Lecture Notes in Bioinformatics 4360

Edited by S. Istrail, P. Pevzner, and M. Waterman

Subseries of Lecture Notes in Computer Science

Werner Dubitzky Assaf Schuster
Peter M.A. Sloot Michael Schroeder
Mathilde Romberg (Eds.)

Distributed, High-Performance and Grid Computing in Computational Biology

International Workshop, GCCB 2006
Eilat, Israel, January 21, 2007
Proceedings

Springer

Volume Editors

Werner Dubitzky
Mathilde Romberg

University of Ulster, School of Biomedial Sciences, Coleraine, UK
E-mail: {w.dubitzky, me.romberg}@ulster.ac.uk

Assaf Schuster
Technion – Israel Institute of Technology, Haifa, Israel
E-mail: assaf@cs.technion.ac.il

Peter M.A. Sloot
University of Amsterdam, Section of Computational Science, The Netherlands
E-mail: sloot@science.uva.nl

Michael Schroeder
Biotec, Dresden University of Technology, Germany
E-mail: Michael.Schroeder@biotec.tu-dresden.de

Library of Congress Control Number: 2006939917

CR Subject Classification (1998): F.1, F.2, J.3, D.1, D.2, D.4, C.2, C.4, H.4

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-540-69841-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-69841-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11976158 06/3142 5 4 3 2 1 0

Preface

This volume of the Springer *Lecture Notes in Computer Science* series contains the contributions presented at the International Workshop on Distributed, High-Performance and Grid Computing in Computational Biology 2006 (GCCB 2006) held in Eilat, January 21, 2007 in conjunction with the fifth European Conference on Computational Biology (ECCB 2006).

Modern computational biology and bioinformatics are characterized by large and complex-structured data and by applications requiring considerable computing resources, such as processing units, storage elements and software programs. In addition, these disciplines are intrinsically geographically distributed in terms of their instruments, communities and computing resources. Tackling the computational challenges in computational biology and bioinformatics increasingly requires high-end and distributed computing infrastructures, systems and tools. The main objective of this workshop is to bring together researchers and practitioners from these areas to discuss ideas and experiences in developing and applying distributed, high-performance and grid computing technology to problems in computational biology and bioinformatics.

The challenges in distributed, high-performance and grid computing in biology and biotechnology are inherently more complicated than those in such domains as physics, engineering and conventional business areas. Some of the added complexities arise from the:

- Conceptual complexity of biological knowledge and the methodologies used in biology and biotechnology
- Need to understand biological systems and processes at a detailed mechanistic, systemic and quantitative level across several levels of organization (ranging from molecules to cells, populations, and the environment)
- Growing availability of high-throughput data from genomics, transcriptomics, proteomics, metabolomics and other high-throughput methods
- Widespread use of image data in biological research and development (microscopy, NMR, MRI, PET, X-ray, CT, etc.)
- Increasing number of investigations studying the properties and dynamic behavior of biological systems and processes using computational techniques (molecular dynamics, QSAR/QSPR, simulation of gene-regulatory, signaling and metabolic networks, protein folding/unfolding, etc)
- Requirement to combine data, information and compute services (e.g., sequence alignments) residing on systems that are distributed around the world
- Variety of different technologies, instruments, infrastructures and systems used in life science R&D
- Huge variety of information formats and frequent addition of new formats arising from new experimental protocols, instruments and phenomena to be studied

- Large and growing number of investigated biological and biomedical phenomena
- Fact that life science R&D is based heavily on the use of distributed and globally accessible computing resources (databases, knowledge bases, model bases, instruments, text repositories, compute-intensive services)

The GCCB workshop brought together computational biologists, bioinformaticians and life scientists who have researched and applied distributed, high-performance and grid computing technologies in the context of computational biology and bioinformatics. The workshop discussed innovative work in progress and important new directions. By sharing the insights, discussing ongoing work and the results that have been achieved, we hope the workshop participants conveyed a comprehensive view of the state of the art in this area and identified emerging and future research issues. We believe that the GCCB workshop made a valuable contribution in encouraging and shaping future work in the field of distributed, high-performance and grid computing in computational biology.

