Lille
- C. Lemaitre gave an invited talk at “Séminaire BioSticker”, LINA (Nantes)

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : A. Mucherino, R. Andonov, P. Peterlongo, Graph algorithms, 130h, L3, Univ. Rennes 1, Rennes France.

Licence : R. Andonov, Algorithmics, 20h, L3, Univ. Rennes 1, Rennes France.

Licence : D. Lavenier, Computer Architecture and system , 70 h, L3 - ENS Magister, Rennes, France

Master : R. Andonov, Advanced algorithmics, 20h, M1, Univ. Rennes 1, Rennes France.

Master : R. Andonov, A. Mucherino, Operations research, 94h, M1, Univ. Rennes 1, France.

Master : G. Chapuis, Compilation, 41h, M1, Univ. Rennes 1, France.

Master : D. Lavenier, Bioinformatics, 20h, M2, Angers, France

Master : A. Mucherino, Initiation to systems and networks, 41h, M2, Univ. Rennes 1, France

Master : C. Lemaitre, P. Peterlongo, Text algorithmics for Bioinformatics, 40 h, M1, Univ. Rennes 1, France.

Master : C. Lemaitre, Dynamical systems for biological networks, 23h, M2, Univ. Rennes 1, France

Master : F. Legeai, Bioinformatics, 4h, M2, Univ. Rennes 1, France.

Master : A. Mucherino, R. Andonov, P. Peterlongo, Sequence and structure algorithms, 50h, M2, Univ. Rennes 1, France

9.2.2. Supervision

PhD Defense: Rayan Chikhi, *Computational methods for de novo assembly of next-generation sequencing data* [10] , ENS Cachan, defended on July 2nd 2012, supervised by D. Lavenier [online manuscript: http://tel.archives-ouvertes.fr/tel-00752033/]

PhD in progress : Guillaume Chapuis, *Bioinformatique parallèle*, Univ. Rennes 1, started in October 2010, supervised by D. Lavenier and R. Andonov

PhD in progress : Mathilde Le Boudic-Jamin, *Through Flexible Protein-Protein Docking*, Univ. Rennes 1, started in October 2011, supervised by R. Andonov

PhD in progress : Nicolas Maillet, *Algorithme pour l'assemblage de données NGS de métagénomique*, Univ. Rennes 1, started in November 2010, supervised by D. Lavenier and P. Peterlongo

PhD in progress : Erwan Scaon, *Modèles et algorithmes pour l'assemblage de novo de génomes à forte redondance*, Univ. Rennes 1, started in October 2012, supervised by D. Lavenier and C. Lemaitre

PhD in progress : François Moreews, *Environnement intégré de conception et d'exécution de workflows en bioinformatique: du prototypage au calcul intensif. Applications à la recherche de motifs de régulation dans les génomes*, Univ. Rennes 1, started in November 2012, supervised by D. Lavenier and S. Lagarigue

9.2.3. Juries

- *President of Ph-D thesis jury*. N. Abbas, ENS Cachan [R. Andonov]
- *Member of Ph-D thesis juries*. I. Wohlers, Univ. Amsterdam [R. Andonov], A. Nicolas, Université de Rennes 1 [R. Andonov], R. Saidi, Université de Clermont-Ferrand [R. Andonov], O. Gaci Université du Havre [R. Andonov]

- *Referee of Ph-D thesis*. T. Tan Truan, Université de Lille [D. Lavenier]

9.3. Popularization

- Participation to the event "A la découverte de la recherche" (presentation of the research activity to high school students) [D. Lavenier]

10. Bibliography

Major publications by the team in recent years

[1] R. ANDONOV, S. BALEV, N. YANEV. *Protein Threading: From Mathematical Models to Parallel Implementations*, in "INFORMS Journal on Computing", 2004, vol. 16, n^{o} 4, p. 393-405 [*DOI :* 10.1287/IJOC.1040.0092], http://joc.journal.informs.org/content/16/4/393.abstract.

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

[3] R. CHIKHI, G. RIZK. *Space-efficient and exact de Bruijn graph representation based on a Bloom filter*, in "WABI 2012", Ljubljana, Slovenia, September 2012, vol. 7534, p. 236-248 [*DOI :* 10.1007/978-3-642-33122-0_19], http://hal.inria.fr/hal-00753930.

[4] F. LEGEAI, G. RIZK, T. WALSH, O. EDWARDS, K. GORDON, D. LAVENIER, N. LETERME, A. MEREAU, 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, n^{o} 1, 281 [*DOI :* 10.1186/1471-2164-11-281], http://www.hal.inserm.fr/inserm-00482283.

