

22nd International Workshop on Algorithms in Bioinformatics

WABI 2022, September 5–7, 2022, Potsdam, Germany

Edited by

Christina Boucher

Sven Rahmann



Editors

Christina Boucher

Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA

cboucher@cise.ufl.edu

Sven Rahmann

Department of Computer Science and Center for Bioinformatics, Saarland University, Saarbrücken, Germany

<https://www.rahmannlab.de/people/rahmann>

sven.rahmann@uni-saarland.de

ACM Classification 2012

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

ISBN 978-3-95977-243-3

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/978-3-95977-243-3>.

Publication date

September, 2022

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.2022.0

ISBN 978-3-95977-243-3

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>Christina Boucher and Sven Rahmann</i>	0:vii–0:viii
Program Committee	
.....	0:ix–0:xi

Invited Talk

Efficient Solutions to Biological Problems Using de Bruijn Graphs	
<i>Leena Salmela</i>	1:1–1:2


Regular Papers

Eulertigs: Minimum Plain Text Representation of k -mer Sets Without Repetitions in Linear Time	
<i>Sebastian Schmidt and Jarno N. Alanko</i>	2:1–2:21
Predicting Horizontal Gene Transfers with Perfect Transfer Networks	
<i>Alitzel López Sánchez and Manuel Lafond</i>	3:1–3:22
Haplotype Threading Using the Positional Burrows-Wheeler Transform	
<i>Ahsan Sanaullah, Degui Zhi, and Shaoije Zhang</i>	4:1–4:14
Non-Binary Tree Reconciliation with Endosymbiotic Gene Transfer	
<i>Mathieu Gascon and Nadia El-Mabrouk</i>	5:1–5:20
Constructing Founder Sets Under Allelic and Non-Allelic Homologous Recombination	
<i>Konstantinn Bonnet, Tobias Marschall, and Daniel Doerr</i>	6:1–6:23
Automated Design of Dynamic Programming Schemes for RNA Folding with Pseudoknots	
<i>Bertrand Marchand, Sebastian Will, Sarah J. Berkemer, Laurent Bulteau, and Yann Ponty</i>	7:1–7:24
Fast and Accurate Species Trees from Weighted Internode Distances	
<i>Baqiao Liu and Tandy Warnow</i>	8:1–8:24
On Weighted k -mer Dictionaries	
<i>Giulio Ermanno Pibiri</i>	9:1–9:20
Accurate k -mer Classification Using Read Profiles	
<i>Yoshihiko Suzuki and Gene Myers</i>	10:1–10:20
New Algorithms for Structure Informed Genome Rearrangement	
<i>Eden Ozery, Meirav Zehavi, and Michal Ziv-Ukelson</i>	11:1–11:19
Fast Gapped k -mer Counting with Subdivided Multi-Way Bucketed Cuckoo Hash Tables	
<i>Jens Zentgraf and Sven Rahmann</i>	12:1–12:20

22nd International Workshop on Algorithms in Bioinformatics (WABI 2022).

