

Genome analysis

DNAVis: interactive visualization of comparative genome annotations

Mark W. E. J. Fiers¹, Huub van de Wetering², Tim H. J. M. Peeters², Jarke J. van Wijk² and Jan-Peter Nap^{1,3,*}

¹Applied Bioinformatics, Plant Research International, Wageningen University and Research Centre, Wageningen, The Netherlands, ²Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, Eindhoven, The Netherlands and ³Centre for BioSystems Genomics, Wageningen, The Netherlands

Received on October 9, 2005; revised on November 27, 2005; accepted on November 28, 2005

Advance Access publication December 6, 2005

Associate Editor: Chris Stoeckert

ABSTRACT

Summary: The software package DNAVis offers a fast, interactive and real-time visualization of DNA sequences and their comparative genome annotations. DNAVis implements advanced methods of information visualization such as linked views, perspective walls and semantic zooming, in addition to the display of heterologous data in dot plot-like matrix views.

Availability: The software is freely available at www.win.tue.nl/dnavis; the source code is available upon request.

Contact: janpeter.nap@wur.nl

Supplementary information: Figures are available at *Bioinformatics* online; documentation and data are available at www.win.tue.nl/dnavis

1 INTRODUCTION

The growing size of current sequence and annotation datasets requires appropriate and novel tools to explore and to retrieve biological insights from such data. Visual exploration that presents information interactively and in real-time is appealing to the intuition of a biologically skilled user and exploits the natural abilities to build mental maps of visually presented data. Previously defined features for the advanced visual exploration of comparative annotation include the ability to visualize and to move interactively between different scales (nucleotide to chromosome) as well as the ability to compare heterologous annotations (Peeters *et al.*, 2004). Appropriate visualization concepts to accomplish such features are semantic zooming (Perlin and Fox, 1993), perspective walls (Mackinlay *et al.*, 1991), focus + context (Spence, 2000) and linked views (Furnas and Bederson, 1995), all in combination with dot plot-like matrix views for comparisons of data types. Current genome viewers, such as the Generic Genome Browser (Stein *et al.*, 2002), the Artemis comparison tool (Carver *et al.*, 2005) or the Microbial Genome Viewer (Kerkhoven *et al.*, 2004) lack one or more of these advanced possibilities. We here present the software suite DNAVis that implements these advanced methods for the visualization of genome annotations.

2 SYSTEMS AND METHODS

Continuous, real-time interaction with data is a key issue to be able to view and to investigate multiple genomic regions simultaneously

with any annotation or user-defined measure of similarity. Equally essential is a high flexibility of scale: the ability to explore complete chromosomes with an arbitrary number of annotations from the lowest scale (individual nucleotide) to the whole genome. DNAVis implements several advanced methods for a fast, real-time interaction with genome sequences and annotations. This will help the exploration and comparison of annotated DNA sequences by biologically skilled users. Semantic zooming improves the display from the nucleotide scale to the whole genome level. Perspective walls provide context for the current area of interest and help to place any detailed analysis in the context of information from a substantially larger area. Linked views translate operations in one view to another view and facilitate co-inspection of areas. A dot plot-like view (Sonnhammer and Durbin, 1995), with all possibilities of the linear views in two dimensions, is chosen as method to display heterologous data simultaneously for easy comparisons of genomic regions and annotations.

3 IMPLEMENTATION

DNAVis is written in C++ and employs the OpenGL library for visualization. It runs on both Linux and Microsoft Windows operating systems. To comfortably use the software, notably with larger datasets, a computer with at least 512 Mb memory and modern accelerated 3D graphics hardware is recommended. All sequence data should be in the FASTA file format and precomputed annotation data require files in the General Feature Format (GFF) as defined by the Sanger Centre. Two figures that present the look and feel of DNAVis are available from the Bioinformatics Online website. Supplementary material that is available for download on the DNAVis website (www.win.tue.nl/dnavis) offers a comprehensive software manual and various precomputed datasets to experience and to experiment with the DNAVis software package.

