18th International Workshop on Algorithms in Bioinformatics

WABI 2018, August 20–22, 2018, Helsinki, Finland

Edited by
Laxmi Parida
Esko Ukkonen
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 (Chair, Gran Sasso Science Institute and Reykjavik University)
- Susanne Albers (TU München)
- Christel Baier (TU Dresden)
- Javier Esparza (TU München)
- 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

https://www.dagstuhl.de/lipics
## Contents

Preface

Laxmi Parida and Esko Ukkonen ................................................ 0:vii

A Duality-Based Method for Identifying Elemental Balance Violations in Metabolic Network Models

Hooman Zabeti, Tamon Stephen, Bonnie Berger, and Leonid Chindelevitch ....... 1:1–1:13

Prefix-Free Parsing for Building Big BWTs

Christina Boucher, Travis Gagie, Alan Kuhnle, and Giovanni Manzini ............ 2:1–2:16

Detecting Mutations by eBWT

Nicola Prezza, Nadia Pisanti, Marinella Sciortino, and Giovanna Rosone ........ 3:1–3:15

Haplotype-aware graph indexes


Reconciling Multiple Genes Trees via Segmental Duplications and Losses

Riccardo Dondi, Manuel Lafond, and Celine Scornavacca .......................... 5:1–5:16

Protein Classification with Improved Topological Data Analysis

Tamal K. Dey and Sayan Mandal .................................................. 6:1–6:13

A Dynamic Algorithm for Network Propagation

Barak Sternberg and Roded Sharan .............................................. 7:1–7:13

New Absolute Fast Converging Phylogeny Estimation Methods with Improved Scalability and Accuracy

Qiuyi (Richard) Zhang, Satish Rao, and Tandy Warnow .......................... 8:1–8:12

An Average-Case Sublinear Exact Li and Stephens Forward Algorithm

Yohei M. Rosen and Benedict J. Paten .......................................... 9:1–9:13

External memory BWT and LCP computation for sequence collections with applications

Lavinia Egidi, Felipe A. Louza, Giovanni Manzini, and Guilherme P. Telles ..... 10:1–10:14

Kermit: Guided Long Read Assembly using Coloured Overlap Graphs

Riku Walve, Pasi Rastas, and Leena Salmela ..................................... 11:1–11:11

A Succinct Solution to Rmap Alignment

Martin D. Muggli, Simon J. Puglisi, and Christina Boucher ........................ 12:1–12:16

Spalter: A Meta Machine Learning Approach to Distinguish True DNA Variants from Sequencing Artefacts

Till Hartmann and Sven Rahmann .................................................. 13:1–13:8

Essential Simplices in Persistent Homology and Subtle Admixture Detection

Saugata Basu, Filippo Utro, and Laxmi Parida ..................................... 14:1–14:10

Minimum Segmentation for Pan-genomic Founder Reconstruction in Linear Time

Tuukka Norri, Bastien Cazaux, Dmitry Kosolobov, and Veli Mäkinen ............ 15:1–15:15
Multiple-Choice Knapsack for Assigning Partial Atomic Charges in Drug-Like Molecules
  Martin S. Engler, Bertrand Caron, Lourens Veen, Daan P. Geerke, Alan E. Mark, and Gunnar W. Klau .................................................. 16:1–16:13

$\ell_1$-Penalised Ordinal Polytomous Regression Estimators with Application to Gene Expression Studies
  Stéphane Chrétien, Christophe Guyeux, and Serge Moulin .................. 17:1–17:13

Differentially Mutated Subnetworks Discovery
  Morteza Chalabi Hajkarim, Eli Upfal, and Fabio Vandin .................... 18:1–18:14

$DGEN$: A Test Statistic for Detection of General Introgression Scenarios
  Ryan A. Leo Elworth, Chabrielle Allen, Travis Benedict, Peter Dulworth, and Luay Nakhleh .......................................................... 19:1–19:13

PRINCE: Accurate Approximation of the Copy Number of Tandem Repeats
  Mehrdad Mansouri, Julian Booth, Margaryta Vityaz, Cedric Chauve, and Leonid Chindelevitch ......................................................... 20:1–20:13

Degenerate String Comparison and Applications

A Multi-labeled Tree Edit Distance for Comparing “Clonal Trees” of Tumor Progression

Heuristic Algorithms for the Maximum Colorful Subtree Problem
  Kai Dührkop, Marie A. Lataretu, W. Timothy J. White, and Sebastian Böcker ... 23:1–23:14

Parsimonious Migration History Problem: Complexity and Algorithms
  Mohammed El-Kebir .............................................................................. 24:1–24:14

The Wasserstein Distance as a Dissimilarity Measure for Mass Spectra with Application to Spectral Deconvolution
  Szymon Majewski, Michal Aleksander Ciach, Michal Startek, Wanda Niemyska, Błażej Miasojedow, and Anna Gambin .............................. 25:1–25:21
Preface

This proceedings volume contains papers presented at the 18th Workshop on Algorithms in Bioinformatics (WABI 2018), which was held at Aalto University, Helsinki, Finland, from August 20–22, 2018. WABI 2018 was one of seven conferences that were organized as part of ALGO 2018.

