

20th International Workshop on Algorithms in Bioinformatics

WABI 2020, September 7–9, 2020, Pisa, Italy (Virtual
Conference)

Edited by

Carl Kingsford

Nadia Pisanti



Editors

Carl Kingsford 

Carnegie Mellon University, USA
carlk@cs.cmu.edu

Nadia Pisanti 

University of Pisa, Italy
pisanti@di.unipi.it

ACM Classification 2012

Mathematics of computing → Discrete mathematics; Applied computing → Computational biology;
Mathematics of computing → Information theory; Theory of computation → Design and analysis of algorithms

ISBN 978-3-95977-161-0

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at <https://www.dagstuhl.de/dagpub/978-3-95977-161-0>.

Publication date

August, 2020

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 <https://portal.dnb.de>.

License

This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0):
<https://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.2020.0

ISBN 978-3-95977-161-0

ISSN 1868-8969

<https://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 (*Chair*, Gran Sasso Science Institute and Reykjavik University)
- Christel Baier (TU Dresden)
- Mikolaj Bojanczyk (University of Warsaw)
- Roberto Di Cosmo (INRIA and University Paris Diderot)
- Javier Esparza (TU München)
- Meena Mahajan (Institute of Mathematical Sciences)
- Dieter van Melkebeek (University of Wisconsin-Madison)
- Anca Muscholl (University Bordeaux)
- Luke Ong (University of Oxford)
- Catuscia Palamidessi (INRIA)
- Thomas Schwentick (TU Dortmund)
- Raimund Seidel (Saarland University and Schloss Dagstuhl – Leibniz-Zentrum für Informatik)

ISSN 1868-8969

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

■ Contents

Preface	
<i>Carl Kingsford and Nadia Pisanti</i>	0:vii–0:viii
Program Committee	
.....	0:ix–0:x
Regular Paper	
Approximate Search for Known Gene Clusters in New Genomes Using PQ-Trees	
<i>Galia R. Zimmerman, Dina Svetlitsky, Meirav Zehavi, and Michal Ziv-Ukelson</i>	1:1–1:24
An Interpretable Classification Method for Predicting Drug Resistance in M. Tuberculosis	
<i>Hooman Zabeti, Nick Dexter, Amir Hosein Safari, Nafiseh Sedaghat, Maxwell Libbrecht, and Leonid Chindelevitch</i>	2:1–2:18
Natural Family-Free Genomic Distance	
<i>Diego P. Rubert, Fábio V. Martinez, and Marília D. V. Braga</i>	3:1–3:23
Fast Lightweight Accurate Xenograft Sorting	
<i>Jens Zentgraf and Sven Rahmann</i>	4:1–4:16
Phyolin: Identifying a Linear Perfect Phylogeny in Single-Cell DNA Sequencing Data of Tumors	
<i>Leah L. Weber and Mohammed El-Kebir</i>	5:1–5:14
The Longest Run Subsequence Problem	
<i>Sven Schrinner, Manish Goel, Michael Wulfert, Philipp Spohr, Korbinian Schneeberger, and Gunnar W. Klau</i>	6:1–6:13
Linear Time Construction of Indexable Founder Block Graphs	
<i>Veli Mäkinen, Bastien Cazaux, Massimo Equi, Tuukka Norri, and Alexandru I. Tomescu</i>	7:1–7:18
GraphBin2: Refined and Overlapped Binning of Metagenomic Contigs Using Assembly Graphs	
<i>Vijini G. Mallawaarachchi, Anuradha S. Wickramarachchi, and Yu Lin</i>	8:1–8:21
Fast and Efficient Rmap Assembly Using the Bi-Labelled de Bruijn Graph	
<i>Kingshuk Mukherjee, Massimiliano Rossi, Leena Salmela, and Christina Boucher</i>	9:1–9:16
Economic Genome Assembly from Low Coverage Illumina and Nanopore Data	
<i>Thomas Gatter, Sarah von Löhneysen, Polina Drozdova, Tom Hartmann, and Peter F. Stadler</i>	10:1–10:22
A Graph-Theoretic Barcode Ordering Model for Linked-Reads	
<i>Yoann Dufresne, Chen Sun, Pierre Marijon, Dominique Lavenier, Cedric Chauve, and Rayan Chikhi</i>	11:1–11:17
Exact Transcript Quantification Over Splice Graphs	
<i>Cong Ma, Hongyu Zheng, and Carl Kingsford</i>	12:1–12:19

20th International Workshop on Algorithms in Bioinformatics (WABI 2020).

