Lecture Notes in Computer Science 2066

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Olivier Gascuel Marie-France Sagot (Eds.)

Computational Biology

First International Conference on Biology,
Informatics, and Mathematics, JOBIM 2000
Montpellier, France, May 3–5, 2000
Selected Papers

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Olivier Gascuel
Laboratoire d'Informatique, de Robotique et de Microelectronique de Montpellier
161 rue Ada, 34392 Montpellier Cedex 5, France
E-mail: gascuel@lirmm.fr

Marie-France Sagot
Institut Pasteur, Laboratoire d'Algorithmique Combinatoire
28, rue du Dr. Roux, 75724 Paris Cedex 15, France
E-mail: sagot@pasteur.fr

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Computational biology : selected papers / First International Conference on Biology, Informatics, and Mathematics, JOBIM 2000, Montpellier, France, May 3 - 5, 2000. Olivier Gascuel ; Marie-France Sagot (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2001
(Lecture notes in computer science ; Vol. 2066)
ISBN 3-540-42242-0

CR Subject Classification (1998): F.2.2, E.4, E.1, G.2, J.3

ISSN 0302-9743
ISBN 3-540-42242-0 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

© Springer-Verlag Berlin Heidelberg 2001
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna
Printed on acid-free paper SPIN: 10781721 06/3142 5 4 3 2 1 0

Preface

The papers contained in this volume were presented at the first annual JOBIM conference ("Journées Ouvertes: Biologie, Informatique et Mathématique" – Open Days: Biology, Informatics and Mathematics), held May 3–5, 2000, at the ENSA-M ("École Nationale Supérieure d'Agronomie de Montpellier"), Montpellier, France. The beautiful park-like ENSA-M campus contains trees that are several hundred years old, and is only a ten minute walk from the center of the old town.

The idea of JOBIM was born two years ago, at the Place Bellecour in Lyon. We were strolling towards one of Lyon's fine restaurants. It was late one evening in April, during the RECOMB'98 conference. It seemed important to us to organize a Francophone conference that would be open to all people working at the frontier between biology, computer science, mathematics, and physics, and interested in the analysis, comparison, and exploitation of genomic data. JOBIM covers word algorithms, comparative genomics, evolution and phylogeny, data and knowledge bases, functional analysis, genome annotation, graph theory and combinatorics applied to biology, macromolecular structures (RNA and proteins), metabolic pathways and regulatory networks, statistics and classification.

We wanted JOBIM to be an annual meeting which would take place in surroundings facilitating exchanges between people, and where younger researchers in particular would be encouraged to present their work. The idea was greeted very favourably, and a program committee was rapidly set up. The initiative was well timed because at the start of the summer of 1999 the French Ministry of Research, through the Genome program, had decided to give new incentives to work in this interdisciplinary area. The IMPG project (Informatics, Mathematics, and Physics for Genomics) was created as a consequence of that decision.

The IMPG project provided an important part of the funding for JOBIM, established a precise scientific context for the conference, and assisted with publicity. Following the call for papers, we received 67 contributions. We chose to select as many papers as possible. There were 60 presentations in total, 20 long presentations for works of general interest and 40 short presentations for specialized contributions. This approach enabled a large number of researchers to present and discuss their results. There were also five invited talks, from Philipp Bucher, Christian Gautier, Gene Myers, David Sankoff, and Mike Steel. The presence of such guests contributed greatly to the scientific interest of the conference and all are gratefully thanked.

A second much stricter selection was then made for the papers that would appear in an English version of the proceedings. These are the 13 papers the reader will find in this volume. The subjects range over as broad an area as the 60 originally selected.

JOBIM was an unexpectedly huge success, with around 330 participants, showing that the field is rapidly maturing in France and the other French-speaking

countries. This includes Belgium, Canada, and Switzerland. The next JOBIM will take place in Toulouse, in the south of France, May 30 to June 1, 2001. It is hoped that JOBIM 2001 will attract an even greater number of participants than the first JOBIM.

JOBIM 2000 was organized by the LIRMM at the University of Montpellier II. The Local Organizing Committee of JOBIM 2000 was headed by Gilles Caraux. The efforts of all, and general support of the French computational biology and bioinformatics community, are gratefully acknowledged.

March 2001

Olivier Gascuel
Marie-France Sagot
Program Chairs

Program Committee

Joël Alexandre
Philipp Bucher
Maxime Crochemore
Pierre Darlu
Laurent Duret
Patrick Forterre
Olivier Gascuel
Jean-François Gibrat
Alain Guénoche
Bernard Jacq
Hans-Peter Lenhof
François Major
Hervé Philippe
François Rechenmann
Éric Rivals
Claude Thermes
Alain Viari
Éric Westhof
Vincent Berry
Dominique Cellier
Antoine Danchin
Gilbert Deléage
Gwenaëlle Fichant
Nicolas Galtier
Christine Gaspin
Manolo Gouy
Jacques Haiech
Richard Lavery
Arthur Lesk
Eugene W. Myers
Bernard Prum
Jean-Loup Risler
Pierre Rouzé
Marie-France Sagot
Tandy Warnow

Organizing Committee

Gilles Caraux – OC Chair
Vincent Berry
Éric Rivals
Philippe Vismara
Isabelle Mougenot
Véronique Sals-Vettorel
Corine Ziegler

Additional referees for the LNCS volume

Andreas Dress
Alain Jean-Marie
Bruno Leclerc
Jacques Nicolas
Mike Steel
Roderic Guigó
Ned Lamb
Vincent Moulton
Cecilia Saccone
Denis Thieffry

Table of Contents

Speeding Up the DIALIGN Multiple Alignment Program by Using the 'Greedy Alignment of BIOlogical Sequences LIBrary' (GABIOS-LIB) 1
Saïd Abdeddaïm, Burkhard Morgenstern

GeMCore, a Knowledge Base Dedicated to Mapping Mammalian Genomes 12
Gisèle Bronner, Bruno Spataro, Christian Gautier, François Rechenmann

Optimal Agreement Supertrees 24
David Bryant

Segmentation by Maximal Predictive Partitioning According to Composition Biases 32
Laurent Guéguen

Can We Have Confidence in a Tree Representation? 45
Alain Guénoche, Henri Garreta

Bayesian Approach to DNA Segmentation into Regions with Different Average Nucleotide Composition 57
Vsevolod Makeev, Vasily Ramensky, Mikhail Gelfand, Mikhail Roytberg, Vladimir Tumanyan

Exact and Asymptotic Distribution of the Local Score of One i.i.d. Random Sequence 74
Sabine Mercier, Dominique Cellier, François Charlot, Jean-Jacques Daudin

Phylogenetic Reconstruction Algorithms Based on Weighted 4-Trees 84
Vincent Ranwez, Olivier Gascuel

Computational Complexity of Word Counting 99
Mireille Régnier

EuGène: An Eukaryotic Gene Finder That Combines Several Sources of Evidence 111
Thomas Schiex, Annick Moisan, Pierre Rouzé

Tree Reconstruction via a Closure Operation on Partial Splits 126
Charles Semple, Mike Steel

InterDB, a Prediction-Oriented Protein Interaction Database for *C. elegans* 135
Nicolas Thierry-Mieg, Laurent Trilling

Application of Regulatory Sequence Analysis and Metabolic Network Analysis to the Interpretation of Gene Expression Data 147
Jacques van Helden, David Gilbert, Lorenz Wernisch, Michael Schroeder, Shoshana Wodak

Author Index **165**