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 exploring directions of future research.

WABI 2018 is grateful for the support to the sponsors of ALGO 2018: Federation of Finnish Learned Societies, Helsinki Institute for Information Technology HIIT, City of Helsinki, and European Association for Theoretical Computer Science (EATCS).

In 2018, a total of 57 manuscripts were submitted to WABI from which 25 were selected for presentation at the conference and are included in this proceedings volume as full papers. 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 25 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 phylogenetic studies, network analysis, sequence and genome analysis, topological data analysis and analysis of mass spectrometry data.

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 thank all the conference participants and speakers who contributed to a great scientific program. In particular, we are indebted to the keynote speaker of the conference, Mihai Pop, for his presentation. We also thank Niina Haiminen and Jarkko Toivonen as well as Päivi Lindeman and Pirjo Mulari for their help with editing the proceedings and setting up the poster session. Finally, we thank the ALGO Organizing Committee and Parinya Chalermsook, Petteri Kaski, and Jukka Suomela in particular for their hard work in making all of the local arrangements.

June 2018

Laxmi Parida
Esko Ukkonen
Organization

Program Chairs

Laxmi Parida  
Esko Ukkonen

IBM T. J. Watson Research Center
University of Helsinki

Program Committee

Tatsuya Akutsu  
Lars Arvestad  
Anne Bergeron  
Paola Bonizzoni  
Alessandra Carbone  
Benny Chor  
Daniel Doerr  
Nadia El-Mabrouk  
Liliana Florea  
Anna Gambin  
Niina Haiinen  
Iman Hajirasouliha  
Bjarni Halldorsson  
Katharina Huber  
Carl Kingsford  
Gunnar W. Klau  
Harri Lähdesmäki  
Timo Lassmann  
Yu Lin  
Zsuzsanna Liptak  
Veli Mäkinen  
Paul Medvedev  
Bernard Moret  
Burkhard Morgenstern  
Vincent Moulton  
Robert Patro  
Christian N. Storm Pedersen  
Solon Pissis  
Cinzia Pizzi  
Alex Pothen  
Teresa Przytycka  
Sven Rahmann  
Knut Reinert  
Eric Rivals  
Simona Rombo  
Marie-France Sagot  
Yasubumi Sakakibara  
Mikaël Salson

Kyoto University
Stockholm University
Université du Québec a Montréal
Università di Milano-Bicocca
Université Pierre et Marie Curie
Tel-Aviv Univ
Bielefeld University
University of Montreal
Johns Hopkins University
Institute of Informatics, Warsaw University
IBM T. J. Watson Research Center
Cornell University
deCODE genetics and Reykjavik University
University of East Anglia
Carnegie Mellon University
Heinrich Heine University Düsseldorf
Aalto University
Teleton Kids
Australian National University
University of Verona
University of Helsinki
Pennsylvania State University
EPFL
University of Goettingen
University of East Anglia
Stony Brook
Aarhus University
King’s College London
Univ Padova
Purdue Univ
NIH
University of Duisburg-Essen
FU Berlin
CNRS & Univ Montpellier
Univ Palermo
INRIA Grenoble Rhône-Alpes and Univ de Lyon 1
Keio University
Université Lille 1

18th International Workshop on Algorithms in Bioinformatics (WABI 2018).
Editors: Laxmi Parida and Esko Ukkonen
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
Organization

Michael Schatz  Cold Spring Harbor Laboratory
Marcel Schulz  Saarland University
Russell Schwartz  Carnegie Mellon University
Kana Shimizu  Waseda University
Rahul Siddharthan  Institute of Mathematical Sciences, Chennai
Alexandros Stamatakis  Heidelberg Inst., for Theoretical Studies
Jens Stoye  Bielefeld University
Hélène Touzet  University of Lille and INRIA
Lusheng Wang  City University of Hong Kong
Tandy Warnow  University of Illinois at Urbana-Champaign
Yufeng Wu  Univ Connecticut
Louxin Zhang  National University of Singapore
Shaojie Zhang  University of Central Florida
Michal Ziv-Ukelson  Ben Gurion University of the Negev

