24th International Workshop on Algorithms in Bioinformatics

WABI 2024, September 2–4, 2024, Royal Holloway, London, United Kingdom

Edited by

Solon P. Pissis

Wing-Kin Sung

Editors

Solon P. Pissis
CWI, Amsterdam, The Netherlands
Vrije Universiteit, Amsterdam, The Netherlands
solon.pissis@cwi.nl

Wing-Kin Sung
The Chinese University of Hong Kong, China
Hong Kong Genome Institute, Hong Kong Science Park, China
kwksung@cuhk.edu.hk

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 978-3-95977-340-9

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/978-3-95977-340-9.

Publication date
August, 2024

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.2024.0

ISBN 978-3-95977-340-9 **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 (Reykjavik University, IS and Gran Sasso Science Institute, IT)
- Christel Baier (TU Dresden, DE)
- Roberto Di Cosmo (Inria and Université Paris Cité, FR)
- Faith Ellen (University of Toronto, CA)
- Javier Esparza (TU München, DE)
- Daniel Král' (Masaryk University, Brno, CZ)
- Meena Mahajan (*Chair*, Institute of Mathematical Sciences, Chennai, IN)
- Anca Muscholl (University of Bordeaux, FR)
- Chih-Hao Luke Ong (Nanyang Technological University, SG)
- 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)
- Pierre Senellart (ENS, Université PSL, Paris, FR)

ISSN 1868-8969

https://www.dagstuhl.de/lipics

Contents

Preface
Solon P. Pissis and Wing-Kin Sung 0:vii

WABI 2024 Committees
..... 0:ix

Invited Talk

Bioinformatics of Pathogens
Tomáš Vinař 1:1–1:2

Regular Papers

On the Complexity of the Median and Closest Permutation Problems
Luís Cunha, Ignasi Sau, and Uéverton Souza 2:1–2:23

An Efficient Algorithm for the Reconciliation of a Gene Network and Species Tree
Yao-ban Chan 3:1–3:17

MEM-Based Pangenome Indexing for k-mer Queries
Stephen Hwang, Nathaniel K. Brown, Omar Y. Ahmed, Katharine M. Jenike, Sam Kovaka, Michael C. Schatz, and Ben Langmead 4:1–4:17

Memoization on Shared Subtrees Accelerates Computations on Genealogical Forests
Lukas Hübner and Alexandros Stamatakis 5:1–5:22

Cosine Similarity Estimation Using FracMinHash: Theoretical Analysis, Safety Conditions, and Implementation
Mahmudur Rahman Hera and David Koslicki 6:1–6:16

Sapling: Inferring and Summarizing Tumor Phylogenies from Bulk Data Using Backbone Trees
Yuanyuan Qi and Mohammed El-Kebir 7:1–7:19

Applying the Safe-And-Complete Framework to Practical Genome Assembly
Sebastian Schmidt, Santeri Toivonen, Paul Medvedev, and Alexandru I. Tomescu 8:1–8:16

Orientability of Undirected Phylogenetic Networks to a Desired Class: Practical Algorithms and Application to Tree-Child Orientation
Tsuyoshi Urata, Manato Yokoyama, and Momoko Hayamizu 9:1–9:17

b-move: Faster Bidirectional Character Extensions in a Run-Length Compressed Index
Lore Depuydt, Luca Renders, Simon Van de Vyver, Lennart Veys, Travis Gagie, and Jan Fostier 10:1–10:18

The mod-minimizer: A Simple and Efficient Sampling Algorithm for Long k-mers
Ragnar Groot Koerkamp and Giulio Ermanno Pibiri 11:1–11:23

24th International Workshop on Algorithms in Bioinformatics (WABI 2024).
Editors: Solon P. Pissis and Wing-Kin Sung
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Reconstructing Rearrangement Phylogenies of Natural Genomes
Leonard Bohnenkämper, Jens Stoye, and Daniel Dörr 12:1–12:16

PLA-index: A k-mer Index Exploiting Rank Curve Linearity
Md. Hasin Abrar and Paul Medvedev 13:1–13:18

A Unifying Taxonomy of Pattern Matching in Degenerate Strings and Founder Graphs
Rocco Ascone, Giulia Bernardini, Alessio Conte, Massimo Equi, Esteban Gabory, Roberto Grossi, and Nadia Pisanti 14:1–14:21

Swiftly Identifying Strongly Unique k-Mers
Jens Zentgraf and Sven Rahmann 15:1–15:15

