APPLICATION NOTE

Visualization for genomics: the Microbial Genome Viewer

Robert Kerkhoven1,*, Frank H. J. van Enckevort1,2, Jos Boekhorst1, Douwe Molenaar2,3, Roland J. Siezen1,2,3

1Centre for Molecular and Biomolecular Informatics (CMBI), University of Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.
2NIZO Food Research, P.O. Box 20, 6710 BA Ede, The Netherlands.
3Wageningen Centre for Food Sciences (WCFS), P.O. Box 557, 6700 AN Wageningen, The Netherlands.
* To whom correspondence should be addressed.
(Running head: Microbial Genome Viewer)

ABSTRACT

Summary: A web-based visualization tool, the Microbial Genome Viewer, is presented which allows the user to combine complex genomic data in a highly interactive way. This web tool enables the interactive generation of chromosome wheels and linear genome maps from genome annotation data stored in a MySQL database. The generated images are in Scalable Vector Graphics (SVG) format, which is suitable for creating high-quality scalable images and dynamic web representations. Gene-related data such as transcriptome and time-course microarray experiments can be superimposed on the maps for visual inspection.

Availability: The Microbial Genome Viewer 1.0 is freely available at http://www.cmbi.kun.nl/MGV

Contact: R.Kerkhoven@cmbi.kun.nl

INTRODUCTION

The ongoing data explosion in the field of genomics, transcriptomics and metabolomics has created the need for data-handling tools which are able to reveal complex relations and patterns. In the field of genomics, much effort has been paid to structuring and standardization of data. The need for world-wide standards for data exchange is clearly represented by the numerous projects defining protocols for information storage in XML. Besides the growing need for structuring the data explosion, visualization of this data is an essential step in the flow of knowledge. As yet, not many free software tools are available for interactive visualization of circular or linear genome maps. Programs like GenomePlot (Gibson et al., 2003) and GenoMap (Sato et al, 2003) are used for creating chromosome wheels and linear maps of genomic data stored in tab-delimited or GenBank/EMBL format respectively. Both are stand-alone programs; the latter can also plot microarray expression or other types of quantitative data. Disadvantages of both programs are the lack of interactivity with the pictures and the requirement of specific input formats. A non-interactive structural DNA analysis and visualization is provided for all sequenced genomes by GenomeAtlas (Pedersen & Jensen et al., 2000) through a dedicated web-interface. Additionally, commercial software exists like GenoStar and DNASTAR, which include genome visualization and analysis.
IMPLEMENTATION

Here we present the Microbial Genome Viewer (MGV), a web-based tool for interactive visualization of annotation and transcriptome data on chromosome wheels and linear genome maps. Genome annotation data is retrieved from a local MySQL database enabling rapid visualization. Data includes all the relevant information on the complete genomes from the Genbank/EMBL/DDBJ databases. Several supplementary annotation methods have been automated in order to enrich this database. Terminator structures were added with TransTerm (Ermolaeva et al., 2000) and all organisms were scanned for COG (Tatusov et al., 2001) and Pfam domains (Bateman et al., 2002) using a Paracel GeneMatcher2 machine.

Scalable Vector Graphics (SVG) was chosen as the format for visualization. SVG is a language for describing two-dimensional images in XML and has been recommended by the W3C consortium (http://www.w3c.org/SVG). The possibility of adding animation, interaction and the supplementation of scripting languages makes SVG highly suitable for the visualization of complex data generated by the genomics community. Other advantages are the small file size and scalability for obtaining high-resolution images for posters and publications. SVG plugins for a wide range of platforms are available (http://www.adobe.com/SVG).

Genes can be colored according to different annotation methods. The classification of COG-domains in functional categories enables functional coloring of the genome maps (Tatusov et al., 2001). Numerical data like GC%, GC-skew, AT-skew and data from microarray experiments can be visualized with a color gradient. Microarray data is accepted in multiple formats. Additionally, manual selection and coloring can be made based on ORF numbers, keywords or customized functional categories. Gene annotation data and external databases like the NCBI are cross-linked with the genome maps by mouse-over functions.

Examples of pictures generated with the Microbial Genome Viewer are shown in Fig. 1; more examples can be found at http://www.cmbi.kun.nl/MGV/examples.

FUTURE

Subsequent versions of the Microbial Genome Viewer will enable the combinatorial visualization of the genome maps with metabolic pathways and gene regulatory networks. Different levels of visualization will ultimately be connected by transcriptome and metabolome data.

ACKNOWLEDGEMENTS

Supported by the Netherlands Ministry of Economic Affairs, IOP Programme, grant IGE01018, and the Netherlands Organisation of Scientific Research (NWO) BioMolecular Informatics Programme, grants 050.50.206 & 700.51.103 (Paracel GeneMatcher2).

REFERENCES


Fig. 1 A: Chromosome wheel from *Lactobacillus plantarum* WCFS1 (Kleerebezem et al., 2003) including COG-domain coloring of genes, GC% (red), GC-skew (blue) and AT-skew (green). B: Linear genome map of replication origin from *Escherichia coli* K-12. Genes are colored according to COG functional categories. C: Transcriptome experiment (glucose repression) from *Bacillus subtilis* 168 (Yoshida et al., 2001). Red: down-regulation; green: up-regulation; yellow: intermediate regulation; black: rRNA genes.