METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Bacterial Pangenomics

Methods and Protocols

Edited by

Alessio Mengoni

Department of Biology, University of Florence, Florence, Italy

Marco Galardini

EMBL-EBI, Cambridge, UK

Marco Fondi

Department of Biology, University of Florence, Florence, Italy

Humana Press

Editors
Alessio Mengoni
Department of Biology
University of Florence
Florence, Italy

Marco Fondi
Department of Biology
University of Florence
Florence, Italy

Marco Galardini
EMBL-EBI
Cambridge, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
ISBN 978-1-4939-1719-8 ISBN 978-1-4939-1720-4 (eBook)
DOI 10.1007/978-1-4939-1720-4
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014951682

Printed on acid-free paper

Humana Press is a brand of Springer
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

From a pioneering field a decade ago, now bacterial genomics is a mature research interdisciplinary field, which is approached by ecologists, geneticists, bacteriologists, molecular biologists, and evolutionary biologists working in medical, industrial, and basic science. The high diffusion of bacterial genomics in many different fields has been helped by the low costs of genome and transcriptome sequencing performed by the so-called Next Generation Sequencing (NGS) technologies. Now, the cost of a draft bacterial genome sequence is as low as few hundreds of Euro (or Dollars). This low cost is allowing many laboratories to perform genome sequencing of virtually every "interesting" bacterial strain they have in hand. In parallel, bioinformatic analysis of the data has grown and the specialized bioinformatician is an obliged professional figure in every laboratory that is interested in genome sequencing.

One of the most striking differences of bacterial genomics with respect to the genomics of eukaryotic multicellular organisms is the concept of pangenome, which was introduced in the late 2005 by researchers working on bacterial pathogenic species. The pangenome is defined as a genomic approximation to describe a species' genome in terms of the sum of core (conserved in all strains) and dispensable (variable among strains) genes. For bacterial species, the pangenome concept is particularly relevant since closely related strains usually show large differences in gene content between them. Consequently, when speaking about bacterial genomics, often people are referring to comparative analysis of bacterial genomes and then to what we can call "bacterial pangenomics." Understanding which genetic components of this large pangenomic variability are functionally, clinically, or evolutionary relevant is a challenging task; in fact, a large fraction of the dispensable genome is found to have a poor functional characterization. The availability of powerful and precise analysis tools is therefore of paramount importance.

Thanks to the large diffusion of bacterial genome analysis (or bacterial pangenomic studies), the present book is intended to provide the most recent methodologies about the study of bacterial pangenomes. Three major areas are covered, namely the experimental methods for approaching bacterial pangenomics ("Preparing the bacterial pangenome"), the bioinformatic pipelines for analysis and annotation of sequence data ("Defining the pangenome"), and finally the methods for inferring functional and evolutionary features from the pangenome ("Interpreting the pangenome"). In each of these sections, researchers from both academia and private leading companies of NGS and bioinformatic analysis (as Beijing Genome Institute, Life Technologies, Era7 Bioinformatics) are providing the most up-to-date protocols and procedures for bacterial genome analysis, from assessment of genome size and structure to the analysis of raw sequence data and their annotation and biological interpretation in terms of gene activity and metabarcoding diversity and genome evolution.

The aim of the present book is then to serve as a "field guide" both for qualified investigators on bacterial genomics who want to update their technical knowledge and for less-experienced researchers who want to start working with bacterial genomics and pangenomics.

Additionally, the book could serve to graduate students as a manual of methods used in bacterial pangenomics and as a supplemental textbook in classes of genomics and bioinformatics.

Florence, Italy — *Alessio Mengoni*
Florence, Italy — *Marco Fondi*
Cambridge, UK — *Marco Galardini*

Contents

Contributors

Rosa Alduina • *Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, University of Palermo, Palermo, Italy*

Alexey Alekhin • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*

M. Ancora • *Istituto Zooprofilattico Sperimentale dell'Abruzzo e Molise "G. Caporale", National and OIE Reference Laboratory for Brucellosis, Teramo, Italy*

Giovanni Bacci • *Department of Biology, University of Florence, Florence, Italy; Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per lo Studio delle Relazioni tra Pianta e Suolo (CRA-RPS), Rome, Italy*