Finding Maximum Common Contractions Between Phylogenetic Networks
Bertrand Marchand, Nadia Tahiri, Olivier Tremblay-Savard, and Manuel Lafond 16:1–16:24

A*PA2: Up to 19× Faster Exact Global Alignment
Ragnar Groot Koerkamp 17:1–17:25

RNA Triplet Repeats: Improved Algorithms for Structure Prediction and Interactions
Kimon Boehmer, Sarah J. Berkemer, Sebastian Will, and Yann Ponty 18:1–18:23

RNA Inverse Folding Can Be Solved in Linear Time for Structures Without Isolated Stacks or Base Pairs
Théo Boury, Laurent Bulteau, and Yann Ponty 19:1–19:23

The Path-Label Reconciliation (PLR) Dissimilarity Measure for Gene Trees
Alitzel López Sánchez, José Antonio Ramírez-Rafael, Alejandro Flores-Lamas, Maribel Hernández-Rosales, and Manuel Lafond 20:1–20:21

McDag: Indexing Maximal Common Subsequences in Practice
Giovanni Buzzega, Alessio Conte, Roberto Grossi, and Giulia Punzi 21:1–21:18

Anchorage Accurately Assembles Anchor-Flanked Synthetic Long Reads
Xiaofei Carl Zang, Xiang Li, Kyle Metcalfe, Tuval Ben-Yehezkel, Ryan Kelley, and Mingfu Shao 22:1–22:17

AlfaPang: Alignment Free Algorithm for Pangenome Graph Construction
Adam Cicherski, Anna Lisiecka, and Norbert Dojer 23:1–23:18

Preface

This proceedings volume contains papers and abstracts presented at the 24th Workshop on Algorithms in Bioinformatics (WABI 2024), which was held in Egham, United Kingdom, from September 2nd to 4th, 2024.

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 serves as a forum for presenting new insights about discrete algorithms and machine-learning methods that address important problems in biology, particularly those based on molecular data and phenomena. These algorithms are founded on sound models, exhibit computational efficiency, and 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 for future research. Over the 24 instances of WABI, computational biology has significantly grown in importance. Some of these methods have been applied in different biological and clinical domains like vaccine development.

WABI 2024 was co-located with ALGO 2024. A total of 46 manuscripts were submitted to WABI 2024, of which 22 were selected for presentation at the conference and included in this proceedings volume as full papers. The 22 papers chosen for the conference underwent a rigorous peer review process, involving at least four independent reviewers per submitted paper. Following this, discussions took place among the WABI 2024 Program Committee members. The reviews were conducted by both Program Committee members and additional reviewers selected based on their expertise relevant to specific papers. The selected papers encompass a wide range of topics, including phylogenetic trees and networks, cancer phylogenetics, sequence alignment and assembly, gene and genomic-level evolution, genome rearrangement, sequence and genome analysis, RNA structure, pattern matching, data compression, and more. The conference program also featured an invited talk. The EasyChair system was utilized for handling submissions and reviews.

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.

We extend our gratitude to all the authors of the submitted papers, whose contributions made this conference possible. Special thanks are due to all the members of the WABI 2024 Program Committee and their sub-reviewers for their diligent efforts in reviewing the manuscripts and engaging in comprehensive discussions. These discussions informed the decision-making process and resulted in constructive review reports for the authors. We are grateful to the WABI Steering Committee for their availability, assistance, and guidance. Our thanks go to all the conference participants, session chairs, and speakers who contributed to a highly successful scientific program. In particular, we express our gratitude to the conference's keynote speaker, Tomáš Vinař (Comenius University in Bratislava, Slovakia), for his presentation titled "Bioinformatics of Pathogens". Furthermore, we would like to express our special appreciation to the ALGO 2024 Organizing Committee for their dedicated efforts and generosity in organizing the event.

Previous WABI proceedings appeared in LNCS/LNBI volumes as follows: 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

24th International Workshop on Algorithms in Bioinformatics (WABI 2024).
Editors: Solon P. Pissis and Wing-Kin Sung
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

(WABI 2013, Sophia Antipolis), 8701 (WABI 2014, Wroclaw), 9289 (WABI 2015, Atlanta), and 9838 (WABI 2016, Aarhus). Beginning in 2017, they appeared in LIPIcs volumes: 88 (WABI 2017, Boston), 113 (WABI 2018, Helsinki), 143 (WABI 2019, Boston), 172 (WABI 2020, held virtually in Pisa), 201 (WABI 2021, held virtually in Chicago), 242 (WABI 2022, Potsdam) and 273 (WABI 2023, Houston).

