

17th International Workshop on Algorithms in Bioinformatics

WABI 2017, August 21–23, 2017, Boston, MA, USA

Edited by

Russell Schwartz

Knut Reinert



Editors

Russell Schwartz	Knut Reinert
Carnegie Mellon University	Freie Universität Berlin
Pittsburgh, USA	Berlin, Germany
russells@andrew.cmu.edu	knut.reinert@fu-berlin.de

ACM Classification 1998

F.2.2 Nonnumerical Algorithms and Problems, F.2.2 Pattern Matching, I.1.2 Algorithms, J.3 Life and Medical Sciences

ISBN 978-3-95977-050-7

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/978-3-95977-050-7>.

Publication date

August, 2017

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/LIPIcs.WABI.2017.0

ISBN 978-3-95977-050-7

ISSN 1868-8969

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

LIPICs – Leibniz International Proceedings in Informatics

LIPICs is a series of high-quality conference proceedings across all fields in informatics. LIPICs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Luca Aceto (Reykjavik University)
- Susanne Albers (TU München)
- Chris Hankin (Imperial College London)
- Deepak Kapur (University of New Mexico)
- Michael Mitzenmacher (Harvard University)
- Madhavan Mukund (Chennai Mathematical Institute)
- Anca Muscholl (University Bordeaux)
- Catuscia Palamidessi (INRIA)
- Raimund Seidel (Saarland University and Schloss Dagstuhl – Leibniz-Zentrum für Informatik)
- Thomas Schwentick (TU Dortmund)
- Reinhard Wilhelm (Saarland University)

ISSN 1868-8969

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

■ Contents

Disentangled Long-Read De Bruijn Graphs via Optical Maps <i>Bahar Alipanahi, Leena Salmela, Simon J. Puglisi, Martin Muggli, and Christina Boucher</i>	1:1–1:14
Gene Tree Parsimony for Incomplete Gene Trees <i>Md. Shamsuzzoha Bayzid and Tandy Warnow</i>	2:1–2:13
Better Greedy Sequence Clustering with Fast Banded Alignment <i>Brian Brubach, Jay Ghurye, Mihai Pop, and Aravind Srinivasan</i>	3:1–3:13
Optimal Computation of Overabundant Words <i>Yannis Almirantis, Panagiotis Charalampopoulos, Jia Gao, Costas S. Iliopoulos, Manal Mohamed, Solon P. Pissis, and Dimitris Polychronopoulos</i>	4:1–4:14
Detecting Locus Acquisition Events in Gene Trees <i>Michał Aleksander Ciach, Anna Muszewska, and Paweł Górecki</i>	5:1–5:13
An IP Algorithm for RNA Folding Trajectories <i>Amir H. Bayegan and Peter Clote</i>	6:1–6:16
Fast Spaced Seed Hashing <i>Samuele Girotto, Matteo Comin, and Cinzia Pizzi</i>	7:1–7:14
A General Framework for Gene Tree Correction Based on Duplication-Loss Reconciliation <i>Nadia El-Mabrouk and Aïda Ouangraoua</i>	8:1–8:14
Towards Distance-Based Phylogenetic Inference in Average-Case Linear-Time <i>Maxime Crochemore, Alexandre P. Francisco, Solon P. Pissis, and Cátia Vaz</i>	9:1–9:14
Yanagi: Transcript Segment Library Construction for RNA-Seq Quantification <i>Mohamed K. Gunady, Steffen Cornwell, Stephen M. Mount, and Héctor Corrada Bravo</i>	10:1–10:14
Shrinkage Clustering: A Fast and Size-Constrained Algorithm for Biomedical Applications <i>Chenyue W. Hu, Hanyang Li, and Amina A. Qutub</i>	11:1–11:13
Sparsification Enables Predicting Kissing Hairpin Pseudoknot Structures of Long RNAs in Practice <i>Hosna Jabbari, Ian Wark, Carlo Montemagno, and Sebastian Will</i>	12:1–12:13
Vaquita: Fast and Accurate Identification of Structural Variation Using Combined Evidence <i>Jongkyu Kim and Knut Reinert</i>	13:1–13:14
Assessing the Significance of Peptide Spectrum Match Scores <i>Anastasiia Abramova and Anton Korobeynikov</i>	14:1–14:11
abSNP: RNA-Seq SNP Calling in Repetitive Regions via Abundance Estimation <i>Shunfu Mao, Soheil Mohajer, Kannan Ramachandran, David Tse, and Sreeram Kannan</i>	15:1–15:14