Dariusz Bartosik • *Institute of Microbiology, Department of Bacterial Genetics, University of Warsaw, Warsaw, Poland*

Marco Bazzicalupo • *Department of Biology, University of Florence, Florence, Italy*

Emanuele Bosi • *Department of Biology, University of Florence, Florence, Italy*

C. Cammà • *Istituto Zooprofilattico Sperimentale dell'Abruzzo e Molise "G. Caporale", National and OIE Reference Laboratory for Brucellosis, Teramo, Italy*

Duccio Cavalieri • *Department of Computational Biology, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy*

Alice Checcucci • *Department of Biology, University of Florence, Florence, Italy*

Lei Cheng • *BGI, Shenzen, China*

F. Del Chierico • *Unit of Parasitology and Unit of Metagenomics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy*

Salvatore Conti • *Thermo Fisher Scientific, Monza, Italy*

Bruno Contreras-Moreira • *Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico; Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (EEAD-CSIC), Zaragoza, Spain; Fundación ARAID, Zaragoza, Spain*

Francesca Decorosi • *Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente (DISPAA), University of Florence, Florence, Italy*

Lukasz Dziewit • *Department of Bacterial Genetics, Institute of Microbiology, University of Warsaw, Warsaw, Poland*

Kristof Engelen • *Department of Computational Biology, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy*

Renato Fani • *Department of Biology, University of Florence, Florence, Italy*

Elio Fantini • *Italian National Agency for New technologies, Energy and Sustainable development, Rome, Italy*

Alessia Fiore • *Italian National Agency for New technologies, Energy and Sustainable development, Rome, Italy*

Marco Fondi • *Department of Biology, University of Florence, Florence, Italy*

Marco Galardini • *EMBL-EBI, Cambridge, UK*

Giulio Gianese • *Italian National Agency for New technologies, Energy and Sustainable development, Rome, Italy*

Luciana Giovannetti • *Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente (DISPAA), University of Florence, Florence, Italy*

Giovanni Giuliano • *Italian National Agency for New technologies, Energy and Sustainable development, Rome, Italy*
Zhenfei Hu • *BGI, Shenzen, China*
Evdokim Kovach • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*
Kris Laukens • *Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium; Biomedical Informatics Research Center Antwerp (biomina), Antwerp University Hospital, University of Antwerp, Edegem, Belgium*
Pietro Liò • *Computer Laboratory, University of Cambridge, Cambridge, UK*
Marina Manrique • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*
M. Marcacci • *Istituto Zooprofilattico Sperimentale dell'Abruzzo e Molise "G. Caporale", National and OIE Reference Laboratory for Brucellosis, Teramo, Italy*
Emmanuela Marchi • *Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente (DISPAA), University of Florence, Florence, Italy*
Alessio Mengoni • *Department of Biology, University of Florence, Florence, Italy*
Pieter Meysman • *Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium; Biomedical Informatics Research Center Antwerp (biomina), Antwerp University Hospital, University of Antwerp, Edegem, Belgium*
Stefano Mocali • *Consiglio per la Ricerca e la sperimentazione in Agricoltura, Centro di Ricerca per l'Agrobiologia e la Pedologia (CRA-ABP), Florence, Italy*
Marco Moretto • *Department of Computational Biology, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy*
Valerio Orlandini • *Department of Biology, University of Florence, Florence, Italy; Department of Protein Biochemistry, National Research Council, Napoli, Italy*
Eduardo Pareja • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*
Pablo Pareja-Tobes • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*
Eduardo Pareja-Tobes • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*
Annalisa Pisciotta • *Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, University of Palermo, Palermo, Italy*
L. Putignani • *Unit of Parasitology and Unit of Metagenomics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy*
Ermanno Rizzi • *Institute for Biomedical Technologies (ITB), National Research Council (CNR), Segrate, MI, Italy*
Paolo Sonego • *Department of Computational Biology, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy*
Raquel Tobes • *Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain*
Pablo Vinuesa • *Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico*
Roberto Viola • *Department of Computational Biology, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy*
Carlo Viti • *Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente (DISPAA), University of Florence, Florence, Italy*
Hai Wang • *BGI, Shenzen, China*
Zhuofei Xu • *Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark*
Rui Zhou • *Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark*