

# 18th International Workshop on Algorithms in Bioinformatics

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

Edited by

Laxmi Parida

Esko Ukkonen



*Editors*

Laxmi Parida	Esko Ukkonen
IBM T. J. Watson Research Center	University of Helsinki
parida@us.ibm.com	esko.ukkonen@helsinki.fi

*ACM Classification 2012*

Applied computing → Bioinformatics, Theory of computation → Design and analysis of algorithms, Mathematics of computing → Probabilistic inference problems, Computing methodologies → Machine learning approaches

**ISBN 978-3-95977-082-8**

*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/978-3-95977-082-8>.

*Publication date*

August, 2018

*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://dnb.d-nb.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.2018.0

ISBN 978-3-95977-082-8

ISSN 1868-8969

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

## ■ Contents

Preface	
<i>Laxmi Parida and Esko Ukkonen</i> .....	0:vii
A Duality-Based Method for Identifying Elemental Balance Violations in Metabolic Network Models	
<i>Hooman Zabeti, Tamon Stephen, Bonnie Berger, and Leonid Chindelevitch</i> .....	1:1–1:13
Prefix-Free Parsing for Building Big BWTs	
<i>Christina Boucher, Travis Gagie, Alan Kuhnle, and Giovanni Manzini</i> .....	2:1–2:16
Detecting Mutations by eBWT	
<i>Nicola Prezza, Nadia Pisanti, Marinella Sciortino, and Giovanna Rosone</i> .....	3:1–3:15
Haplotype-aware graph indexes	
<i>Jouni Sirén, Erik Garrison, Adam M. Novak, Benedict Paten, and Richard Durbin</i> .....	4:1–4:13
Reconciling Multiple Genes Trees via Segmental Duplications and Losses	
<i>Riccardo Dondi, Manuel Lafond, and Celine Scornavacca</i> .....	5:1–5:16
Protein Classification with Improved Topological Data Analysis	
<i>Tamal K. Dey and Sayan Mandal</i> .....	6:1–6:13
A Dynamic Algorithm for Network Propagation	
<i>Barak Sternberg and Roded Sharan</i> .....	7:1–7:13
New Absolute Fast Converging Phylogeny Estimation Methods with Improved Scalability and Accuracy	
<i>Qiuyi (Richard) Zhang, Satish Rao, and Tandy Warnow</i> .....	8:1–8:12
An Average-Case Sublinear Exact Li and Stephens Forward Algorithm	
<i>Yohei M. Rosen and Benedict J. Paten</i> .....	9:1–9:13
External memory BWT and LCP computation for sequence collections with applications	
<i>Lavinia Egidi, Felipe A. Louza, Giovanni Manzini, and Guilherme P. Telles</i> .....	10:1–10:14
Kermit: Guided Long Read Assembly using Coloured Overlap Graphs	
<i>Riku Walve, Pasi Rastas, and Leena Salmela</i> .....	11:1–11:11
A Succinct Solution to Rmap Alignment	
<i>Martin D. Muggli, Simon J. Puglisi, and Christina Boucher</i> .....	12:1–12:16
Spalter: A Meta Machine Learning Approach to Distinguish True DNA Variants from Sequencing Artefacts	
<i>Till Hartmann and Sven Rahmann</i> .....	13:1–13:8
Essential Simplices in Persistent Homology and Subtle Admixture Detection	
<i>Saugata Basu, Filippo Utro, and Laxmi Parida</i> .....	14:1–14:10
Minimum Segmentation for Pan-genomic Founder Reconstruction in Linear Time	
<i>Tuukka Norri, Bastien Cazaux, Dmitry Kosolobov, and Veli Mäkinen</i> .....	15:1–15:15

Multiple-Choice Knapsack for Assigning Partial Atomic Charges in Drug-Like Molecules <i>Martin S. Engler, Bertrand Caron, Lourens Veen, Daan P. Geerke, Alan E. Mark, and Gunnar W. Klau</i> .....	16:1–16:13
$\ell_1$ -Penalised Ordinal Polytomous Regression Estimators with Application to Gene Expression Studies <i>Stéphane Chrétien, Christophe Guyeux, and Serge Moulin</i> .....	17:1–17:13
Differentially Mutated Subnetworks Discovery <i>Morteza Chalabi Hajkarim, Eli Upfal, and Fabio Vandin</i> .....	18:1–18:14
$D_{\text{GEN}}$ : A Test Statistic for Detection of General Introgression Scenarios <i>Ryan A. Leo Elworth, Chabrielle Allen, Travis Benedict, Peter Dulworth, and Luay Nakhleh</i> .....	19:1–19:13
PRINCE: Accurate Approximation of the Copy Number of Tandem Repeats <i>Mehrdad Mansouri, Julian Booth, Margaryta Vityaz, Cedric Chauve, and Leonid Chindelevitch</i> .....	20:1–20:13
Degenerate String Comparison and Applications <i>Mai Alzamel, Lorraine A. K. Ayad, Giulia Bernardini, Roberto Grossi, Costas S. Iliopoulos, Nadia Pisanti, Solon P. Pissis, and Giovanna Rosone</i> .....	21:1–21:14
A Multi-labeled Tree Edit Distance for Comparing “Clonal Trees” of Tumor Progression <i>Nikolai Karpov, Salem Malikic, Md. Khaledur Rahman, and S. Cenk Sahinalp</i> ....	22:1–22:19
Heuristic Algorithms for the Maximum Colorful Subtree Problem <i>Kai Dührkop, Marie A. Lataretu, W. Timothy J. White, and Sebastian Böcker</i> ...	23:1–23:14
Parsimonious Migration History Problem: Complexity and Algorithms <i>Mohammed El-Kebir</i> .....	24:1–24:14
The Wasserstein Distance as a Dissimilarity Measure for Mass Spectra with Application to Spectral Deconvolution <i>Szymon Majewski, Michał Aleksander Ciach, Michał Startek, Wanda Niemyska, Błażej Miasojedow, and Anna Gambin</i> .....	25:1–25:21