Editors: Christina Boucher and Sven Rahmann

Leibniz International Proceedings in Informatics

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

A Linear Time Algorithm for an Extended Version of the Breakpoint Double Distance	
<i>Marília D. V. Braga, Leonie R. Brockmann, Katharina Klerx, and Jens Stoye</i>	13:1–13:16
Efficient Reconciliation of Genomic Datasets of High Similarity	
<i>Yoshihiro Shibuya, Djamel Belazzougui, and Gregory Kucherov</i>	14:1–14:14
WGSUniFrac: Applying UniFrac Metric to Whole Genome Shotgun Data	
<i>Wei Wei and David Koslicki</i>	15:1–15:22
Reconstructing Phylogenetic Networks via Cherry Picking and Machine Learning	
<i>Giulia Bernardini, Leo van Iersel, Esther Julien, and Leen Stougie</i>	16:1–16:22
Feasibility of Flow Decomposition with Subpath Constraints in Linear Time	
<i>Daniel Gibney, Sharma V. Thankachan, and Srinivas Aluru</i>	17:1–17:16
Prefix-Free Parsing for Building Large Tunnelled Wheeler Graphs	
<i>Adrián Goga and Andrej Baláz</i>	18:1–18:12
Pangenomic Genotyping with the Marker Array	
<i>Taher Mun, Naga Sai Kavya Vaddadi, and Ben Langmead</i>	19:1–19:17
Suffix Sorting via Matching Statistics	
<i>Zsuzsanna Lipták, Francesco Masillo, and Simon J. Puglisi</i>	20:1–20:15
A Maximum Parsimony Principle for Multichromosomal Complex Genome Rearrangements	
<i>Pijus Simonaitis and Benjamin J. Raphael</i>	21:1–21:22
Locality-Sensitive Bucketing Functions for the Edit Distance	
<i>Ke Chen and Mingfu Shao</i>	22:1–22:14
phyBWT: Alignment-Free Phylogeny via eBWT Positional Clustering	
<i>Veronica Guerrini, Alessio Conte, Roberto Grossi, Gianni Liti, Giovanna Rosone, and Lorenzo Tattini</i>	23:1–23:19
Gene Orthology Inference via Large-Scale Rearrangements for Partially Assembled Genomes	
<i>Diego P. Rubert and Marília D. V. Braga</i>	24:1–24:22
Toward Optimal Fingerprint Indexing for Large Scale Genomics	
<i>Clément Agret, Bastien Cazaux, and Antoine Limasset</i>	25:1–25:15

■ Preface

This proceedings volume contains papers presented at the 22nd Workshop on Algorithms in Bioinformatics (WABI 2022), which was held in Potsdam, Germany, September 5–7, 2022, and followed by a workshop on Computational Pangenomics, September 7–9 (without published proceedings).

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 focus of the meeting is on recent research results, including significant work-in-progress, as well as identifying and exploring directions of future research.

Over the 22 instances of WABI, computational biology has grown significantly in importance, and now computational analysis methods – some furthered significantly over the years at WABI – have been crucial for the global response to the CoViD-19 pandemics and for the development of vaccines. After two years of workshops strongly affected by the pandemic, the community was happy to meet again in person.

WABI 2022 was organized within the ALGO federation of conferences that in 2022 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), IPEC (International Symposium on Parameterized and Exact Computation), and WAOA (Workshop on Approximation and Online Algorithms).

In 2022, a total of 44 manuscripts were submitted to WABI from which 24 were selected for presentation at the conference and are included in this proceedings volume as full papers. While the number of submissions is still lower than in some previous years, it has increased in comparison to the past (pandemic) years, and many rejected submissions were of high quality and solely had to be rejected due to lack of further time slots at the conference. 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 24 papers selected for the conference underwent a thorough peer review, involving at least three (and most frequently four) 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 2022 Program Committee and their subreviewers for their participation in a very active review process with numerous exchanges that culminated in constructive review reports for the authors. We are also grateful to the WABI Steering Committee for their availability, help and advice. We thank all the conference participants, session chairs, and speakers who contributed to a great scientific program.

22nd International Workshop on Algorithms in Bioinformatics (WABI 2022).
Editors: Christina Boucher and Sven Rahmann



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

In particular, we are indebted to the keynote speaker of the conference, Leena Salmela (University of Helsinki), for her presentation “Efficient solutions to biological problems using de Bruijn graphs”. WABI 2022 is grateful for the support of the University of Potsdam and of the Hasso Plattner Institute, Potsdam. We thank the ALGO 2022 Organizing Committee for setting up the event in these times of uncertainty.

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, Wrocław), 9289 (WABI 2015, Atlanta), and 9838 (WABI 2016, Aarhus). As of 2017, they appeared in LIPIcs volumes 88 (WABI 2017, Boston), 113 (WABI 2018, Helsinki), 143 (WABI 2019, Boston), 172 (WABI 2020, virtually in Pisa) and 201 (WABI 2021, virtually in Chicago).

