Bernhard Haubold
Thomas Wiehe

Introduction to Computational Biology

An Evolutionary Approach

Birkhäuser Verlag
Basel · Boston · Berlin

Bernhard Haubold
Department of Biotechnology and Informatics
University of Applied Sciences Weihenstephan
85350 Freising
Germany

Thomas Wiehe
Institut für Genetik
Universität zu Köln
Zülpicher Strasse 47
50674 Köln
Germany

A CIP catalogue record for this book is available from the Library of Congress, Washington D.C., USA

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 10: 3-7643- 6700-8
ISBN 13: 978-3-7643-6700-8

Birkhäuser Verlag, Basel – Boston – Berlin

The publisher and editor can give no guarantee for the information on drug dosage and administration contained in this publication. The respective user must check its accuracy by consulting other sources of reference in each individual case.
The use of registered names, trademarks etc. in this publication, even if not identified as such, does not imply that they are exempt from the relevant protective laws and regulations or free for general use.
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of use, permission of the copyright owner must be obtained.

© 2006 Birkhäuser Verlag, P.O. Box 133, CH-4010 Basel, Switzerland
Part of Springer Science+Business Media
Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Printed in Germany
Cover illustration: Simulation of gene genealogies under the Wright-Fisher model of evolution. Each gene (dot) is linked to exactly one ancestral gene in the proceding generation. In addition, it may be linked to one or more descendants in subsequent generations.
ISBN 10: 3-7643-6700-8 e-ISBN 10: 3-7643-7387-3
ISBN 13: 978-3-7643-6700-8 e-ISBN 13: 978-3-7643-7387-0

9 8 7 6 5 4 3 2 1 www.birkhauser.ch

To Angelika and Claudia

Preface

In 1982, the first release of the GenBank sequence database contained 601,438 residues. By 2005, this number had grown beyond 10^{11} and continues to increase exponentially. Far from regarding this as "information overload", we believe the free availability of so much precise and fundamental data on the ultimate constituents of life to be the hallmark of a golden age in biomedical research. Computational biology is concerned with helping to understand these data.

The aim of this book is to give a first introduction to the computational aspects of genome-scale molecular biology, also known as genomics. The interpretation of biological data is often contingent on an understanding of the evolutionary history that has generated it. Hence, we explain evolutionary models as well as classical sequence analysis.

Our intended audience is primarily students of bioinformatics, as well as researchers and students in neighboring disciplines including molecular biology, genetics, medicine, physics, mathematics, and computer science. As background, we assume familiarity with basic general and molecular biology as well as elementary probability theory. We also expect an interest in computers and their programming.

In writing this book we have benefited from the expertise and support of a number of colleagues. Clemens Beckstein invited us in 1999 to give our first lecture series on computational biology at Jena University. Wolfgang Stephan and Monty Slatkin encouraged us to turn the lecture notes accumulated in Jena into a textbook. Steffi Gebauer-Jung helped with some of the algorithms we present. Claudia Acquisti, Frank Leßke, Peter Pfaffelhuber, Karl Schmid, and Daniel Zivkovic commented on earlier versions of the manuscript. Our students improved our teaching of computational biology over the years. Finally, we owe a huge debt of gratitude to Angelika Börsch-Haubold, who edited the entire book, compiled the index and guided this project through the production stage. Without her contribution there would be no book.

Freising and Köln,
March 2006

Bernhard Haubold
Thomas Wiehe

Contents