

Cambridge University Press

978-1-107-07983-0 - Bacterial Genomics: Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee

Frontmatter

[More information](#)

Bacterial Genomics

Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-1-107-07983-0 - Bacterial Genomics: Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee

Frontmatter

[More information](#)

CAMBRIDGE
UNIVERSITY PRESS

Cambridge House, 4381/4 Ansari Road, Daryaganj, Delhi 110002, India

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/9781107079830

© Aswin Sai Narain Seshasayee 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 India

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

Library of Congress Cataloging-in-Publication Data

Seshasayee, AswinSaiNarain, author.

Bacterial genomics : genome organization and gene expression tools /

AswinSaiNarainSeshasayee.

p. ; cm.

Includes bibliographical references and index.

Summary: "Discusses the application of genomic tools in the study of bacterial adaptation and provides review of recent research in the field of bacterial research"--Provided by publisher.

ISBN 978-1-107-07983-0 (hardback)

I. Title.

[DNLM: 1. Bacteria--genetics. 2. Genome, Bacterial--genetics. 3. Gene Expression Regulation, Bacterial. 4. Gene Expression. 5. Models, Genetic. QW 51]

QH434

572.8'6293--dc23

2014027746

ISBN 978-1-107-07983-0 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.

Cambridge University Press

978-1-107-07983-0 - Bacterial Genomics: Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee

Frontmatter

[More information](#)

To

National Centre for Biological Sciences, Tata Institute of
Fundamental Research, Bangalore, India

and

Dedicated to the memory of the ever-enchanted
Professor K. S. Krishnan

Cambridge University Press

978-1-107-07983-0 - Bacterial Genomics: Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee

Frontmatter

[More information](#)

Contents

<i>List of Figures</i>	<i>ix</i>
<i>Acknowledgements</i>	<i>xiii</i>
1. Introduction: Bacterial Genomes and Gene Expression	1
2. Comparative Genomics in the Era of Sanger Sequencing	4
2.1 Introduction	4
2.2 The process of assembling and annotating bacterial genomes	6
2.2.1 Genome assembly and gap closure	7
2.2.2 Genome-scale computational identification of features	8
2.2.3 Annotating genes with functions	13
2.3 Case studies	16
2.3.1 The <i>Escherichia coli</i> complex and large-scale horizontal gene acquisition	16
2.3.2 Genome reduction in intracellular pathogens, endosymbionts and marine α -proteobacteria	20
2.3.3 The dynamic genomes of <i>Helicobacter pylori</i> and <i>Campylobacter jejuni</i>	26
2.4 Some lessons learnt from studying 2,000 bacterial genomes	27
2.4.1 Genome size	28
2.4.2 Coding density	29
2.4.3 Gene order conservation	30
2.4.4 Comparative genomics of gene functions: Systematic annotation	31
2.4.5 Comparative genomics of gene functions: Scaling laws	33
Summary	35
3. Studying Bacterial Genome Variation with Microarrays	36
3.1 Introduction	36
3.2 DNA microarrays: The concept	36

3.3	DNA microarrays: From fluorescence intensities to information	38
3.3.1	Background correction	39
3.3.2	Normalisation	41
3.3.3	Differences in signal from the same probe between two samples	44
3.4	Comparative genome hybridisation and bacterial phylogenomics	45
3.5	Case studies	46
3.5.1	Comparative genome hybridisation studies of <i>Escherichia coli</i>	47
3.5.2	Comparative genome hybridisation studies of <i>Staphylococcus aureus</i>	52
3.5.3	Comparative genome hybridisation studies of <i>Helicobacter pylori</i>	56
	Summary	59
4.	Studying Bacterial Genomes using Next-Generation Sequencing	60
4.1	Introduction	60
4.2	Next-generation sequencing technologies	61
4.2.1	Template preparation strategies	62
4.2.2	Sequencing strategies	64
4.3	Sequencing data processing for genome sequencing and re-sequencing	66
4.3.1	Genome assembly	66
4.3.2	Aligning short reads to long genomes	70
4.4	Case studies	73
4.4.1	Pyrosequencing-enabled complete genome sequence of <i>Acinetobacter baumannii</i>	74
4.4.2	On the track of pandemics: The genome of the aetiological agent of Black Death	75
4.4.3	From community genomes to complete genomes to single-cell genomes	80
4.4.4	Bacteria evolving in the laboratory	84
4.4.5	Bacteria evolving in their biotic hosts	86
	Summary	88