<i>All Fingers Are Not the Same: Handling Variable-Length Sequences in a Discriminative Setting Using Conformal Multi-Instance Kernels</i> <i>Sarvesh Nikumbh, Peter Ebert, and Nico Pfeifer</i>	16:1–16:14
Forbidden Time Travel: Characterization of Time-Consistent Tree Reconciliation Maps <i>Nikolai Nøjgaard, Manuela Geiß, Daniel Merkle, Peter F. Stadler, Nicolas Wieseke, and Marc Hellmuth</i>	17:1–17:12
Rainbowfish: A Succinct Colored de Bruijn Graph Representation <i>Fatemeh Almodaresi, Prashant Pandey, and Rob Patro</i>	18:1–18:15
ThIEF: Finding Genome-wide Trajectories of Epigenetics Marks <i>Anton Polishko, Md. Abid Hasan, Weihua Pan, Evelien M. Bunnik, Karine Le Roch, and Stefano Lonardi</i>	19:1–19:16
Byte-Aligned Pattern Matching in Encoded Genomic Sequences <i>Petr Procházka and Jan Holub</i>	20:1–20:13
Analysis of Min-Hashing for Variant Tolerant DNA Read Mapping <i>Jens Quedenfeld and Sven Rahmann</i>	21:1–21:13
Efficient and Accurate Detection of Topologically Associating Domains from Contact Maps <i>Abbas Roayaei Ardakany and Stefano Lonardi</i>	22:1–22:11
Outlier Detection in BLAST Hits <i>Nidhi Shah, Stephen F. Altschul, and Mihai Pop</i>	23:1–23:11
Finding Local Genome Rearrangements <i>Pijus Simonaitis and Krister M. Swenson</i>	24:1–24:13
Seed-driven Learning of Position Probability Matrices from Large Sequence Sets <i>Jarkko Toivonen, Jussi Taipale, and Esko Ukkonen</i>	25:1–25:13
Improved De Novo Peptide Sequencing using LC Retention Time Information <i>Yves Frank, Tomas Hruz, Thomas Tschager, and Valentin Venzin</i>	26:1–26:17
Optimal Completion of Incomplete Gene Trees in Polynomial Time Using OCTAL <i>Sarah Christensen, Erin K. Molloy, Pranjal Vachaspati, and Tandy Warnow</i>	27:1–27:14

■ Preface

This proceedings volume contains papers presented at the 17th Workshop on Algorithms in Bioinformatics (WABI 2017), which was held in Boston, MA, USA in conjunction with the 8th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB) from August 21–23, 2017.

The Workshop on Algorithms in Bioinformatics is an annual conference established in 2001 to cover all aspects of algorithmic work in bioinformatics, computational biology, and systems biology. The workshop is intended as a forum for discrete algorithms and machine-learning methods that address important problems in molecular biology, that are founded on sound models, that are computationally efficient, and that have been implemented and tested in simulations and on real datasets. The meeting's focus is on recent research results, including significant work-in-progress, as well as identifying and explore directions of future research.

WABI 2017 is grateful for the support of ACM-BCB in allowing us to cohost the meetings, as well as to ACM-BCB's sponsors: the Association for Computing Machinery (ACM) and ACM's SIGBIO.

In 2017, a total of 55 manuscripts were submitted to WABI from which 27 were selected for presentation at the conference. This year, WABI is adopting a new proceedings form, publishing the conference proceedings through the LIPIcs (Leibniz International Proceedings in Informatics) proceedings series. Extended versions of selected papers will be invited for publication in a thematic series in the journal *Algorithms for Molecular Biology (AMB)*, published by BioMed Central.

The 27 papers were selected based on a thorough peer review, involving at least three independent reviewers per submitted paper, followed by discussions among the WABI Program Committee members. The selected papers cover a wide range of topics, including statistical inference, phylogenetic studies, sequence and genome analysis, comparative genomics, and mass spectrometry data analysis.

We thank all the authors of submitted papers and the members of the WABI Program Committee and their reviewers for their efforts that made this conference possible. We are also grateful to the WABI Steering Committee for their help and advice. We also thank all the conference participants and speakers who contribute to a great scientific program. In particular, we are indebted to the keynote speaker of the conference, Tandy Warnow, for her presentation. We also thank Christopher Pockrandt for setting up the WABI webpage and Umit Acar for his help with coordinating the WABI and ACM-BCB pages. Finally, we thank the ACM-BCB Organizing Committee, especially Nurit Haspel and Lenore Cowen, for their hard work in making all of the local arrangements and working closely with us to ensure a successful and exciting WABI and ACM-BCB.