Christina Boucher & Sven Rahmann

■ WABI 2022 Committees

Steering Committee

Bernard Moret
EPFL, Switzerland

Vincent Moulton
University of East Anglia, UK

Jens Stoye
Bielefeld University, Germany

Tandy Warnow
University of Illinois at Urbana-Champaign

Travis Gagie
Diego Portales University

Anna Gambin
Warsaw University

Shilpa Garg
University of Copenhagen

Bjarni Halldorsson
deCODE genetics and Reykjavik University

Katharina Huber
University of East Anglia

PC Chairs

Christina Boucher
University of Florida, USA

Sven Rahmann
Saarland University, Germany

Carl Kingsford
Carnegie Mellon University

Gregory Kucherov
CNRS and LIGM

Manuel Lafond
Université de Sherbrooke

Program Committee

Tatsuya Akutsu
Kyoto University

Marco Antoniotti
Università degli Studi di Milano-Bicocca

Jasmijn Baaijens
TU Delft

Anne Bergeron
Université du Québec à Montréal

Paola Bonizzoni
Università di Milano-Bicocca

Lenore Cowen
Tufts University

Gianluca Della Vedova
Univ. degli Studi Milano-Bicocca

Daniel Doerr
Heinrich Heine University Düsseldorf

Mohammed El-Kebir
University of Illinois at Urbana-Champaign

Nadia El-Mabrouk
University of Montreal

Elodie Laine
Sorbonne Université

Yu Lin
The Australian National University

Stefano Lonardi
UC Riverside

Camille Marchet
CNRS

Guillaume Marçais
Carnegie Mellon University

Erin Molloy
University of Maryland at College Park

Vincent Moulton
University of East Anglia

Francesca Nadalin
EMBL

William Stafford Noble
University of Washington

Aida Ouangraoua
Université de Sherbrooke

22nd International Workshop on Algorithms in Bioinformatics (WABI 2022).

Editors: Christina Boucher and Sven Rahmann



Leibniz International Proceedings in Informatics

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

Ion Petre
University of Turku

Nadia Pisanti
University of Pisa

Solon Pissis
CWI Amsterdam

Alberto Policriti
University of Udine

Mihai Pop
University of Maryland

Teresa Przytycka
National Center of Biotechnology
Information, NLM, NIH

Knut Reinert
FU Berlin

Eric Rivals
CNRS and LIRMM Montpellier

Sebastien Roch
UW Madison

Giovanna Rosone
Università di Pisa

Alexander Schoenhuth
Bielefeld University

Mingfu Shao
Carnegie Mellon University

Krister Swenson
CNRS and Université de Montpellier

Ewa Szczurek
University of Warsaw

Sharma V. Thankachan
University of Central Florida

Hélène Touzet
CNRS and CRISAL, Lille

Tomas Vinar
Comenius University in Bratislava

Prudence Wong
University of Liverpool

Simone Zaccaria
UCL Cancer Institute

Louxin Zhang
National University of Singapore

Michael Ziv-Ukelson
Ben Gurion University of the Negev

Subreviewers

Paniz Abedin

Giulia Bernardini

Konstantinn Bonnet

Broňa Brejová

Laurent Bulteau

Sakshar Chakravarty

Ibrahim Chegrane

Ke Chen

Wei Chen

Seth Commichaux

Dan DeBlasio

Luca Denti

Abigail Djossou

Norbert Dojer

Hannes P. Eggertsson

Estéban Gabory

Askar Gafurov

Ali Ghaffaari

Daniel Gibney

Krzysztof Gogolewski

Khodor Hannoush

Quang Minh Hoang

Guillaume Holley

Dmitry Kosolobov

Benjamin Langmead

Felipe A. Louza

Anna Macioszek

Amirsadra Mohseni

Harihara Subrahmaniam Muralidharan

Njagi Mwaniki

Wend Yam Donald Davy Ouédraogo

Fabio Pardi

Luca Parmigiani

Murray Patterson

Yuri Pirola

Nicola Prezza

Simone Procaccia

Yutong Qiu

Raffaella Rizzi

Johannes Schlüter

Tizian Schulz

Remy Schwab

Saleh Sereshki

Yihang Shen

Michał Startek

Aakash Sur

Laura Tung

Gianvito Urgese

Sebastian Wild

Roland Wittler

Ran Zhang

Hongyu Zheng

ALGO Organizing Committee

Tobias Friedrich (chair)

Simon Krogmann

Timo Kötzing

Gregor Lagodzinski

Pascal Lenzner

(all at Hasso Plattner Institute,
University of Potsdam)

