

Genome-Scale Algorithm Design

Biological Sequence Analysis in the Era of High-Throughput Sequencing

High-throughput sequencing has revolutionized the field of biological sequence analysis. Its application has enabled researchers to address important biological questions, often for the first time.

This book provides an integrated presentation of the fundamental algorithms and data structures that power modern sequence analysis workflows. The topics covered range from the foundations of biological sequence analysis (alignments and hidden Markov models), to classical index structures (k-mer indexes, suffix arrays, and suffix trees), Burrows–Wheeler indexes, graph algorithms, and a number of advanced omics applications. The chapters feature numerous examples, algorithm visualizations, exercises, and problems, each chosen to reflect the steps of large-scale sequencing projects, including read alignment, variant calling, haplotyping, fragment assembly, alignment-free genome comparison, transcript prediction, and analysis of metagenomic samples. Each biological problem is accompanied by precise formulations and complexity analyses, providing graduate students and researchers in bioinformatics and computer science with a powerful toolkit for the emerging applications of high-throughput sequencing.

The book is accompanied by a website (www.genome-scale.info) offering LaTeX source files for the exercises, along with relevant links.

Veli Mäkinen is a Professor of Computer Science at the University of Helsinki, Finland, where he heads a research group working on genome-scale algorithms as part of the Finnish Center of Excellence in Cancer Genetics Research. He has taught advanced courses on string processing, data compression, biological sequence analysis, along with introductory courses on bioinformatics.

Djamal Belazzougui is a postdoctoral researcher at the University of Helsinki. His research topics include hashing, succinct and compressed data structures, and string algorithms.

Fabio Cunial is a postdoctoral researcher at the University of Helsinki. His research focuses on string algorithms and genome analysis.

Alexandru I. Tomescu is a postdoctoral researcher at the University of Helsinki. His current research interests lie at the intersection of computational biology and computer science.

Genome-Scale Algorithm Design

Biological Sequence Analysis in the Era
of High-Throughput Sequencing

VELI MÄKINEN
DJAMAL BELAZZOUGUI
FABIO CUNIAL
ALEXANDRU I. TOMESCU

University of Helsinki, Finland



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107078536

© V. Mäkinen, D. Belazzougui, F. Cunial and A. I. Tomescu 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Mäkinen, Veli, author.

Genome-scale algorithm design : biological sequence analysis in the era of high-throughput sequencing / Veli Mäkinen, Djamal Belazzougui, Fabio Cunial, and Alexandru I. Tomescu.
p. ; cm.

Includes bibliographical references and index.

ISBN 978-1-107-07853-6 (hardback : alk. paper)

I. Belazzougui, Djamal, author. II. Cunial, Fabio, author. III. Tomescu, Alexandru I., author.

IV. Title.

[DNLM: 1. Genomics. 2. Algorithms. 3. Sequence Analysis--methods. QU 460]

QH447

572.8'629--dc23 2014045039

ISBN 978-1-107-07853-6 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.