

German Conference on Bioinformatics 2013

GCB'13, September 10–13, 2013, Göttingen, Germany

Edited by

Tim Beißbarth

Martin Kollmar

Andreas Lehä

Burkhard Morgenstern

Anne-Kathrin Schultz

Stephan Waack

Edgar Wingender



Editors

Tim Beißbarth
Department of Medical Statistics
University Medical Center Göttingen
Tim.Beissbarth@med.uni-goettingen.de

Martin Kollmar
NMR Based Structural Biology
MPI for Biophysical Chemistry, Göttingen
mako@nmr.mpibpc.mpg.de

Andreas Leha
Department of Medical Statistics
University Medical Center Göttingen
andreas.leha@med.uni-goettingen.de

Burkhard Morgenstern
Department of Bioinformatics (IMG)
University of Göttingen
bmorgen@gwdg.de

Anne-Kathrin Schultz
Department of Bioinformatics (IMG)
University of Göttingen
aschult2@gwdg.de

Stephan Waack
Institute of Computer Science
University of Göttingen
waack@informatik.uni-goettingen.de

Edgar Wingender
Institute of Bioinformatics
University Medical Center Göttingen
edgar.wingender@bioinf.med.uni-goettingen.de

ACM Classification 1998
J.3 Life and Medical Sciences

ISBN 978-3-939897-59-0

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern,
Germany. Online available at <http://www.dagstuhl.de/dagpub/978-3-939897-59-0>.

Publication date
September, 2013

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed
bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

License
This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0):
<http://creativecommons.org/licenses/by/3.0/legalcode>.



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work
under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/OASlcs.GCB.2013.i

ISBN 978-3-939897-59-0

ISSN 2190-6807

<http://www.dagstuhl.de/oasics>

OASlcs – OpenAccess Series in Informatics

OASlcs aims at a suitable publication venue to publish peer-reviewed collections of papers emerging from a scientific event. OASlcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Daniel Cremers (TU München, Germany)
- Barbara Hammer (Universität Bielefeld, Germany)
- Marc Langheinrich (Università della Svizzera Italiana – Lugano, Switzerland)
- Dorothea Wagner (*Editor-in-Chief*, Karlsruher Institut für Technologie, Germany)

ISSN 2190-6807

www.dagstuhl.de/oasics

■ Contents

On the estimation of metabolic profiles in metagenomics <i>Kathrin Petra Aßhauer and Peter Meinicke</i>	1
On Weighting Schemes for Gene Order Analysis <i>Matthias Bernt, Nicolas Wieseke, and Martin Middendorf</i>	14
Alignment-free sequence comparison with spaced k -mers <i>Marcus Boden, Martin Schöneich, Sebastian Horwege, Sebastian Lindner, Chris Leimeister, and Burkhard Morgenstern</i>	24
PanCake: A Data Structure for Pangenomes <i>Corinna Ernst and Sven Rahmann</i>	35
Reconstructing Consensus Bayesian Network Structures with Application to Learning Molecular Interaction Networks <i>Holger Fröhlich and Gunnar W. Klau</i>	46
Efficient Interpretation of Tandem Mass Tags in Top-Down Proteomics <i>Anna Katharina Hildebrandt, Ernst Althaus, Hans-Peter Lenhof, Chien-Wen Hung, Andreas Tholey, and Andreas Hildebrandt</i>	56
GEDEVO: An Evolutionary Graph Edit Distance Algorithm for Biological Network Alignment <i>Rashid Ibragimov, Maximilian Malek, Jiong Guo, and Jan Baumbach</i>	68
Dinucleotide distance histograms for fast detection of rRNA in metatranscriptomic sequences <i>Heiner Klingenberg, Robin Martinjak, Frank Oliver Glöckner, Rolf Daniel, Thomas Lingner, and Peter Meinicke</i>	80
Utilization of ordinal response structures in classification with high-dimensional expression data <i>Andreas Leha, Klaus Jung, and Tim Beißbarth</i>	90
Extended Sunflower Hidden Markov Models for the recognition of homotypic <i>cis</i> -regulatory modules <i>Ioana M. Lemnian, Ralf Eggeling, and Ivo Grosse</i>	101
Avoiding Ambiguity and Assessing Uniqueness in Minisatellite Alignment <i>Benedikt Löwes and Robert Giegerich</i>	110
Aligning Flowgrams to DNA Sequences <i>Marcel Martin and Sven Rahmann</i>	125

■ Preface

This proceedings volume contains original research papers presented at the German Conference on Bioinformatics 2013 (GCB'13) held at Georg-August-University, Göttingen, Germany, September 11–13, 2013.

The GCB is an annual, international conference devoted to all areas of bioinformatics. Recent meetings attracted a multinational audience with 250 – 300 participants each year.

GCB'13 is organized by the bioinformatics groups at *Göttingen Research Campus* in cooperation with the *German Society for Chemical Engineering and Biotechnology (DE-CHEMA)*, the *Society for Biochemistry and Molecular Biology (GBM)* and the *Special Interest Group on Informatics in Biology* of the *German Society of Computer Science (GI)*.

Five internationally renowned speakers agreed to give keynote talks at GCB'13: Manfred Eigen, Gene Myers, Erwin Neher, Terry Speed and Sarah Teichmann. Four satellite workshops were held on 10 September 2013 on *Statistical Methods in Bioinformatics*, *Computational Methods for Metagenomics and Meta-Omics*, *Alignment-Free Sequence Comparison* and *Methods for Integrated Analysis of Multi-Level Datasets*.

