Genome Comparison of Bacterial Pathogens

T.M. Wassenaar[a,b] · J. Bohlin[c] · T.T. Binnewies[a,d] · D.W. Ussery[a]

[a]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark; [b]Molecular Microbiology and Genomics Consultants, Zotzenheim, Germany; [c]Norwegian School of Veterinary Science, Epi-Center, Department of Food Safety and Infection Biology, and National Veterinary Institute, Section of Epidemiology, Oslo, Norway; [d]Roche Diagnostics Ltd., Rotkreuz, Switzerland

Abstract

Bacterial pathogens are being sequenced at an increasing rate. To many microbiologists, it appears that there simply is not enough time to digest all the information suddenly available. In this chapter we present several tools for comparison of sequenced pathogenic genomes, and discuss differences between pathogens and non-pathogens. The presented tools allow comparison of large numbers of genomes in a hypothesis-driven manner. Visualization of the results is very important for clear presentation of the results and various ways of graphical representation are introduced.

Copyright © 2009 S. Karger AG, Basel

The first complete sequence of a bacterial genome was published in 1995 [1]. Since then, more than 800 bacterial and archaeal genomes have been fully sequenced and published, and in addition for more than a thousand genomes a near-to complete sequence has become publicly available. The rate at which completed bacterial genome sequences are added to the public domain is increasing with time (fig. 1, left panel). These statistics were obtained from the NCBI Genome Project web pages [2]. Pathogens comprise a large fraction of the sequenced bacterial genomes and since many of these belong to the Proteobacteria, this and a few other bacterial phyla are highly overrepresented in the available genome sequences (fig. 1, right panel). This should be borne in mind when interpreting BLAST E-values, as that program assumes an equal chance for any homology to be found by chance, whereas that chance greatly increases when searching with genes from, e.g., Proteobacteria or Firmicutes.

In this chapter we compare the sequenced genomes of pathogenic bacteria amongst each other and with non-pathogenic bacteria, using some common and relatively simple methods of comparison. Instead of zooming in on a single given genome sequence, we use tools to compare genomes within a well-defined group of related organisms, such as bacteria sharing a particular life style, or belonging to a particular species,