■ List of Authors

Anastasiia Abramova
Saint Petersburg State University
St. Petersburg, Russia
abramova.asya93@gmail.com

Bahar Alipanahi
University of Florida
Gainesville, USA
bahar.panahie@gmail.com

Yannis Almirantis
National Center for Scientific Research
Demokritos
Athens, Greece
yalmir@bio.demokritos.gr

Fatemeh Almodaresi
Stony Brook University
New York, USA
falmodaresit@cs.stonybrook.edu

Stephen Altschul
National Institutes of Health
Bethesda, USA
altschul@ncbi.nlm.nih.gov

Amir H. Bayegan
Boston College
Boston, USA
bayegan@bc.edu

Shamsuzzoha Bayzid
Bangladesh University of Engineering and
Technology
Dhaka, Bangladesh
shams.bayzid@gmail.com

Christina Boucher
University of Florida
Gainesville, USA
christinaboucher@ufl.edu

Hector Bravo
University of Maryland
College Park, USA
hcorrada@umiacs.umd.edu

Brian Brubach
University of Maryland
College Park, USA
bbrubach@cs.umd.edu

Evelyn Bunnik
University of Texas
Houston, USA
bunnik@uthscsa.edu

Pannagiotos Charalampopoulos
King's College London
London, UK
panagiotis.charalampopoulos@kcl.ac.uk

Sarah Christensen
University of Illinois
Urbana-Champaign, USA
sac2@illinois.edu

Michal Ciach
University of Warsaw
Warsaw, Poland
m_ciach@student.uw.edu.pl

Peter Clote
Boston College
Boston, USA
clote@bc.edu

Matteo Comin
University of Padova
Padova, Italy
comin@dei.unipd.it

Steven Cornwell
University of Pennsylvania
Philadelphia, USA
steffenc@seas.upenn.edu

Maxime Crochemore
King's College London
London, UK
maxime.crochemore@kcl.ac.uk

Peter Ebert
Max Planck Institute for Informatics
Saarbrücken, Germany
pebert@mpi-inf.mpg.de

17th International Workshop on Algorithms in Bioinformatics (WABI 2017).
Editors: Russell Schwartz and Knut Reinert



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Nadia El-Mabrouk
University of Montreal
Montreal, Canada
mabrouk@iro.umontreal.ca

Alexandre Francisco
Universidade de Lisboa
Lisboa, Portugal
apl@ist.utl.pt

Yves Frank
ETH Zurich
Zurich, Switzerland
yfrank@student.ethz.ch

Jia Gao
King's College London
London, UK
jia.gao@kcl.ac.uk

Manuela Geiß
University of Leipzig
Leipzig, Germany
manuela@bierdepot.bioinf.uni-leipzig.de

Samuele Girotto
University of Padova
Padova, Italy
samuele.girotto@gmail.com

Jay Ghurye
University of Maryland
College Park, USA
jayg@cs.umd.edu

Pavel Górecki
University of Warsaw
Warsaw, Poland
gorecki@mimuw.edu.pl

Mohamed Gunady
University of Maryland
College Park, USA
mgunady@cs.umd.edu

Md Abid Hasan
University of California, Riverside
Riverside, USA
mhasa006@ucr.edu

Marc Hellmuth
University of Greifswald
Greifswald, Germany
mhellmuth@mailbox.org

Jan Holub
Czech Technical University in Prague
Prague, Czech Republic
Jan.Holub@fit.cvut.cz

Thomas Hruz
ETH Zurich
Zurich, Switzerland
tomas.hruz@inf.ethz.ch

Chenyue Hu
Rice University
Houston, USA
wendyhu001@gmail.com

Costas Iliopoulos
King's College London
London, UK
costas.iliopoulos@kcl.ac.uk

Hosna Jabbari
University of Alberta
Edmonton, Canada
jabbari@ualberta.ca

Sreeran Kannan
University of Washington
Seattle, USA
ksreeram@uw.edu

Jonkyu Kim
Freie Universität Berlin
Berlin, Germany
j.kim@fu-berlin.de

Anton Korobeynikov
Saint Petersburg State University
St. Petersburg, Russia
anton@korobeynikov.info

Karine Le Roch
University of California, Riverside
Riverside, USA
karine.leroch@ucr.edu

Hanyang Li
Rice University
Houston, USA
hl43@rice.edu

Stefano Lonardi
University of California, Riverside
Riverside, USA
stelo@cs.ucr.edu

Mohamed Manal
King's College London
London, UK
manal.mohamed@kcl.ac.uk

Shunfu Mao
University of Washington
Seattle, USA
shunfu@uw.edu

Daniel Merkle
University of Southern Denmark
Odense, Denmark
daniel@imada.sdu.dk

Sohail Mohajer
University of Minnesota
Rochester, USA
soheil@umn.edu

Erin K. Molloy
University of Illinois
Urbana-Champaign, USA
emolloy2@illinois.edu