WABI Steering Committee

Bernard Moret  EPFL
Vincent Moulton  University of East Anglia
Jens Stoye  Bielefeld University
Tandy Warnow  University of Illinois at Urbana-Champaign

ALGO Organizing Committee Chairs

Parinya Chalermsook  Aalto University
Petteri Kaski  Aalto University
Jukka Suomela  Aalto University

Additional Reviewers

Said Sadique Adi  Krzysztof Gogolewski  Murray Patterson
Ilan Ben-Bassat  Mauricio Soto Gomez  Leonardo Pellegrina
Sebastian Böcker  Aldo Guzmán-Sáenz  Pierre Peterlongo
Aritra Bose  Arnaldur Gylfason  Christopher Pockrandt
Christina Boucher  Martheinn Hardarson  Raffaella Rizzi
Bastien Cazaux  Lin He  Giovanna Rosone
Annie Chateau  Markus Heinonen  Yutaka Saito
Rayan Chikhi  Shahidul Islam  Mingfu Shao
Neo Christopher Chung  Vasanthan Jayakumar  Blerina Sinaimeri
Simone Ciccolella  Dominik Kempa  Jean-Stéphane Varré
Leandro I. S. De Lima  Kiavash Kianfar  Roland Wittler
Meleskho Dmitrii  Kuo-Ching Liang  Weronika Wronowska
Yoann Dufresne  Simone Linz  Zohar Yakhini
Hannes P. Eggertsson  Guillaume Marçais  Guangyu Yang
Jamal Elkhader  Ardalan Naseri  Hongyu Zheng
Giuditta Franco  Laurent Noé
List of Authors

Chabrielle Allen
caa6@rice.edu
Rice University
United States

Mai Alzamel
mai.alzamel@kcl.ac.uk
King’s College London
United Kingdom

Lorraine A. K. Ayad
lorraine.ayad@kcl.ac.uk
King’s College London
United Kingdom

Saugata Basu
sbasu@purdue.edu
Purdue University
United States

Travis Benedict
trb6@rice.edu
Rice University
United States

Bonnie Berger
bab@csail.mit.edu
MIT
United States

Giulia Bernardini
giuliabernardini.morg@gmail.com
University of Milan-Bicocca
Italy

Sebastian Böcker
sebastian.boecker@uni-jena.de
Friedrich Schiller University Jena
Germany

Julian Booth
julius_booth@sfu.ca
Simon Fraser University
Canada

Christina Boucher
cboucher@cise.ufl.edu
University of Florida
United States

Bertrand Caron
b.caron@uq.edu.au
School of Chemistry & Molecular Biosciences,
The University of Queensland, St Lucia
Australia

Bastien Cazaux
bastien.cazaux@ilrm.fr
University of Helsinki
Finland

Cedric Chauve
cedric_chauve@sfu.ca
Simon Fraser University
Canada

Leonid Chindelevitch
leonid@sfu.ca
Simon Fraser University
Canada

Stéphane Chrétien
stephane.chretien@npl.co.uk
National Physical Laboratory
United Kingdom

Michał Aleksander Ciach
m_ciach@student.uw.edu.pl
Institute of Mathematics, Informatics and Mechanics, University of Warsaw
Poland

Tamal K. Dey
dey.8@osu.edu
The Ohio State University
United States

Riccardo Dondi
riccardo.dondi@unibg.it
Università degli Studi di Bergamo
Italy

Kai Dührkop
Kai.Duehrkop@uni-jena.de
Friedrich-Schiller-University Jena
Germany
Peter Dulworth  
psd2@rice.edu  
Rice University  
United States

Richard Durbin  
rd@sanger.ac.uk  
Department of Genetics, University of Cambridge  
United Kingdom

Lavinia Egidi  
lavinia.egidi@mfn.unipmn.it  
Computer Science Institute, University of Eastern Piedmont  
Italy

Mohammed El-Kebir  
melkebir@illinois.edu  
University of Illinois at Urbana-Champaign  
United States

Ryan A. Leo Elworth  
r.a.leo.elworth@rice.edu  
Rice University  
United States

Martin S. Engler  
martin.engler@cwi.nl  
Life Sciences and Health Group, Centrum Wiskunde & Informatica, Amsterdam  
Netherlands

