Curriculum Vitæ — Géraldine Jean

Personal Details

Name:	Jean Géraldine
Date and place of birth:	02/23/1982 in Cholet, France
Citizenship:	French
Address:	LaBRI, Université Bordeaux 1 351, Cours de la Libération F-33 405 Talence Cedex, FRANCE
Phone (work):	+33 540 00 36 11
Phone (cell):	+33 609 92 09 60
Email:	jean@labri.fr
Web:	http//www.labri.fr/∼jean

Education

2005–2008 **Ph.D. in Computer Science**
Thesis: *"In silico methods for genome rearrangement analysis: from identification of common markers to ancestral reconstruction"*
Advisers: Serge Dulucq and Macha Nikolski
LaBRI, University of Bordeaux - Talence, France

2004–2005 **M.Sc. in Computer Science**
Specialisation: *Models and Algorithms*
M.Sc. thesis: *"Genomic rearrangements: state of the art and constrained reversal distance"*
Adviser: Serge Dulucq
University of Bordeaux - Talence, France

2000–2004 **B.S. in Computer Science**
Faculty of Sciences and Techniques of Nantes - Nantes, France

Professional Experience

2008–2009 **Assistant Professor in Computer Science**
Computer Science department of IUT Bordeaux 1 - Bordeaux, France

2005–2008 **Teaching assistant in Computer Science**
Computer Science department of IUT Bordeaux - France
Computer Science department of University of Bordeaux - France
ENSEIRB, Engineer School of Computer Science - France

Competence

Systems:	GNU/Linux, Windows
Languages:	C, C++, Pascal, Shell, Visual Basic, Java, SQL, Prolog, Caml, XML/XSLT, python, perl, UML
Fields:	algorithm theory, combinatorics, graph theory, comparative genomics, bioinformatics
English:	fluent
Spanish:	basic
French:	native

References

Pr. Jean-Luc Souciet
Université de Strasbourg and CNRS
Institut de Botanique, 28 rue Goethe, 67000 Strasbourg, France
+ 33 3 90 24 18 17
jean-luc.souciet@gem.u-strasbg.fr

Dr. Macha Nikolski
LaBRI, Université Bordeaux 1
351 cours de la Libération, 33405 Talence Cedex, France
+33 5 40 00 35 06
macha@labri.fr

Pr. Bernard Moret
Laboratory for Computational Biology and Bioinformatics
EPFL, 1015 Lausanne, Switzerland
+41 21 69 31391
bernard.moret@epfl.ch

Publications

Journals

The Génolevures Consortium. *Comparative genomics of protoploid genomes of Saccharomycetaceae defines the orthologous gene set and basic yeast proteome repertoire*, Genome Research, online access ahead of print.

Géraldine Jean, David James Sherman, Macha Nikolski, *Mining the semantics of genome super-blocks to infer ancestral architectures*, Journal of Computational Biology, to appear.

Géraldine Jean, Macha Nikolski, *Genome rearrangements: a correct algorithm for optimal capping*, Information Processing Letters, Volume 104, Issue 1, Pages 14-20, October 2007.

Géraldine Jean, David James Sherman, Macha Nikolski, *SyDiG: Uncovering Synteny in Distant Genomes*, submitted to Int. J. of Bioinformatics Research and Applications.

Refereed conferences

Géraldine Jean, David James Sherman, Macha Nikolski, *Reconstruction and visualization of genome rearrangements within the Kluyveromyces*, ESF-EMBO Symposium on Comparative Genomics of Eukaryotic Microorganisms: Eukaryotic Genome Evolution, Sant Feliu de Guixols, Spain, 20-25 October 2007.

Workshops

Géraldine Jean, *Reconstruction and visualization of genome rearrangements within the Kluyveromyces*, First German/French/European meeting yeast and Filamentous Fungi, Strasbourg, France, 29-31 May 2008.

Géraldine Jean, Macha Nikolski, *Reconstruction of ancestral genomes within Hemiascomycetes*, Yeast Genome 10th anniversary, Académie des Sciences, Brussels, Belgium, 7-8 September 2006.