5	Genome-Scale Analysis of Gene Expression and its Regulation in Bacteria	89
5.1	Introduction	89
5.2	The process of transcription and the regulation of its initiation: An overview	90
5.3	Measuring gene expression on a genomic scale: Technologies	99
5.4	Next-generation sequencing for gene expression measurements: Data analysis	101
5.4.1	Transcriptome assembly	104
5.4.2	Measuring gene expression levels	110
5.5	Gene expression at high temporal resolution using fluorescent reporters	114
5.6	Constructing transcriptional regulatory networks: ChIP-chip and ChIP-seq	117
5.7	Case studies	122
5.7.1	Experimental annotation of bacterial genomes	122
5.7.2	Bioinformatic analysis of bacterial promoters	129
5.7.3	DNA topology and its interplay with gene expression	134
5.7.4	RNA polymerase occupancy and the σ -factors	145
5.7.5	Transcription factors and transcriptional regulatory networks	154
5.7.6	Transcriptional control by the small-molecule alarmone ppGpp	186
5.7.7	RNA chaperones and their regulons	188
	Summary	190
6	DNA Methylation in Bacteria: A Case for Bacterial Epigenetics	191
6.1	Introduction	191
6.2	DNA methyltransferases in bacteria: From restriction–modification systems	192
6.3	Identifying sites of DNA methylation on a genomic scale	195
6.3.1	Methylated DNA immunoprecipitation	195
6.3.2	Bisulphite sequencing	195
6.3.3	DNA cytosine methylation in laboratory <i>E. coli</i>	197

viii Contents

6.4	Detecting DNA methylation by single-molecule real-time sequencing	200
6.4.1	DNA adenine methylation in pathogenic <i>E. coli</i> by SMRT sequencing	202
6.4.2	Insight into the epigenetic control of <i>Caulobacter crescentus</i> cell cycle from SMRT sequencing	203
	Summary	205
	Index	207

List of Figures

2.1	A schematic overview of the process of sequencing a bacterial genome.	7
2.2	Base composition patterns.	9
2.3	A toy Markov model differentiating between genes and random sequences.	11
2.4	Regulatory RNA secondary structures.	12
2.5	A schematic outline of the process of annotating a bacterial genome.	13
2.6	The genome of <i>E. coli</i> K12 MG1655.	17
2.7	The two major forces of large-scale bacterial evolution.	19
2.8	Gene content of Rickettsia.	21
2.9	Gene content of Mycobacteria.	23
2.10	Metabolic potential of two co-operating bacteria.	25
2.11	Clusters of simple repeats in the genome of <i>H. pylori</i> genome.	27
2.12	Genome size and gene density in bacteria.	28
2.13	Gene order conservation.	30
2.14	Genes conserved across <i>S. aureus</i> prophages.	32
2.15	Scaling laws.	34
3.1	Steps involved in a typical microarray data processing pipeline.	38
3.2	Background subtraction for microarray data.	40
3.3	Methods for within-array normalisation of two-colour microarray data.	42
3.4	Quantile normalisation.	43
3.5	A schematic representation of a CGH experiment.	46
3.6	Evolution of <i>E. coli</i> gene content.	48
3.7	Designing a multi-genome oligonucleotide probe for CGH experiments.	52
3.8	Gene conservation in <i>Staphylococcus aureus</i> .	53
3.9	Gene content of <i>Helicobacter pylori</i> genomes.	57

x [List of Figures](#)

