Visualization of pathogenicity regions in bacteria

Carsten Friis, Lars Juhl Jensen & David W. Ussery

Center for Biological Sequence Analysis, Department of Biotechnology, Building 208, The Technical University of Denmark, DK-2800 Lyngby, Denmark; *Author for correspondence (Phone: (+45) 45 25 24 88; Fax: (+45) 45 93 15 85; E-mail: dave@cbs.dtu.dk)

Key words: base composition, Escherichia coli, pathogenicity islands, Salmonella typhimurium DT104, Yersenia pestis

Abstract

We show here how pathogenicity islands can be analysed using GenomeAtlases, which is a method for visualising repeats, DNA structural characteristics, and base composition of chromosomes and plasmids. We have applied this method to the E. coli plasmid pO157, and the Y. pestis plasmid pPCP1. In both cases pathogenic genes were shown to differ in A+C content and structural properties. Furthermore, examination of an antibiotic resistance gene cluster from S. typhimurium showed that the same was true for genes encoding antibiotic resistance.

Introduction

With the availability of DNA sequences from pathogenic organisms, computational analysis of these organisms and their toxicity has become feasible. We present here a method for visualizing pathogenicity islands so that repetitive sequences and anomalies in base composition or DNA structure become visible. The GenomeAtlas is a wheelplot summarizing such different properties of DNA (Jensen, Friis & Ussery, 1999). This method can be applied to both small plasmids and gene clusters as seen in this work, but also to complete microbial chromosomes. We have created GenomeAtlases for all the fully sequenced microbial chromosomes that are publicly available. These atlases are available on the internet at http://www.cbs.dtu.dk/services/GenomeAtlas/.

To illustrate the usefulness of the GenomeAtlas for finding genes responsible for pathogenicity, three GenBank entries from different organisms were examined. pO157, a 92077 bp plasmid from the pathogenic E. coli strain O157:H7 (GenBank accession number AF074613), was chosen because it is in part responsible for the hosts pathogenicity and is believed to encode at least one toxic protein (Burland et al., 1998). pPCP1, a plasmid from Y. pestis which carries two known virulence genes, was also selected. Whereas infection of humans with strains of Y. enterocolitica and Y. pseudotuberculosis typically result in diarrhea and abdominal pains, Y. pestis with the plasmid pPCP1 is the cause of the bubonic plague (Hu et al., 1998). Finally, a cluster of five antibiotic resistance genes from S. typhimurium DT104 was selected (GenBank accession number AF071555) (Briggs & Fratamico, 1999). A recent outbreak of this multidrug resistant Salmonella in Denmark resulted in the death of two people (Molback et al., 1999). Multidrug resistant Salmonella are becoming increasingly difficult to treat with antibiotics, and may become a major health concern in the future.

Methods

To generate wheelplots, a number of parameters are calculated for the DNA double helix based on the nucleotide sequence. These parameters belong to three categories: repeats, structural parameters, and parameters directly related to the base composition. An atlas in which these parameters are visualized as colored circles is made for each of these three categories; in addition the combined GenomeAtlas summarizing the most informative parameters is constructed.
Structural parameters

A number of measures for the local structure of DNA have been devised, most of which are based on simple lookup tables of dinucleotide or trinucleotide values that have been obtained by fitting either experimental results or theoretical estimates (Pedersen et al., 2000).

Intrinsic curvature is a property of DNA that is closely related to anomalous gel mobility, as DNA fragments with high intrinsic curvature will migrate slower on polyacrylamide gels than markers with the same length. In this work we have used the CURVATURE programme (Shpigelman, Trifonov & Bolshoy, 1993), which is based on a wedge model (Trifonov & Sussman, 1980; Ulanovsky, Bodner & Trifonov, 1986), for prediction of intrinsic curvature. From a set of dinucleotide values for the twist, wedge, and direction angles the three-dimensional path of a 21-bp fragment is calculated. Curvature profiles for longer sequences can thus be calculated using a 21-bp running window. Curves are often encountered upstream of highly expressed genes such as the toxin genes in pathogenicity islands (Bracco et al., 1989).

Next is the stacking energy, which relates to the interaction energy between adjacent basepairs in the DNA double helix. The total stacking energy of a DNA segment can be estimated from the set of dinucleotide values determined by quantum mechanical calculations on crystal structures (Ornstein et al., 1978). All stacking energies are negative since base stacking is an energetically favourable interaction that serves to stabilise the double helix. This means that regions with large stacking energies are strongly stabilised and therefore less likely to destack or melt than regions with less negative stacking energies.

The position preference is a measure of helix flexibility based on a set of 32 trinucleotide values giving the log-odds of the minor groove facing