[5] A. MUCHERINO, C. LAVOR, L. LIBERTI, N. MACULAN. *The Discretizable Molecular Distance Geometry Problem*, in "Computational Optimization and Applications", 2012, vol. 52, p. 115-146, http://hal.inria.fr/hal-00756940.

[6] V. H. NGUYEN, D. LAVENIER. *PLAST: parallel local alignment search tool for database comparison*, in "Bmc Bioinformatics", October 2009, vol. 10, 329, http://hal.inria.fr/inria-00425301.

[7] P. PETERLONGO, R. CHIKHI. *Mapsembler, targeted and micro assembly of large NGS datasets on a desktop computer*, in "BMC Bioinformatics", March 2012, vol. 13, n^{o} 48 [*DOI :* 10.1186/1471-2105-13-48], http://hal.inria.fr/hal-00675974.

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

[9] G. A. T. SACOMOTO, J. KIELBASSA, R. CHIKHI, R. URICARU, P. ANTONIOU, M.-F. SAGOT, P. PETERLONGO, V. LACROIX. *KisSplice: de-novo calling alternative splicing events from RNA-seq data*, in "BMC Bioinformatics", March 2012, http://hal.inria.fr/hal-00681995.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] R. CHIKHI. *Computational methods for de novo assembly of next-generation genome sequencing data*, École normale supérieure de Cachan - ENS Cachan, July 2012, http://hal.inria.fr/tel-00752033.

Articles in International Peer-Reviewed Journals

[11] J. JAQUIÉRY, S. STOECKEL, P. NOUHAUD, L. MIEUZET, F. MAHÉO, F. LEGEAI, N. BERNARD, A. BONVOISIN, R. VITALIS, J. C. SIMON. *Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex.*, in "Molecular Ecology", November 2012, vol. 21, n^{o} 21, p. 5251-64 [*DOI :* 10.1111/MEC.12048], http://hal.inria.fr/hal-00753439.

[12] G. LE TRIONNAIRE, S. JAUBERT-POSSAMAI, J. BONHOMME, J.-P. GAUTHIER, G. GUERNEC, A. LE CAM, F. LEGEAI, J. MONFORT, D. TAGU. *Transcriptomic profiling of the reproductive mode switch in the pea aphid in response to natural autumnal photoperiod.*, in "Journal of Insect Physiology", December 2012, vol. 58, n^{o} 12, p. 1517–1524 [*DOI :* 10.1016/J.JINSPHYS.2012.07.009], http://hal.inria.fr/hal-00753429.

[13] P. LEROY, N. GUILHOT, H. SAKAI, A. BERNARD, F. CHOULET, S. THEIL, S. REBOUX, N. AMANO, T. FLUTRE, C. PELEGRIN, H. OHYANAGI, M. SEIDEL, F. GIACOMONI, M. REICHSTADT, M. ALAUX, E. GICQUELLO, F. LEGEAI, L. CERUTTI, H. NUMA, T. TANAKA, K. MAYER, T. ITOH, H. QUESNEVILLE, C. FEUILLET. *TriAnnot: A Versatile and High Performance Pipeline for the Automated Annotation of Plant Genomes.*, in "Front Plant Sci", 2012, vol. 3, 5 [*DOI :* 10.3389/FPLS.2012.00005], http://hal.inria.fr/hal-00753407.

[14] A. MUCHERINO, C. LAVOR, L. JON, J. LEE S., L. LIBERTI, M. SVIRIDENKO. *Discretization Orders for Distance Geometry Problems*, in "Optimization Letters", 2012, vol. 6, n^{o} 4, p. 783-796, http://hal.inria.fr/hal-00756941.

[15] A. MUCHERINO, C. LAVOR, L. LIBERTI, N. MACULAN. *Recent Advances on the Discretizable Molecular Distance Geometry Problem*, in "European Journal of Operational Research", 2012, vol. 219, p. 698-706, http://hal.inria.fr/hal-00756942.

[16] A. MUCHERINO, C. LAVOR, L. LIBERTI, N. MACULAN. *The Discretizable Molecular Distance Geometry Problem*, in "Computational Optimization and Applications", 2012, vol. 52, p. 115-146, http://hal.inria.fr/hal-00756940.

[17] A. MUCHERINO, C. LAVOR, L. LIBERTI. *Exploiting Symmetry Properties of the Discretizable Molecular Distance Geometry Problem*, in "Journal of Bioinformatics and Computational Biology", 2012, vol. 10, n^{o} 3, 1242009(1-15), http://hal.inria.fr/hal-00756939.

[18] A. MUCHERINO, C. LAVOR, L. LIBERTI. *The Discretizable Distance Geometry Problem*, in "Optimization Letters", 2012, vol. 6, n^{o} 8, p. 1671-1686, http://hal.inria.fr/hal-00756943.