Solon P. Pissis and Wing-Kin Sung

WABI 2024 Committees

Steering Committee

Vincent Moulton
University of East Anglia, UK

Nadia Pisanti
University of Pisa, Italy

Jens Stoye
Bielefeld University, Germany

Mona Singh
Princeton University, USA

PC Chairs

Solon P. Pissis
CWI, The Netherlands; Vrije Universiteit, The Netherlands

Wing-Kin Sung
Chinese University of Hong Kong, Hong Kong

Program Committee

Tatsuya Akutsu
Kyoto University, Japan

Jarno Alanko
University of Helsinki, Finland

Lorraine Ayad
Brunel University London, UK

Jasmijn Baaijens
TU Delft, The Netherlands

Giulia Bernardini
Università di Trieste, Italy

Broňa Brejová
Comenius University in Bratislava, Slovakia

Panagiotis Charalampopoulos
Birkbeck, University of London, UK

Rayan Chikhi
CNRS, France

Lenore Cowen
Tufts University, USA

Manuel Cáceres
Aalto University, Finland

24th International Workshop on Algorithms in Bioinformatics (WABI 2024).
Editors: Solon P. Pissis and Wing-Kin Sung
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Daniel Doerr
Heinrich Heine University Düsseldorf, Germany

Mohammed El-Kebir
University of Illinois at Urbana-Champaign, USA

Jonas Ellert
ENS - PSL, France

Anna Gambin
Institute of Informatics, Warsaw University, Poland

Shilpa Garg
Technical University of Denmark, Denmark

Bjarni Halldorsson
deCODE genetics and Reykjavik University, Iceland

Wing-Kai Hon
National Tsing Hua University, Taiwan

Carl Kingsford
Carnegie Mellon University, USA

Gregory Kucherov
CNRS/LIGM, France

Dominik Köppl
University of Yamanashi, Japan

Manuel Lafond
Université de Sherbrooke

Francesca Nadalin
EMBL, UK

Yuri Pirola
DISCo, Univ. degli Studi di Milano-Bicocca, Italy

Nicola Prezza
Ca' Foscari University, Italy

Mingfu Shao
The Pennsylvania State University, USA

Sharma V. Thankachan
North Carolina State University, USA

Prudence Wong
University of Liverpool, UK

Simone Zaccaria
UCL Cancer Institute, UK

Meirav Zehavi
Ben-Gurion University, Israel

Jie Zheng
ShanghaiTech University, China

Subreviewers

Eloi Araujo
Rocco Ascone
Andrej Baláž
Elena Biagi
Maria Bochenek
Christina Boucher
Davide Cenzato
Cedric Chauve
Ke Chen
Neo Christopher Chung
Davide Cozzi
Gianluca Della Vedova
Riccardo Dondi
Shiyi Du
Shane Elder
Estéban Gabory
Arnab Ganguly
Krzysztof Gogolewski
Khodor Hannoush
Peter Heringer
Guillaume Holley
Aaron Hong
Ferdoos Hossein Nezhad
Jesper Jansson
Mark Jones
Rebecca Katharina Pfeil
Sung-Hwan Kim
Bojana Kodric
Aleksandr Koshkarov
Oded Lachish
Benjamin Langmead
Téo Lemane
Jiayi Li
Xiang Li
Josipa Lipovac
Alitzel López Sánchez
Ting Lu
Guillaume Marçais
Bertrand Marchand
Camille Marchet
Pierre Marijon
Robert Mercas
Manal Mohamed
Rahul Nihalani
Hannes P. Eggertsson
Gunnar Pálsson
Karol Pokorski
Simon Puglisi
Daniel Puttini
Yuanyuan Qi
Jakub Radoszewski
Nicola Rizzo
Duncan Robertson
Yingfan Rui
Kristoffer Sahlin
Yoshifumi Sakai
Jannik Schestag
Francisco Sena
Rahul Shah
Yihang Shen
Qian Shi
Jouni Sirén
Yuyao Song
Michał Startek
Nadia Tahiri
Siyu Tao
Haotian Teng
Yue Teng
Carlo Tosoni
Jasper van Bemmelen
Leo van Iersel
Anna Vathrakokoili Pournara
Hilde Verbeek
Riccardo Vicedomini
Tomas Vinar
Mathias Weller
Qingyu Yang
Jiale Yu
Xin Yuan
Tasfia Zahin
Xiaofei Zang