4.1	Sample preparation strategies for next-generation sequencers.	63
4.2	Sequencing strategies for next-generation sequencers.	65
4.3	de Bruijn graph for genome assembly.	67
4.4	Paired-end and mate-pair sequencing.	69
4.5	Burrows–Wheeler transform for read mapping.	71
4.6	Standard applications of next-generation sequencing.	73
4.7	A phylogenetic tree of <i>Yersinia pestis</i> .	77
4.8	Rapid genomic characterisation of an <i>E. coli</i> outbreak.	79
4.9	From communities to single-cell genomes.	82
4.10	Pathogen evolution in hosts.	87
5.1	Transcription in <i>E. coli</i> .	91
5.2	The promoter structure of <i>E. coli</i> .	92
5.3	DNA supercoiling.	94
5.4	σ -factors in <i>E. coli</i> .	95
5.5	Gene regulation by transcription factors.	96
5.6	Gene regulation by the small molecule-alarmon ppGpp.	97
5.7	Mechanisms of gene regulation by riboswitches.	98
5.8	Quantitative nature of deep-sequencing data.	103
5.9	Transcriptome assembly.	104
5.10	Gene expression quantification from RNA-seq experiment.	111
5.11	Fluorescence for promoter activity.	115
5.12	Chromatin immunoprecipitation.	118
5.13	Directionality in read alignments in ChIP-seq data.	121
5.14	Gene expression and RNA polymerase occupancy.	123
5.15	Overview of functional elements of <i>E. coli</i> from Ecocyc and RegulonDB.	130
5.16	Effect of DNA supercoiling on gene expression in <i>E. coli</i> .	136
5.17	Schematic representation of a chromosome conformation capture protocol.	140
5.18	Contact probability map of <i>Caulobacter crescentus</i> genome.	142
5.19	The σ -factor regulatory network of <i>Geobacter sulfurreducens</i> .	150
5.20	Transcriptional regulatory network of <i>E. coli</i> from the RegulonDB database.	157
5.21	Transcription factor binding site position and type of regulation.	158
5.22	Degree distributions in the transcriptional regulatory network of <i>E. coli</i> .	159
5.23	Transcription regulatory network motifs in <i>E. coli</i> .	163

5.24	Reconstructing transcriptional regulatory networks from a compendium of gene expression data.	167
5.25	ChIP-seq signals for three sequence-specific nucleoid-associated proteins in <i>E. coli</i> .	176
5.26	Agreement between the binding and measurable transcriptional effects of NAPs.	176
5.27	Comparison of ChIP-seq profiles of two homologous NAPs – IHF and HU.	178
5.28	Binding of H-NS to the <i>Salmonella Typhimurium</i> genome.	181
5.29	A NAP regulatory network of <i>E. coli</i> .	183
5.30	RNA targets of the RNA chaperone Hfq.	189
6.1	Restriction modification systems.	192
6.2	A model for the evolution of DNA methyltransferases in bacteria.	194
6.3	A schematic representation of deriving methylation calls from bisulphite sequencing.	196
6.4	Variation in extent of cytosine methylation with growth phase in <i>E. coli</i> .	198
6.5	IPD ratio distribution DNA methylation by SMRT sequencing.	201

Cambridge University Press

978-1-107-07983-0 - Bacterial Genomics: Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee

Frontmatter

[More information](#)

Acknowledgements

I would like to thank my host institute, the National Centre for Biological Sciences (Bangalore, India), for providing me the resources and support to not only pursue research, but also to work on this book. All the members of my laboratory have been helpful and understanding during the times when I committed more hours to writing the book than to the lab. I would like to acknowledge the efforts of the two academic genomics core facilities that have enabled my research. These are the GeneCore facility at the European Molecular Biology Laboratory, Heidelberg, and the Next Generation Genomics Facility at the Centre for Cellular and Molecular Platforms, Bangalore. Thanks to the Department of Science and Technology (their research grants and the Ramanujan Fellowship scheme) and the Department of Biotechnology, Government of India, and CEFIPRA for funding our work.

Special thanks to Avantika Lal in the laboratory for designing and making several illustrations in this book. Aalap Mogre and Vittore Scolari from the lab also made illustrations for this work. Thanks to Hardik Gala (Institute of Stem Cell Biology and Regenerative Medicine, Bangalore) for contributing an illustration from his unpublished data. Many thanks to the large number of scientists worldwide who have published their high-quality work in open access journals, permitting free reproduction of important figures, which I could have never made otherwise. These, and others who gave permission to reuse copyrighted illustrations or data, are all acknowledged in the appropriate locations in the text. Of course, thanks to the entire community of researchers working in genomics and bacteriology!

Dasaradhi Palakodeti, Madan Babu Mohan, Subhajyoti De and Marco Cosentino Lagomarsino gave feedback on parts of the manuscript. This work would not have been possible without the efforts of Manish Chaudhari, my editor at CUP India.

Big thanks to Nicholas Luscombe, my PhD supervisor, for every great thing that a fantastic PhD supervisor does to his ward, in particular, give unbridled freedom to pursue one's interests in research.

Cambridge University Press

978-1-107-07983-0 - Bacterial Genomics: Genome Organization and Gene Expression Tools

Aswin Sai Narain Seshasayee

Frontmatter

[More information](#)

xiv

[Acknowledgements](#)

My twin daughters were three months old when I started working on this book. They are now over two years old. A lot of credit to their impeccable behaviour and my wife Gayathri's efforts, which ensured that my work was not particularly affected by what was happening at home! Thanks, as always, to my parents for always being there.