Submissions to GCB'13 were possible as *Regular Papers*, *i.e.* original research papers, *Highlight Papers*, usually reporting on work published during the last year, or poster abstracts.

Overall, we received 26 submissions for *Regular Papers* and 19 Submissions for *Highlight Papers*. After a careful reviewing procedure and discussions in the *Program Committee*, 12 out of the 26 *Regular* submissions and 8 out of the 19 *Highlight* submission were selected for oral presentation at the conference. This proceedings volume contains revised versions of the 12 selected *Regular Papers*.

We would like to thank all authors, members of the *Program Committee* and subreviewers as well as the members of the local *Organizing Committee* and the support team for their work. In particular, we are indebted to Dr. Anne-Kathrin Schultz for doing most of the organization work for GCB'13. We thank Andreas Leha for organizing the production of this proceedings volume and Britta Leinemann for administrative support.

Göttingen, September 2013

Burkhard Morgenstern and Edgar Wingender

■ Program Committee

Program Chairs

Burkhard Morgenstern
Edgar Wingender

Program Committee

Mario Albrecht	Christoph Kaleta	Matthias Rarey
Rolf Backofen	Gunnar W. Klau	Knut Reinert
Jan Baumbach	Ina Koch	Uwe Scholz
Michael Beckstette	Oliver Kohlbacher	Dietmar Schomburg
Niko Beerenwinkel	Martin Kollmar	Falk Schreiber
Tim Beissbarth	Antje Krause	Michael Schroeder
Sebastian Böcker	Stefan Kurtz	Stefan Schuster
Erich Bornberg-Bauer	Thomas Lengauer	Torsten Schwede
Thomas Dandekar	Hans-Peter Lenhof	Joachim Selbig
Andreas Dress	Thomas Lingner	Rainer Spang
Mareike Fischer	Manja Marz	Peter Stadler
Dmitrij Frishman	Alice Mchardy	Mario Stanke
Holger Froehlich	Peter Meinicke	Jens Stoye
Georg Fuellen	Irmtraud Meyer	Martin Vingron
Robert Giegerich	Axel Mosig	Arndt Von Haeseler
Ivo Grosse	Eugene Myers	Stephan Waack
Volker Heun	Steffen Neumann	Thomas Werner
Andreas Hildebrandt	Kay Nieselt	Ralf Zimmer
Daniel Huson	Sven Rahmann	

Additional Referees

Volker Helms	Reinhard Guthke	Patrick Trampert
Dirk Willrodt	Walton White	Sascha Winter
Alexander Kel	Kousik Kundu	Anne Hildebrandt
Michaela Bayerlova	Christian Colmsee	Eva Grafahrend-Belau
Michael Love	Martin Engler	Christoph Kaleta
Frank Kramer	Sascha Steinbiss	Jochen Singer
Juliane Siebourg-Polster	Dragos Sorescu	Tobias Petri
Anja Hartmann	Jonathan Goeke	Anne-Christin Hauschild

■ Supporters and Sponsors

Supporting Scientific Institutions



DECHEMA Gesellschaft für Chemische Technik
und Biotechnologie e.V.
<http://www.dechema.de>



GBM Gesellschaft für Biochemie und
Molekularbiologie e.V.
<http://www.gbm-online.de>



Fachgruppe "Informatik in den Biowissenschaften"
der GI
<http://www.cebitec.uni-bielefeld.de/groups/fg402>



Max-Planck-Institute for Biophysical Chemistry
<http://www.mpibpc.mpg.de>



University of Göttingen
<http://www.uni-goettingen.de>



University Medical Center Göttingen
<http://www.med.uni-goettingen.de>



GWDG: IT in der Wissenschaft
<http://www.gwdg.de>

German Conference on Bioinformatics 2013 (GCB'13).

Editors: T. Beißbarth, M. Kollmar, A. Leha, B. Morgenstern, A.-K. Schultz, S. Waack, E. Wingender
OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Sponsors and Donors



geneXplain: From genes to drugs
<http://genexplain.com>



KWS: Saatgutspezialisten für Landwirte
<http://www.kws.de>



Speise- & Schankwirtschaft Bullerjahn
<http://www.bullerjahn.info>



MoBiTec: Innovative Tools for Molecular and Cell Biology
<http://www.mobitec.com>

■ Index of Authors

A	
Althaus, Ernst	56
Aßhauer, Kathrin	1
B	
Baumbach, Jan	68
Beißbarth, Tim	90
Bernt, Matthias	14
Boden, Marcus	24
D	
Daniel, Rolf	80
E	
Eggeling, Ralf	101
Ernst, Corinna	35
F	
Fröhlich, Holger	46
G	
Giegerich, Robert	110
Glöckner, Frank	80
Grosse, Ivo	101
Guo, Jiong	68
H	
Hildebrandt, Andreas	56
Hildebrandt, Anna	56
Horwege, Sebastian	24
Hung, Chien-Wen	56
I	
Ibragimov, Rashid	68
J	
Jung, Klaus	90
K	
Klau, Gunnar	46
Klingenberg, Heiner	80
L	
Löwes, Benedikt	110
Leha, Andreas	90
Leimeister, Chris	24
Lemnian, Ioana	101
Lenhof, Hans-Peter	56
Lindner, Sebastian	24
Lingner, Thomas	80
M	
Malek, Maximilian	68
Martin, Marcel	125
Martinjak, Robin	80
Meinicke, Peter	1, 80
Middendorf, Martin	14
Morgenstern, Burkhard	24
R	
Rahmann, Sven	35, 125
S	
Schöneich, Martin	24
T	
Tholey, Andreas	56
W	
Wieseke, Nicolas	14