

# 21st International Workshop on Algorithms in Bioinformatics

WABI 2021, August 2–4, 2021, Virtual Conference

Edited by

Alessandra Carbone  
Mohammed El-Kebir



*Editors*

**Alessandra Carbone** 

UMR 7238 CNRS, Sorbonne Université, Paris, France  
alessandra.carbone@sorbonne-universite.fr

**Mohammed El-Kebir** 

University of Illinois at Urbana-Champaign, Champaign, IL, USA  
melkebir@illinois.edu

*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 978-3-95977-200-6**

*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/978-3-95977-200-6>.

*Publication date*

July, 2021

*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 4.0 International license (CC-BY 4.0):  
<https://creativecommons.org/licenses/by/4.0/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.2021.0

ISBN 978-3-95977-200-6

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*, Reykjavik University, IS and Gran Sasso Science Institute, IT)
- Christel Baier (TU Dresden, DE)
- Mikolaj Bojanczyk (University of Warsaw, PL)
- Roberto Di Cosmo (Inria and Université de Paris, FR)
- Faith Ellen (University of Toronto, CA)
- Javier Esparza (TU München, DE)
- Daniel Král' (Masaryk University - Brno, CZ)
- Meena Mahajan (Institute of Mathematical Sciences, Chennai, IN)
- Anca Muscholl (University of Bordeaux, FR)
- Chih-Hao Luke Ong (University of Oxford, GB)
- Phillip Rogaway (University of California, Davis, US)
- Eva Rotenberg (Technical University of Denmark, Lyngby, DK)
- Raimund Seidel (Universität des Saarlandes, Saarbrücken, DE and Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Wadern, DE)

**ISSN 1868-8969**

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



## ■ Contents

Preface	
<i>Alessandra Carbone and Mohammed El-Kebir</i> .....	0:vii–0:viii
Program Committee	
.....	0:ix–0:x

### Regular Papers

The Most Parsimonious Reconciliation Problem in the Presence of Incomplete Lineage Sorting and Hybridization Is NP-Hard	
<i>Matthew LeMay, Yi-Chieh Wu, and Ran Libeskind-Hadas</i> .....	1:1–1:10
Efficient Privacy-Preserving Variable-Length Substring Match for Genome Sequence	
<i>Yoshiki Nakagawa, Satsuya Ohata, and Kana Shimizu</i> .....	2:1–2:23
Making Sense of a Cophylogeny Output: Efficient Listing of Representative Reconciliations	
<i>Yishu Wang, Arnaud Mary, Marie-France Sagot, and Blerina Sinaimeri</i> .....	3:1–3:18
Perplexity: Evaluating Transcript Abundance Estimation in the Absence of Ground Truth	
<i>Jason Fan, Skylar Chan, and Rob Patro</i> .....	4:1–4:22
The Maximum Duo-Preservation String Mapping Problem with Bounded Alphabet	
<i>Nicolas Boria, Laurent Gourvès, Vangelis Th. Paschos, and Jérôme Monnot</i> .....	5:1–5:12
Treewidth-Based Algorithms for the Small Parsimony Problem on Networks	
<i>Celine Scornavacca and Mathias Weller</i> .....	6:1–6:21
Tree Diet: Reducing the Treewidth to Unlock FPT Algorithms in RNA Bioinformatics	
<i>Bertrand Marchand, Yann Ponty, and Laurent Bulteau</i> .....	7:1–7:23
Space-Efficient Representation of Genomic k-Mer Count Tables	
<i>Yoshihiro Shibuya, Djamel Belazzougui, and Gregory Kucherov</i> .....	8:1–8:19
Parsimonious Clone Tree Reconciliation in Cancer	
<i>Palash Sashittal, Simone Zaccaria, and Mohammed El-Kebir</i> .....	9:1–9:21
An Efficient Linear Mixed Model Framework for Meta-Analytic Association Studies Across Multiple Contexts	
<i>Brandon Jew, Jiajin Li, Sriram Sankararaman, and Jae Hoon Sul</i> .....	10:1–10:17
LRBinner: Binning Long Reads in Metagenomics Datasets	
<i>Anuradha Wickramarachchi and Yu Lin</i> .....	11:1–11:18
Compression of Multiple $k$ -Mer Sets by Iterative SPSS Decomposition	
<i>Kazushi Kitaya and Tetsuo Shibuya</i> .....	12:1–12:17

21st International Workshop on Algorithms in Bioinformatics (WABI 2021).

Editors: Alessandra Carbone and Mohammed El-Kebir



Leibniz International Proceedings in Informatics

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

Compressing and Indexing Aligned Readsets <i>Travis Gagie, Garance Gourdel, and Giovanni Manzini</i> .....	13:1–13:21
BPPart: RNA-RNA Interaction Partition Function in the Absence of Entropy <i>Ali Ebrahimpour-Borojeny, Sanjay Rajopadhye, and Hamidreza Chitsaz</i> .....	14:1–14:24
BISER: Fast Characterization of Segmental Duplication Structure in Multiple Genome Assemblies <i>Hamza Išerić, Can Alkan, Faraz Hach, and Ibrahim Numanagić</i> .....	15:1–15:18
Flow Decomposition with Subpath Constraints <i>Lucia Williams, Alexandru I. Tomescu, and Brendan Mumey</i> .....	16:1–16:15
Conflict Resolution Algorithms for Deep Coalescence Phylogenetic Networks <i>Marcin Wawerka, Dawid Dąbkowski, Natalia Rutecka, Agnieszka Mykowiecka, and Paweł Górecki</i> .....	17:1–17:21
Genome Halving and Aliquoting Under the Copy Number Distance <i>Ron Zeira, Geoffrey Mon, and Benjamin J. Raphael</i> .....	18:1–18:25
Efficient Haplotype Block Matching in Bi-Directional PBWT <i>Ardalan Naseri, William Yue, Shaojie Zhang, and Degui Zh</i> .....	19:1–19:13
Fast Approximate Shortest Hyperpaths for Inferring Pathways in Cell Signaling Hypergraphs <i>Spencer Krieger and John Kececioglu</i> .....	20:1–20:20

## ■ Preface

This proceedings volume contains papers presented at the 21st Workshop on Algorithms in Bioinformatics (WABI 2021), which was virtually held in Chicago, Illinois, USA, August 2–4, 2021.

