

Preface

Overview and Goals

This book describes how to visualize and compare bacterial genomes. Sequencing technologies are becoming so inexpensive that soon going for a cup of coffee will be more expensive than sequencing a bacterial genome. Thus, there is a very real and pressing need for high-throughput computational methods to compare hundreds and thousands of bacterial genomes.

It is a long road from molecular biology to systems biology, and in a sense this text can be thought of as a path bridging these fields. The goal of this book is to provide a coherent set of tools and a methodological framework for starting with raw DNA sequences and producing fully annotated genome sequences, and then using these to build up and test models about groups of interacting organisms within an environment or ecological niche.

Organization and Features

The text is divided into four main parts: Introduction, Comparative Genomics, Transcriptomics and Proteomics, and finally Microbial Communities. The first five chapters are introductions of various sorts. Each of these chapters represents an introduction to a specific scientific field, to bring all readers up to the same basic level before proceeding on to the methods of comparing genomes. First, a brief overview of molecular biology and of the concept of sequences as biological information are given. The equivalent in the post-genomics era of the ‘Central Dogma’ of molecular biology (DNA makes RNA makes protein) is that the genome makes the transcriptome, which makes the proteome. Before going on to the details of this, a historical background is provided that pictures the scene of the origins of molecular biology and biological sequences. After this introduction, Chapter 2 describes sequence alignment, the most common procedure used to compare biological sequences. Instead of going into technical details of how exactly these alignments are calculated, the text focuses on their practical use. Chapter 3 introduces bacterial genomes and Chapter 4 deals with the most important databases, whilst Chapter 5 is

an introduction to the computational background of the tools necessary to analyze all of this information.

The second part, on Comparative Genomics (Chapters 6–8), describes some basic methods of comparing genomes. This section introduces various atlases building up to the ‘Genome Atlas,’ which is our standard visualization tool for representing the DNA sequence of a chromosome in a single figure, mapping the most relevant DNA properties along the chromosome. We have found such atlases very useful for mapping newly sequenced genomes and quickly visualizing regions of potential interest. The value of atlas projections is illustrated by the examples provided.

Part three (Chapters 9–11) takes the reader from genome sequences to RNA sequences (transcriptomics) to proteins (proteomics) and regulation of gene expression. An important overview of experimental results can be obtained by mapping back and visualizing the transcriptomic and proteomic data onto physical chromosomal maps. Examples illustrate how important chromosome location is, and which features can be predicted by careful analysis of genes and their surrounding sequences.

The final part (Chapters 12–14) deals with microbial communities. In a sense this can be thought of as ‘population genomics’ (as opposed to the more traditional ‘population biology’ which often focuses on only one or a few genes). First the concept of ‘pan-genome’ and ‘core genome’ is introduced (Chapter 12), followed by metagenomics (Chapter 13), and then evolution of microbial communities (Chapter 14). From a larger perspective, population genomics can provide a framework for modeling ecosystems in terms of interacting biological systems.

Target Audiences and Required Background Knowledge

The reader should have basic knowledge about computers and be able to use web interfaces. For programmers, some general knowledge of microbiology is assumed, but it is our hope that both programmers and more ‘biology-oriented’ readers will find this book helpful. Details on programming were deliberately left out; instead, the text concentrates on the use and interpretation of publicly available web tools. This book has grown out of lectures for the course in Comparative Microbial Genomics,¹ which DWU has taught since 2001 as a full semester length course at the Technical University of Denmark, and as one-week workshops given in Bangkok, Thailand; in Petropolis, Brazil; and in Oslo, Norway.

This book is in a sense merging different scientific languages. The three authors have different scientific and national backgrounds. DWU is from the U.S., studied biochemistry, worked in molecular biology, and for the last 10 years has led a group

¹ <http://www.cbs.dtu.dk/dtucourse/programme27444.php>

in bioinformatics and genomics. SB is from Italy, studied quantum chemistry with focus on scientific programming, data standardization, and software integration; whereas TMW studied biochemistry and worked in molecular biology and later as a consultant in microbiology. These different backgrounds actually helped to develop a common language in science. The subject area of this textbook is extremely interdisciplinary, covering (bio)chemistry, physics, biology, microbiology, mathematics, and computational science, and by the introduction of concepts (and some jargon) from these various disciplines, the different languages used by specialists are bridged.

This book is meant mainly for people studying bacterial genomes, although of course nearly all of the methods described in the text would work for viral, Archaeal, or Eukaryotic genomes as well. There are two main target audiences. The first is the microbiologist who wants to get the most out of a bacterial genome sequence. This could be a university student, or an experienced laboratory microbiologist who enters the field of genomics. This book enables one to get a handle on how to use high-throughput computational methods to compare only a few, or hundreds of sequenced genomes. The second audience comprises the computer programmers who assist these microbiologists in actually carrying out the analyses. From experience we know there can be communication problems between the experimental bacteriologist who is more laboratory-oriented, and the computer scientist who wants to do everything on computers. Both disciplines are essential in present-day research. This book aims to explain to the computational scientist why and how we want to study bacterial genomes, and what questions we hope to answer. At the same time, it explains to the biologist some of the basics behind the bioinformatic tools that are necessary for research in the field. Bringing these two worlds, scientific interests, and languages together is our ultimate goal.

Notes to the Instructor

There are no exercises or questions at the end of the chapters, although at the end of most chapters textboxes present descriptions of essential methods used. From experience we can say that giving small groups of students a project in which they can choose a recently sequenced bacterial genome and compare it to other similar genomes can produce surprisingly successful results. It is very motivating to work with recently published data (new genome sequence papers are being published on an almost daily basis now), and sometimes the students produce important observations that the authors of the scientific papers had missed! In some occasions, such activities have resulted in a real scientific publication by the students, illustrating how ‘easy’ it is to do these kinds of analyses, as long as one asks relevant questions.

Supplemental Resources

A number of web links are mentioned in the book, and since web addresses are not always stable, a dedicated web page is put up on which all web pages presented in the book are summarized, and as necessary, updated. This can be found at <http://comparativemicrobial.com>.

Lyngby, Denmark
Zurich, Switzerland
Zotzenheim, Germany

David Ussery
Stefano Borini
Trudy Wassenaar

Computing for Comparative Microbial Genomics

Bioinformatics for Microbiologists

Ussery, D.W.; Wassenaar, T.M.; Borini, S.

2009, XIV, 270 p. With online files/update., Hardcover

ISBN: 978-1-84800-254-8