Tandy Warnow
Editor

Bioinformatics and Phylogenetics

Seminal Contributions of Bernard Moret

Springer

Editor
Tandy Warnow
University of Illinois at Urbana-Champaign
Urbana, IL, USA

ISSN 1568-2684 ISSN 2662-2432 (electronic)
Computational Biology
ISBN 978-3-030-10836-6 ISBN 978-3-030-10837-3 (eBook)
https://doi.org/10.1007/978-3-030-10837-3

Library of Congress Control Number: 2018966860

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to Bernard M. E. Moret on his retirement.

Preface

This Festschrift is in honor of Bernard M. E. Moret, whose retirement from Ecole Polytechnique Fédérale de Lausanne (EPFL) in December 2016 culminated a nearly 40-year career. Bernard's research spanned several areas in computer science, including algorithm engineering, high-performance computing, and algorithmic computational biology. Many of Bernard's contributions were concerned with inferring and using phylogenies (i.e., evolutionary trees and phylogenetic networks), especially for large and challenging datasets. His work in genome rearrangement phylogeny is the best known, where he has been one of a small number of leading researchers in establishing theory and developing novel methods and open-source software based on rigorous mathematical theory. In addition, he has also contributed to the understanding and developing divide-and-conquer strategies, supertree construction, absolute fast converging phylogeny estimation, phylogenetic networks, and the use of phylogenies to answer biological questions. He has also trained (both directly and indirectly) many of the leading people in computational biology, including several of the contributors to this volume. For example, Alexandros Stamatakis and Daniel Doerr were postdoctoral researchers of Bernard's; Jijun Tang was his Ph.D. student; and Mark Holder, Luay Nakhleh, and Sébastien Roch were funded by the CIPRES project (see http://www.phylo.org), which Moret directed from 2003–2006.

Two conferences were organized in honor of Bernard's retirement: the first, organized by his current and former students and postdocs, was held at EPFL on November 7–8, 2016 (see https://lcbb.epfl.ch/climb/), and the second, which I organized, was held at Berkeley on June 2, 2017 (see http://tandy.cs.illinois.edu/Moret-Festschrift.html). These two conferences led to this Festschrift.

The chapters in this volume represent some of the areas in which Bernard's work has had an influence, including methods for phylogenetic tree and network estimation, genome rearrangements, cancer phylogeny, species trees, divide-and-conquer strategies, and the use of integer linear programming in computational biology. Each chapter provides an introduction to a cutting-edge problem in computational biology that is computationally and mathematically interesting. Hence, the volume is designed to be useful as a text for a graduate course in

computational biology and bioinformatics, aiming at computer scientists, applied mathematicians, and statisticians. Although much of the work that is described is advanced, each chapter provides references to the literature to enable the reader to obtain additional background, if needed.

The chapters are ordered based on the statistical complexity of the problem they address. The first three chapters address challenges in developing accurate and efficient software for the NP-hard maximum likelihood phylogeny estimation problem. Chapters 1 (by Alexandros Stamatakis) and 2 (by Stéphane Guindon and Olivier Gascuel) focus on optimizing maximum likelihood codes (including but not limited to numerical optimization), while Chap. 3 (by David Bader and Kamesh Madduri) focuses on high-performance computing aspects; together, the three chapters provide complementary insights into how to design codes for this important basic problem in computational phylogenetics. Chapter 4 (by Sébastien Roch) is also about phylogeny estimation from aligned sequences, and addresses a basic statistical question: how much data does a phylogeny estimation method require (or need) to recover the true tree with high probability, as a function of the model tree parameters (e.g., number of leaves and lengths of the branches in the tree)? Roch's chapter comes with a Jupyter notebook and provides scripts for analyses and simulations. Chapter 5 (by Nadia El-Mabrouk and Emmanuel Noutahi) addresses algorithms to infer gene trees when the input includes aligned gene sequences and also a species tree. The inference of species trees is covered in Chaps. 6 and 7. Chapter 6 (which I contributed) is about divide-and-conquer strategies to scale phylogeny estimation methods to large datasets, and specifically addresses limitations in current supertree methods. Chapter 7 (by Benjamin Redelings and Mark Holder) is about taxonomic supertrees and the challenge of constructing them when some taxa in the input have unknown placements within a taxonomic hierarchy. Chapter 8 (by Santos Muños et al.) addresses the inference of ultrametric distances from additive distance matrices, and so is related to the problem of assigning dates to internal nodes in phylogenetic trees. Chapters 9 (by Jiijun Tang) addresses the inference of ancestral genomes under genome rearrangement events. Chapters 10 (by Ron Zeira and Ron Shamir) and 11 (by Russell Schwartz) address complementary approaches to inferring evolutionary histories in cancer; Zeira and Shamir focus on how chromosomal rearrangements can be used as indicators of evolutionary history and Schwartz provides an overview of different approaches to tumor phylogenetics. Chapters 12 (by Louxin Zhang) and 13 (by R. A. Leo Elworth et al.) examine problems in phylogenetic networks, with Zhang focusing on discrete mathematics questions and Elworth et al. addressing statistical estimation issues. Chapter 14 (by Daniel Doerr and Jens Stoye) uses evolution to provide a framework within which to understand comparative and functional genomics. Chapter 15 (by Dan Gusfield) provides an introduction to Integer Linear Programming and its use in computational biology, with specific focus on how ILP can be used to solve the NP-hard Traveling Salesman Problem; his chapter also comes with software and datasets.

While the chapters are designed to be self-contained (and each contains a substantial bibliography to enable the reader to get additional background), some background in computer science (algorithms and running time analysis) and statistics (e.g., the use of probabilistic models and statistical estimation under these models) is assumed. Background in computational phylogenetics is helpful but not required, but many readers may find it helpful to read some of the textbooks in phylogenetics. For a statistical perspective on this research area, see [1] and [3]. A computer science perspective on algorithm design for phylogeny estimation (especially for large datasets) is provided in [2].

Urbana, USA Tandy Warnow

References

1. Felsenstein, J.: Inferring Phylogenies. Sinauer Associates, Sunderland, Massachusetts (2004)
2. Warnow, T.: Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation. Cambridge University Press, Cambridge, UK (2018)
3. Yang, Z.: Molecular Evolution: A Statistical Approach. Oxford University Press, Oxford (2014)

Contents