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.

WABI 2021 was co-located with ACM-BCB 2021. Because of the COVID-19 pandemic, the ACM-BCB Organizing Committee decided to run the affiliated conferences online exclusively. This did not affect the scientific aspects of WABI, with the exception of somewhat lower number of submissions that fortunately maintained the high quality of WABI standards. The activities of the Program 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. This year, for the first time, WABI had a poster session, allowing students to interact with and present their ongoing research to the WABI community.

In 2021, a total of 38 manuscripts were submitted to WABI from which 20 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 20 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 for making this conference possible. A special thanks goes to all the members of the WABI 2021 Program Committee and their subreviewers for their participation in a very active review process with numerous exchanges that culminated in constructive reviews reports for the authors. We are also grateful to the WABI Steering Committee (Bernard Moret, Vincent Moulton, Jens Stoye and Tandy Warnow) for their availability, 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, Mona Singh (Princeton University), for her presentation. We thank the ACM-BCB 2021 Organizing Committee for setting up the event in these complicated times due to the pandemic emergency. Finally, we thank Lechuan Li for maintaining the WABI 2021 website.



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), 143 (WABI 2019, Boston) and 172 (WABI 2020, Pisa).

Alessandra Carbone & Mohammed El-Kebir



## ■ Program Committee

Alessandra Carbone (Co-Chair)  
Sorbonne Universite, France

Mohammed El-Kebir (Co-Chair)  
University of Illinois at Urbana-Champaign,  
USA

Tatsuya Akutsu  
Kyoto University, Japan

Marco Antoniotti  
Universita di Milano-Bicocca, Italy

Anne Bergeron  
Universite du Quebec a Montreal, Canada

Paola Bonizzoni  
Universita di Milano-Bicocca, Italy

Christina Boucher  
University of Florida, USA

Rayan Chikhi  
Institut Pasteur, France

Leonid Chindelevitch  
Imperial College London, UK

Lenore Cowen  
Tufts University, USA

Daniel Doerr  
University of Duesseldorf, Germany

Nadia El-Mabrouk  
University of Montreal, Canada

Anna Gambin  
Warsaw University, Poland

Bjarni Halldorsson  
deCODE genetics and Reykjavik University,  
Iceland

Katharina Huber  
University of East Anglia, UK

Carl Kingsford  
Carnegie Mellon University, USA

Gunnar Klau  
Heinrich Heine University Dusseldorf,  
Germany

Gregory Kucherov  
University of Paris Est, France

Manuel Lafond  
Universite de Sherbrooke, Canada

Elodie Laine  
Sorbonne Universite, France

Yu Lin  
Australia National University, Australia

Stefano Lonardi  
University of California Riverside, USA

Veli Makinen  
University of Helsinki, Finland

Guillaume Marcais  
Carnegie Mellon University, USA

Tobias Marschall  
Heinrich Heine University, Germany

Erin Molloy  
University of Maryland College Park /  
UCLA, US

Bernard Moret  
Ecole Polytechnique Federale de Lausanne,  
Switzerland

Vincent Moulton  
University of East Anglia, UK

Francesca Nadalin  
EMBL-EBI / ITT, Italy, England

Mihai Pop  
University of Maryland, USA

Teresa Przytycka  
NCBI / NIH, USA

William Stafford Noble  
University of Washington, USA

Aida Ouangraoua  
Universite de Sherbrooke, Canada

Ion Petre  
Turku University, Finland

21st International Workshop on Algorithms in Bioinformatics (WABI 2021).

Editors: Alessandra Carbone and Mohammed El-Kebir



Leibniz International Proceedings in Informatics

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

Nadia Pisanti  
University of Pisa, Italy

Solon Pissis  
CWI Amsterdam, the Netherlands

Alberto Policriti  
University of Udine, Italy

Sven Rahmann  
Saarland University, Germany

Knut Reinert  
FU Berlin, Germany

Eric Rivals  
LIRMM - Universite de Montpellier, France

Sebastien Roch  
University of Wisconsin Madison, US

Giovanna Rosone  
University of Pisa, Italy

Marie-France Sagot  
INRIA Rhone-Alpes, France

Alexander Schoenhuth  
Bielefeld University, Germany

Mingfu Shao  
Penn State, US

Jens Stoye  
Bielefeld University, Germany

Krister Swenson  
CNRS and Universite de Montpellier, France

Ewa Szczurek  
Warsaw University, Poland

Sharma Thankachan  
University of Central Florida, USA

Alexandru Tomescu  
University of Helsinki, Finland

Helene Touzet  
CNRS, France

Esko Ukkonen  
University of Helsinki, Finland

Gianluca Della Vedova  
University of Milano-Bicocca, Italy

Tomas Vinar  
Comenius University, Slovakia

Prudence W.H. Wong  
University of Liverpool, UK

Tandy Warnow  
University of Illinois at Urbana-Champaign,  
USA

Simone Zaccaria  
UCL Cancer Institute, UK

Louxin Zhang  
National University of Singapore, Singapore

Xiuwei Zhang  
Georgia Tech, US

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