Carlo Montemagno
University of Alberta
Edmonton, Canada
montemag@ualberta.ca

Steven Mount
University of Maryland
College Park, USA
smount@umd.edu

Martin Muggli
University of Helsinki
Helsinki, Finland
Martin.Muggli@colostate.edu

Anna Muszewsak
Polish Academy of Sciences
Warsaw, Poland
mmusze@ibb.waw.pl

Sarvesh Nikumbh
Max Planck Institute for Informatics
Saarbrücken, Germany
snikumbh@mpi-inf.mpg.de

Nikolai Nøjgaard
University of Greifswald
Greifswald, Germany
nnoej10@student.sdu.dk

Aida Ouangraoua
Sherbrooke university
Quebec, Canada
aida.ouangraoua@usherbrooke.ca

Weihua Pan
University of California, Riverside
Riverside, USA
wpan005@ucr.edu

Prashand Pandey
Stony Brook University
New York, USA
ppandey@cs.stonybrook.edu

Rob Patro
Stony Brook University
New York, USA
Rob.Patro@cs.stonybrook.edu

Nico Pfeifer
University of Tübingen & Max Planck
Institute for Informatics
Saarbrücken, Germany
pfeifer@informatik.uni-tuebingen.de

Solon Pissis
King's College London
London, UK
solon.pissis@kcl.ac.uk

Cinzia Pizzi
University of Padova
Padova, Italy
cinzia.pizzi@dei.unipd.it

Anton Polishko
University of California, Riverside
Riverside, USA
polishka@cs.ucr.edu

Dimitri Polychronopoulos
Imperial College London
London, UK
dpolychr@imperial.ac.uk

Mihai Pop
University of Maryland
College Park, USA
mpop@umiacs.umd.edu

Petr Procházka
Czech Technical University in Prague
Prague, Czech Republic
Petr.Prochazka@fit.cvut.cz

Simon Puglisi
University of Helsinki
Helsinki, Finland
puglisi@cs.helsinki.fi

Jens Quedenfeld
Technical University of Munich
Munich, Germany
jens.quedenfeld@in.tum.de

Amina Qutub
Rice University
Houston, USA
aminaq@rice.edu

Sven Rahmann
University of Duisburg-Essen
Duisburg, Germany
sven.rahmann@uni-due.de

Kannan Ramachandran
University of California, Berkeley
Berkeley, USA
kannanr@eecs.berkeley.edu

Knut Reinert
Freie Universität Berlin
Berlin, Germany
knut.reinert@fu-berlin.de

Abbas Roayaei Ardakany
University of California, Riverside
Riverside, USA
roayaei@gmail.com

Leena Salmela
University of Helsinki
Helsinki, Finland
leena.salmela@cs.helsinki.fi

Pijus Simonaitis
ENS Lyon
Lyon, France
pijus.simonaitis@ens-lyon.fr

Nidhi Shah
University of Maryland
College Park, USA
nidhi@cs.umd.edu

Aravind Srinivasan
University of Maryland
College Park, USA
srin@cs.umd.edu

Peter F. Stadler
University of Leipzig
Leipzig, Germany
studla@bioinf.uni-leipzig.de

Krister Swenson
Université Montpellier
Montpellier, France
swenson@lirmm.fr

Jussi Taipala
Karolinska Institutet
Stockholm, Sweden
jussi.taipale@ki.se

Jarkko Toivonen
University of Helsinki
Helsinki, Finland
jarkko.toivonen@cs.helsinki.fi

Thomas Tschager
ETH Zurich
Zurich, Switzerland
thomas.tschager@inf.ethz.ch

David Tse
Stanford University
Stanford, USA
dntse@stanford.edu

Esko Ukkonen
University of Helsinki
Helsinki, Finland
esko.ukkonen@cs.helsinki.fi

Pranjal Vachaspati
University of Illinois
Urbana-Champaign, USA
vachasp2@illinois.edu

Cátia Vaz
Universidade de Lisboa
Lisboa, Portugal
cvaz@cc.isel.ipl.pt

Valentin Venzin
ETH Zurich
Zurich, Switzerland
vvenzin@student.ethz.ch

Ian Wark
University of Alberta
Edmonton, Canada
wark@ualberta.ca

Tandy Warnow
University of Illinois
Urbana-Champaign, USA
warnow@illinois.edu

Nicolas Wieseke
University of Leipzig
Leipzig, Germany
wieseke@informatik.uni-leipzig.de

Sebastian Will
University of Vienna
Vienna, Austria
will@tbi.univie.ac.at

