Funct Integr Genomics
DOI 10.1007/s10142-006-0027-2

ORIGINAL PAPER

**Tim T. Binnewies · Yair Motro · Peter F. Hallin · Ole Lund · David Dunn · Tom La · David J. Hampson · Matthew Bellgard · Trudy M. Wassenaar · David W. Ussery**

# Ten years of bacterial genome sequencing: comparative-genomics-based discoveries

Received: 20 January 2006 / Revised: 24 February 2006 / Accepted: 7 March 2006


**Abstract** It has been more than 10 years since the first bacterial genome sequence was published. Hundreds of bacterial genome sequences are now available for comparative genomics, and searching a given protein against more than a thousand genomes will soon be possible. The subject of this review will address a relatively straightforward question: "What have we learned from this vast amount of new genomic data?" Perhaps one of the most important lessons has been that genetic diversity, at the level of large-scale variation amongst even genomes of the same species, is far greater than was thought. The classical textbook view of evolution relying on the relatively slow accumulation of mutational events at the level of individual bases scattered throughout the genome has changed. One of the most obvious conclusions from examining the sequences from several hundred bacterial genomes is the enormous amount of diversity—even in different genomes from the same bacterial species. This diversity is generated by a variety of mechanisms, including mobile genetic elements and bacteriophages. An examination of the 20 *Escherichia coli* genomes sequenced so far dramatically illustrates this, with the genome size ranging from 4.6 to 5.5 Mbp; much of the variation appears to be of phage origin. This review also addresses mobile genetic elements, including pathogenicity islands and the structure of transposable elements. There are at least 20 different methods available to compare bacterial genomes. Metagenomics offers the chance to study genomic sequences found in ecosystems, including genomes of species that are difficult to culture. It has become clear that a genome sequence represents more than just a collection of gene sequences for an organism and that information concerning the environment and growth conditions for the organism are important for interpretation of the genomic data. The newly proposed Minimal Information about a Genome Sequence standard has been developed to obtain this information.

**Keywords** Bacterial genomics · Comparative genomics · Bioinformatics · Genomic diversity · Molecular evolution

T. T. Binnewies · P. F. Hallin · O. Lund · D. W. Ussery (✉)
Center for Biological Sequence Analysis,
Technical University of Denmark,
2800 Lyngby, Denmark
e-mail: dave@cbs.dtu.dk

Y. Motro · D. Dunn · M. Bellgard
Center for Bioinformatics and Biological Computing,
Murdoch University,
Murdoch, Western Australia 6150, Australia

T. La · D. J. Hampson
School of Veterinary and Biomedical Sciences,
Murdoch University,
Murdoch, Western Australia 6150, Australia

T. M. Wassenaar
Molecular Microbiology and Genomics Consultants,
Zotzenheim, Germany

## Introduction

The year 1995 marked the publication of two human pathogenic bacterial genome sequences: *Haemophilus influenzae* (Fleischmann et al. 1995, US patent number 6,528,289) and *Mycoplasma genetalium* (Fraser et al. 1995, US patent number 6,537,773). Since then, more than 300 bacterial genomes have been fully sequenced and become publicly available, including the sequence of a virulent form of *H. influenzae* (Harrison et al. 2005); the original *H. influenzae* strain sequenced in 1995 was from an isolate that does not cause disease. Although the majority of these several hundred genomes are from pathogenic organisms, some environmental bacterial genome sequences have also become available. This review article will provide a brief overview of sequenced bacterial genomes, their genomic diversity and some of the insights gained from analysis of this vast amount of data.

Bacteria are microscopic unicellular prokaryotes that inhabit a wide variety of environmental niches, broadly distributed in three ecosystems: the soil, marine environments and other living organisms. Although there are