Editors: Carl Kingsford and Nadia Pisanti



Leibniz International Proceedings in Informatics

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

Shape Decomposition Algorithms for Laser Capture Microdissection <i>Leonie Selbach, Tobias Kowalski, Klaus Gerwert, Maike Buchin, and Axel Mosig</i> .	13:1–13:17
The Bourque Distances for Mutation Trees of Cancers <i>Katharina Jahn, Niko Beerenwinkel, and Louxin Zhang</i>	14:1–14:22
Advancing Divide-And-Conquer Phylogeny Estimation Using Robinson-Foulds Supertrees <i>Xilin Yu, Thien Le, Sarah Christensen, Erin K. Molloy, and Tandy Warnow</i>	15:1–15:17
Disk Compression of k -mer Sets <i>Amatur Rahman, Rayan Chikhi, and Paul Medvedev</i>	16:1–16:18
Near-Linear Time Edit Distance for Indel Channels <i>Arun Ganesh and Aaron Sy</i>	17:1–17:18
Fold Family-Regularized Bayesian Optimization for Directed Protein Evolution <i>Trevor S. Frisby and Christopher J. Langmead</i>	18:1–18:17
Sequence Searching Allowing for Non-Overlapping Adjacent Unbalanced Translocations <i>Domenico Cantone, Simone Faro, and Arianna Pavone</i>	19:1–19:14

■ Preface

This proceedings volume contains papers presented at the 20th Workshop on Algorithms in Bioinformatics (WABI 2020), which was virtually held in Pisa, Italy, September 7–9, 2020.

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 conference is intended as a forum for presentation of new insights about discrete algorithms and machine-learning methods that address important problems in biology (particularly problems based on molecular data and phenomena); 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.

This 20th edition of WABI took place in the year of the COVID-19 pandemic emergency. This presented some logistic challenges, but also served to highlight the importance of the development of advanced computational methodologies for understanding biological problems. Over the 20 instances of WABI, computational biology has grown significantly in importance, and now computational analysis methods — some furthered significantly over the years at WABI — are crucial to quickly understanding the evolution and function of the agent of a global health crisis.

WABI 2020 was organized within the ALGO federation of conferences that in 2020 included WABI, ESA (European Symposium on Algorithms), ALGO CLOUD (International Symposium on Algorithmic Aspects of Cloud Computing), ALGO SENSORS (International Symposium on Algorithms and Experiments for Wireless Sensor Networks), ATMOS (International Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems), and WAOA (Workshop on Approximation and Online Algorithms).

Because of the COVID-19 pandemic, the ALGO Organizing Committee decided to run the affiliated conferences online exclusively. On the other hand, the scientific aspects were not affected, except for a somewhat lower number of submissions that fortunately maintained the high quality of WABI standards. The activity of the Programme Committee of WABI took place as usual: peer review, discussion, selection of accepted papers, and publishing of the proceedings have been accomplished in the same way as in all WABI editions.

In 2020, a total of 38 manuscripts were submitted to WABI from which 19 were selected for presentation at the conference and are included in this proceedings volume as full papers. Extended versions of selected papers have been invited for publication in a thematic series in the journal *Algorithms for Molecular Biology (AMB)*, published by BioMed Central. The 19 papers selected for the conference underwent a thorough peer review, involving at least three (and often four or five) 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 trees and networks, biological network analysis, sequence alignment and assembly, genomic-level evolution, sequence and genome analysis, RNA and protein structure, topological data analysis, and more. They are ordered randomly within this volume.

We thank all the authors of submitted papers and the members of the WABI 2020 Program Committee and their subreviewers 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, session chairs, and speakers who contributed to



a great scientific program. In particular, we are indebted to the keynote speaker of the conference, Dan Gusfield (University of California Davis), for his presentation, and to the two WABI invited speakers Valentina Boeva (ETH Zurich and Institute Cochin Paris) and Erik Garrison (University of California Santa Cruz). WABI 2020 is grateful for the support of the University of Pisa. We thank ALGO 2020 Organizing Committee for setting up the event in these complicated times due to the pandemic emergency.