[19] M. OLLIVIER, T. GABALDÓN, J. POULAIN, F. GAVORY, N. LETERME, J.-P. GAUTHIER, F. LEGEAI, D. TAGU, J. C. SIMON, C. RISPE. *Comparison of gene repertoires and patterns of evolutionary rates in eight aphid species that differ by reproductive mode.*, in "Genome Biol Evol", 2012, vol. 4, n^{o} 2, p. 155-67 [*DOI :* 10.1093/GBE/EVR140], http://hal.inria.fr/hal-00753402.

[20] P. PETERLONGO, R. CHIKHI. *Mapsembler, targeted and micro assembly of large NGS datasets on a desktop computer*, in "BMC Bioinformatics", March 2012, vol. 13, n^{o} 48 [*DOI :* 10.1186/1471-2105-13-48], http://hal.inria.fr/hal-00675974.

[21] G. A. T. SACOMOTO, J. KIELBASSA, R. CHIKHI, R. URICARU, P. ANTONIOU, M.-F. SAGOT, P. PETERLONGO, V. LACROIX. *KisSplice: de-novo calling alternative splicing events from RNA-seq data*, in "BMC Bioinformatics", March 2012, http://hal.inria.fr/hal-00681995.

[22] I. WOHLERS, R. ANDONOV, G. W. KLAU. *Optimal DALI protein structure alignment*, in "IEEE/ACM Transactions on Computational Biology and Bioinformatics", November 2012, 20, http://hal.inria.fr/hal-00685824.

[23] I. WOHLERS, N. MALOD-DOGNIN, R. ANDONOV, G. W. KLAU. *CSA: Comprehensive comparison of pairwise protein structure alignments*, in "Nucleic Acids Research", 2012, p. 303-309, Preprint, submitted to Nucleic Acids Research, http://hal.inria.fr/hal-00667920.

International Conferences with Proceedings

[24] R. CHIKHI, G. RIZK. *Space-efficient and exact de Bruijn graph representation based on a Bloom filter*, in "WABI 2012", Ljubljana, Slovenia, September 2012, vol. 7534, p. 236-248 [*DOI :* 10.1007/978-3-642-33122-0_19], http://hal.inria.fr/hal-00753930.

[25] N. MAILLET, C. LEMAITRE, R. CHIKHI, D. LAVENIER, P. PETERLONGO. *Compareads: comparing huge metagenomic experiments*, in "RECOMB Comparative Genomics 2012", Niterói, Brazil, October 2012, http://hal.inria.fr/hal-00720951.

[26] A. MUCHERINO, C. LAVOR, L. LIBERTI, N. MACULAN. *Finding Low-Energy Homopolymer Conformations by a Discrete Approach*, in "Global Optimization Workshop 2012 (GOW12)", Natal, Brazil, 2012, http://hal.inria.fr/hal-00756946.

[27] A. MUCHERINO, C. LAVOR, L. LIBERTI, N. MACULAN. *On the Discretization of Distance Geometry Problems*, in "Mathematics of Distances and Applications 2012 (MDA12)", Varna, Bulgaria, 2012, http://hal.inria.fr/hal-00756945.

[28] A. MUCHERINO, C. VIRGINIA, C. LUIZ MARIANO, C. LAVOR, N. MACULAN. *On Suitable Orders for Discretizing Molecular Distance Geometry Problems related to Protein Side Chains*, in "IEEE Conference Proceedings, Federated Conference on Computer Science and Information Systems (FedCSIS12), Workshop on Computational Optimization (WCO12)", Warsaw, Poland, 2012, http://hal.inria.fr/hal-00756944.

[29] A. MUCHERINO, G. WARLEY, C. LAVOR, N. MACULAN. *A Parallel BP Algorithm for the Discretizable Distance Geometry Problem*, in "Workshop on Parallel Computing and Optimization (PCO12), 26th IEEE International Parallel & Distributed Processing Symposium (IPDPS12)", Shanghai, China, 2012, http://hal.inria.fr/hal-00756947.

National Conferences with Proceeding

[30] M. LE BOUDIC-JAMIN, A. MUCHERINO, R. ANDONOV. *Modeling protein flexibility by distance geometry*, in "ROADEF 2012", Angers, France, Université d'Angers, 2012, http://hal.inria.fr/hal-00757717.

Conferences without Proceedings

[31] G. HOLLEY, P. PETERLONGO. *BlastGraph: intensive approximate pattern matching in string graphs and de-Bruijn graphs*, in "PSC 2012", Prague, Czech Republic, August 2012, http://hal.inria.fr/hal-00711911.