Research Activity

My research mainly concerns the definition of mathematical models describing the genome at the gene scale and in the development of combinatoric and algorithmic methods for genome comparison. Since the beginning of my PhD, I participate in the Génolevures Consortium (`http://genolevures.org`) in which I directly collaborate with biologists. This collaboration focuses on the application and validation of mathematical methods on real data from yeast genomes. My research interests are centered around three main themes:

- identification of common markers between several genomes,
- computation of rearrangement distance and parsimonious scenarios between two genomes,
- reconstruction of architecture of ancestral genomes.

Mathematical solutions for the problem of the reconstruction of ancestral genomes clearly depend on the sample of compared genomes but also on the identification of common markers in those genomes. This identification in its turn makes it possible to represent the genome by signed permutations. Several methods were developed in order to determine common markers between species such as GRIMM-Synteny (Pevzner and Tesler, 2003), which defines markers for rearrangement analysis. However, these methods are not adapted for yeast genomes I study. So I proposed a new algorithm SyDiG (Synteny in Distant Genomes) that processes complete genome sequences in order to infer cross-species synteny, and has the ability to handle species having a large evolutionary span. This method is based on ADHoRe (Vandepoele et al., 2002) which provides pairwise homologies even for highly divergent chromosomic regions. Then, the algorithm consists in extending certain homologies by transitivity. Algorithm SyDiG and its results on yeast genomes were presented during one international conference and one international workshop and is currently submitted to the journal *Int. J. of Bioinformatics Research and Applications* (IJBRA).

Reconstruction of evolutionary scenarios based on genomic rearrangements makes possible the understanding of evolution of species. For inferring such scenarios *in silico*, genomes are modelled by signed permutations where each element of the permutation represents a common marker. Following parsimony principle, the problem consists in quantifying the minimum number of rearrangements (distance computation) and determining which are the operations on permutations that transform a genome into another (scenarios). *The sorting signed permutations by reversals problem* introduced by Sankoff (Sankoff, 1992) was largely studied in literature and efficient algorithms for unichromosomal and multichromosomal cases have been known since (Hannenhalli and Pevzner, 1995). During my PhD, I studied in detail the Hannenhalli and Pevzner theory for the multichromosomal case as well as its successive corrections by Tesler (Tesler, 2002) and Ozery-Flato (Ozery-Flato and Shamir, 2003) for the computation of distance and a parsimonious scenario. I proposed a single and coherent classification of the notions involved in existing algorithms for solving these two problems. This classification makes it possible to pinpoint the fact that current algorithms present errors. I introduced a correct algorithm for optimal capping with its proof of correction (Jean and Nikolski, 2007).

The study of genome evolution by comparison of contemporary genomes is frustrated by the impossibility of knowing with certainty the architecture of the common ancestral genomes. Constructing plausible hypotheses about these ancestral genomes is a combinatoric problem whose results may provide deep insight both into the past histories of particular genomes and the general mechanisms of their formation. Reconciliating by *in silico* methods contemporary genomes into an ancestral architecture is formulated as *the multiple genome rearrangement problem* (Hannenhalli et al., 1995) (Sankoff et al., 1996): given a set of N contemporary genomes $G_1, .., G_N$ and a distance d, find a tree T with the N genomes as leaf nodes and assign permutations (plausible ancestral architectures) to internal nodes such that $D(T) = \sum_{(\pi,\gamma)\in T} d(\pi, \gamma)$ is minimized. Methods were developed according to different distances (breakpoint distance (Sankoff and Blanchette, 1997), reversal distance (Caprara, 1999) and (Caprara, 2003), rearrangement distance(Bourque and Pevzner, 2002)). In both cases, *the multiple genome rearrangement problem* for $N = 3$ is NP-hard (see (Bryant, 1998) and (Pe'er and Shamir, 1998) for breakpoint distance; (Caprara, 1999) and (Caprara, 2003) for reversal distance). In addition to the computational intractability of this problem, these *in silico* methods provide one single global solution chosen among a multitude of equivalent ones (Eriksen, 1997), that, furthermore, can highly diverge. Given a unique global ancestral architecture is biologically misleading. A more realistic approach is to consider what common structural features of ancestral genomes might be found. I developed with Macha Nikolski a new piece-wise method for the reconstruction of ancestral architectures based on the study of both adjacencies between common markers and rearrangement distances between modern genomes. Moreover, it makes it possible to use biological constraints such as centromere position. This leads to the construction of the set of partial assemblies in which each element, called *super-block*, represents a common ancestral feature. This work is going to appear in *Journal of Computational Biology*.