Travis Gagie  
travis.gagie@gmail.com  
Diego Portales University  
Chile

Anna Gambin  
aniaag@mimuw.edu.pl  
Institute of Mathematics, Informatics and Mechanics, University of Warsaw  
Poland

Erik Garrison  
eg10@sanger.ac.uk  
Wellcome Sanger Institute  
United Kingdom

Daan P. Geerke  
d.p.geerke@vu.nl  
AIMMS Division of Molecular and Computational Toxicology, Vrije Universiteit Amsterdam  
Netherlands

Roberto Grossi  
grossi@di.unipi.it  
University of Pisa  
Italy

Christophe Guyeux  
christophe.guyeux@univ-fcomte.fr  
University of Bourgogne Franche-Comté  
France

Morteza Chalabi Hajkarim  
morteza.hajkarim@bric.ku.dk  
University of Copenhagen  
Denmark

Till Hartmann  
P.Hartmann@uni-due.de  
University of Duisburg-Essen  
Germany

Costas S. Iliopoulos  
c.iliopoulos@kcl.ac.uk  
King’s College London  
United Kingdom

Nikolai Karpov  
nkarpov@iu.edu  
Indiana University Bloomington  
United States

Gunnar W. Klau  
gunnar.klau@hhu.de  
Algorithmic Bioinformatics, Heinrich Heine University Düsseldorf  
Germany

Dmitry Kosolobov  
dkosolobov@mail.ru  
Ural Federal University  
Russia

Alan Kuhnle  
akuhnle418@gmail.com  
University of Florida  
United States
Authors

Manuel Lafond
mlafond2@uOttawa.ca
Université de Sherbrooke
Canada

Marie A. Lataretu
marie.lataretu@uni-jena.de
Friedrich-Schiller-University Jena
Germany

Felipe A. Louza
louza@usp.br
Department of Computing and Mathematics, University of São Paulo
Brazil

Szymon Majewski
smajewski@impan.pl
Institute of Mathematics, Polish Academy of Sciences
Poland

Veli Mäkinen
vmakinen@cs.helsinki.fi
University of Helsinki
Finland

Salem Malikic
smalikic@sfu.ca
Simon Fraser University
Canada

Sayan Mandal
mandal.25@osu.edu
The Ohio State University
United States

Mehrdad Mansouri
mehrdad_mansouri@sfu.ca
Simon Fraser University
Canada

Giovanni Manzini
giovanni.manzini@unipmn.it
Computer Science Institute, University of Eastern Piedmont
Italy

Alan E. Mark
a.e.mark@uq.edu.au
School of Chemistry & Molecular Biosciences, The University of Queensland, St Lucia
Australia

Błażej Miasojedow
bmia@mimuw.edu.pl
Institute of Mathematics, Informatics and Mechanics, University of Warsaw
Poland

Serge Moulin
serge.moulin@hotmail.fr
University of Bourgogne Franche-Comté
France

Martin D. Muggli
martin.muggli@colorstate.edu
Colorado State University
United States

Luay Nakhleh
nakhleh@cs.rice.edu
Rice University
United States

Wanda Niemyska
w.niemyska@mimuw.edu.pl
Institute of Mathematics, Informatics and Mechanics, University of Warsaw
Poland

Tuukka Norri
tsnorri@iki.fi
University of Helsinki
Finland

Adam M. Novak
anovak@soe.ucsc.edu
University of California, Santa Cruz
United States

Laxmi Parida
parida@us.ibm.com
IBM T. J. Watson Research Center
United States

Benedict J. Paten
bpaten@ucsc.edu
University of California, Santa Cruz
United States

Nadia Pisanti
pisanti@di.unipi.it
Dipartimento di Informatica, Universita di Pisa, Italy & Erable Team, INRIA
Italy
Filippo Utro
futro@us.ibm.com
IBM T. J Watson Research Center
United States

Fabio Vandin
fabio.vandin@unipd.it
University of Padova
Italy

Lourens Veen
l.veen@esciencecenter.nl
Netherlands eScience Center, Amsterdam
Netherlands

Margaryta Vityaz
margaryta_vityaz@sfu.ca
Simon Fraser University
Canada

Riku Walve
riku.walve@helsinki.fi
University of Helsinki
Finland

Tandy Warnow
warnow@illinois.edu
University of Illinois at Urbana-Champaign
United States

W. Timothy J. White
tim.white@bihealth.de
Berlin Institute of Health
Germany

Hooman Zabeti
hzabeti@sfu.ca
Simon Fraser University
Canada

Qiuyi (Richard) Zhang
qiuyi@math.berkeley.edu
University of California, Berkeley
United States