Previous proceedings of WABI appeared in LNCS/LNBI volumes 2149 (WABI 2001, Aarhus), 2452 (WABI 2002, Rome), 2812 (WABI 2003, Budapest), 3240 (WABI 2004, Bergen), 3692 (WABI 2005, Mallorca), 4175 (WABI 2006, Zurich), 4645 (WABI 2007, Philadelphia), 5251 (WABI 2008, Karlsruhe), 5724 (WABI 2009, Philadelphia), 6293 (WABI 2010, Liverpool), 6833 (WABI 2011, Saarbrücken), 7534 (WABI 2012, Ljubljana), 8126 (WABI 2013, Sophia Antipolis), 8701 (WABI 2014, Wroclaw), 9289 (WABI 2015, Atlanta), and 9838 (WABI 2016, Aarhus). As of 2016, they appeared in LIPICS volumes 88 (WABI 2017, Boston), 113 (WABI 2018, Helsinki), and 143 (WABI 2019, Boston).

Carl Kingsford & Nadia Pisanti

■ Program Committee

Carl Kingsford (Co-Chair)
Carnegie Mellon University

Nadia Pisanti (Co-Chair)
University of Pisa, Italy

Tatsuya Akutsu
Kyoto University, Japan

Carl Barton
EBI, UK

Anne Bergeron
Universite du Quebec a Montreal, Canada

Paola Bonizzoni
Università di Milano-Bicocca, Italy

Christina Boucher
University of Florida, USA

Alessandra Carbone
Université Pierre et Marie Curie, France

Rayan Chikhi
Pasteur Institute, France

Lenore Cowen
Tufts University, USA

Gianluca Della Vedova
University of Milano-Bicocca, Italy

Nadia El-Mabrouk
University of Montreal, Canada

Anna Gambin
Warsaw University, Poland

Raffaele Giancarlo
University of Palermo, Italy

Dan Gusfield
UC Davis, USA

Bjarni Halldorsson
deCODE genetics and Reykjavik University,
Iceland

Katharina Huber
University of East Anglia, UK

Gunnar Klau
Heinrich Heine University Düsseldorf,
Germany

Gregory Kucherov
University of Paris Est, France

Ritu Kundu
National University of Singapore, Singapore

Manuel Lafond
Université de Sherbrooke, Canada

Stefano Lonardi
University of California Riverside, USA

Veli Makinen
University of Helsinki, Finland

Guillaume Marçais
Carnegie Mellon University, USA

Tobias Marschall
Heinrich Heine University, Düsseldorf,
Germany

Bernard Moret
Ecole Polytechnique Fédérale de Lausanne,
Switzerland

Vincent Moulton
University of East Anglia, UK

Luay Nakhleh
Rice University, USA

William Stafford Noble
University of Washington, USA

Solon Pissis
CWI Amsterdam, the Netherlands

Alberto Policriti
University of Udine, Italy

Mihai Pop
University of Maryland, USA

Teresa Przytycka NCBI
NIH, USA

Sven Rahmann
University of Duisburg-Essen, Germany

20th International Workshop on Algorithms in Bioinformatics (WABI 2020).

Editors: Carl Kingsford and Nadia Pisanti



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

Knut Reinert
FU Berlin, Germany

Eric Rivals
LIRMM – Université de Montpellier, France

Giovanna Rosone
University of Pisa, Italy

Marie-France Sagot
INRIA Rhône-Alpes, France

Alexander Schoenhuth
Bielefeld University, Germany

Jens Stoye
Bielefeld University, Germany

Krister Swenson
CNRS & Université de Montpellier, France

Sharma Thankachan
University of Central Florida, USA

Alexandru Tomescu
University of Helsinki, Finland

Hélène Touzet
CNRS, CRIStAL Lille, France

Esko Ukkonen
University of Helsinki, Finland

Tomas Vinar
Comenius University, Slovakia

Tandy Warnow
University of Illinois at Urbana-Champaign,
USA

Prudence W.H. Wong
University of Liverpool, UK

Louxin Zhang
National University of Singapore, Singapore

Michal Ziv-Ukelson
Ben Gurion University of the Negev, Israel