Acknowledgements

GCCB 2006 was sponsored and organized by the University of Ulster, Coleraine, UK; the Technion - Israel Institute of Technology, Haifa, Israel; the University of Amsterdam, Amsterdam, The Netherlands; the Dresden University of Technology, Dresden, Germany; and the DataMiningGrid Consortium (EC grant IST-2004-004475, www.DataMiningGrid.org). We are indebted to the ECCB 2006 Conference organizers Hershel Safer, Weizmann Institute of Science, and Haim Wolfson, Tel Aviv University. Finally, we would like to extend our gratitude to the members of the GCCB 2006 International Program Committee and to those who helped in the review process. And last but not least we want to thank all authors and workshop participants for their contributions and the valuable discussions.

January 2007

Werner Dubitzky
Assaf Schuster
Peter Sloot
Michael Schroeder
Mathilde Romberg

Organization

Program Chairs

Werner Dubitzky, University of Ulster, Coleraine, UK
Assaf Schuster, Technion - Israel Institute of Technology, Haifa, Israel
Peter M.A. Sloot, University of Amsterdam, Faculty of Sciences, Amsterdam, The Netherlands
Michael Schroeder, Dresden University of Technology, Biotechnological Centre, Dresden, Germany
Mathilde Romberg, University of Ulster, Coleraine, UK

Program Committee

David A. Bader, Georgia Tech, College of Computing, Atlanta, Georgia, USA
Eric Bremer, Children's Memorial Hospital, Northwestern University, Chicago, USA
Rui M. Brito, Universidade de Coimbra, Coimbra, Portugal
Marian Bubak, AGH - University of Science and Technology, Krakow, Poland
Kevin Burrage, The University of Queensland, Australia
Gustavo Deco, Universitat Pompeu Fabra, Barcelona, Spain
Frank Dehne, University of Ottawa, Ottawa, Canada
Guiseppe di Fatta, University of Konstanz, Konstanz, Germany
Werner Dubitzky, University of Ulster, Coleraine, UK
Jordi Vill i Freixa, Universitat Pompeu Fabra, Barcelona, Spain
David Gilbert, University of Glasgow, Glasgow, UK
Carol Goble, University of Manchester, Manchester, UK
Danilo Gonzalez, Universidad de Talca, Talca, Chile
Ulrich Hansmann, Forschungszentrum Juelich, NIC, Juelich, Germany
Des Higgins, University College Dublin, Conway Institute, Dublin, Ireland
Alfons Hoekstra, University of Amsterdam, Amsterdam, The Netherlands
Martin Hoffmann, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Sankt Augustin, Germany
Rod Hose, University of Sheffield, Sheffield, UK
Chun-Hsi (Vincent) Huang, University of Connecticut, Storrs, USA
Akihiko Konagaya, Riken Genomic Sciences Center, Yokohama City, Japan
Miron Livny, University of Wisconsin at Madison, Wisconsin, USA
Uko Maran, University of Tartu, Tartu, Estonia
Hartmut Mix, Dresden University of Technology, Dresden, Germany
Ron Perrot, Queens University, Belfast, UK
Mark Ragan, The University of Queensland, Australia

Stephen Robinson, University of Ulster, Coleraine, UK
Mathilde Romberg, University of Ulster, Coleraine, UK
Michael Schroeder, Dresden University of Technology, Biotechnological Centre, Dresden, Germany
Assaf Schuster, Technion - Israel Institute of Technology, Haifa, Israel
Gadi Schuster, Technion - Israel Institute of Technology, Haifa, Israel
Richard Sinnot, University of Glasgow, UK
Peter Sloot, University of Amsterdam, Faculty of Sciences, Amsterdam, The Netherlands
Craig A. Stewart, Indiana University, Indiana, USA
Domenico Talia, Università della Calabria, DEIS, Rende, Italy
Albert Zomaya University of Sydney, Sydney, Australia

External Reviewers

Pinar Alper, University of Manchester, Manchester, UK
Michael Johnston, Universitat Pompeu Fabra, Barcelona, Spain
David Jones, University of Manchester, Manchester, UK
Peter Kral, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Sankt Augustin, Germany
Dean Kuo, University of Manchester, Manchester, UK
Jan Meinke, Forschungszentrum Juelich, NIC, Juelich, Germany
Javier Tamames, Universitat Pompeu Fabra, Barcelona, Spain

Table of Contents

Session 1a. "Sequence Analysis"

Session 1b. "Grids for Screening and Property Prediction"

Session 2a. "Data Management"

Session 2b. "Collaborative Environments"