DNAVis presents two types of views, a bar view and a matrix view. The two views can be displayed simultaneously. Figure 1 (Supplementary material) shows examples of the bar view and its various possibilities for adaptation of the views. Annotations have different types, e.g. a gene model, a microRNA, paralogs, expression level or protein interaction partners. It is expected that the number and type of annotations available will increase considerably over time. The visualization and layout in a bar view are configurable per annotation type. For example, the number

*To whom correspondence should be addressed.

and position of bars and their color can be set by the user. The bar view can be fluently panned by dragging. Zooming in or out is accomplished by dragging the view vertically. Perspective walls at both the right and the left of the bar show the neighboring areas (the context) of the currently displayed region (the focus). Multiple bar views can be linked and subsequently be panned and zoomed simultaneously. The bar views in Figure 1 (Supplementary material) illustrate how a part of chromosome 4 of *Arabidopsis thaliana* is depicted at different zoom levels with different types of annotations, including perspective walls and the combination with the annotation density of the whole chromosome in a histogram. This allows to inspect the same area at different zoom levels or to view orthologous areas of chromosomes from different organisms. An example could be a comparison of the *Caenorhabditis elegans* and *C. briggsae* genomes.

Comparative genomics information and annotation are presented in a dot plot-like matrix view. See Figure 2 (Supplementary material) for examples of the matrix view and its additional features. The matrix view consists of two axes of perpendicularly placed bar views, as described above, that define the dot plot-like matrix area. All options for a bar view apply to both axes in the matrix view, except for perspective walls. Annotations defining two areas as similar are displayed as a rectangle in the matrix area. The definition of what is similar is defined by the user. Fluent zooming and scrolling are possible. Zooming in to the lowest scale of the individual nucleotide yields classical dot plots for nucleotide comparisons. The matrix views in Figure 2 (Supplementary material) illustrate how DNAVis allows easy analysis of heterologous annotation data by combining sequence similarity according to Blast (Altschul *et al.*, 1990) or Vmatch (see Kurtz *et al.*, 2001, www.vmatch.de) and data on gene expression. With proper scale settings, gene duplications between chromosomes (Bowers *et al.*, 2003) or local co-expression domains (in this case, quadruplets; Ren *et al.*, 2005) are easily discernible in the genome of *A. thaliana*. DNAVis demonstrates that an integrated display of multiple heterologous datasets is feasible and attractive in such a matrix.

4 DISCUSSION AND CONCLUSIONS

DNAVis uses advanced methods for visualization that depend on modern PC graphics hardware. The advanced technology results in smooth and real-time interaction with datasets as large as complete chromosomes or genomes, with large numbers of different types of annotations. The addition of a third dimension in the visualization would result in more possibilities to display data. This could be of use with yet larger or higher dimensional datasets. However, 3D visualizations inherently introduce occlusion and consequently require (interactive) definitions of appropriate view points. The associated problems with interpretation for a user add to the complexity of using 3D in genome visualization. In view of the way DNAVis manages to visualize large datasets without the drawbacks of 3D, it has been a deliberate choice not to implement any 3D visualization. Future improvements, depending on the availability of funding and manpower, will implement interactive database

connections, add mouse-over information on similarity matches, the possibility to edit annotations in a view, more extensive search facilities and other suggestions made by users.

The future of the visualization of genome information is likely to focus on the display of comparative heterologous data for many more genomes. The ability to interactively and in real-time compute and visualize complex relationships between multiple annotations will be a major challenge for the future of genome visualization. This is likely to lead to new insights in genome structure and organization. The careful consideration and use of information visualization technology has resulted in an efficient and effective approach for modern genome exploration. In retrospect, it was surprising for the team involved how little of the standard approaches in advanced visualization science had yet found its place in genome visualization. This opens up future promises for the comparative display and exploration of genomes and their annotations with the help of visualization science.

ACKNOWLEDGEMENTS

We thank the Centre for Biosystems Genomics (CBSG, part of the Dutch Genomics Initiative) for financial support (J.P.N.), Roeland van Ham and other members of the Applied Bioinformatics group of Plant Research International for input and discussions and an anonymous reviewer for suggesting further improvements in the DNAVis software and this paper.

Conflict of Interest: none declared.

REFERENCES

- Altschul,S. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Bowers,J. *et al.* (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433–438.
- Carver,T. *et al.* (2005) ACT: the Artemis comparison tool. *Bioinformatics*, **21**, 3422–3423.
- Furnas,G.W. and Bederson,B.B. (1995) Space-scale diagrams: understanding multi-scale interfaces. In *Proceedings of CHI'95*, Denver, CO, ACM Press, NY, pp. 234–241.
- Kerkhoven,R. *et al.* (2004) Visualization for genomics: the Microbial Genome Viewer. *Bioinformatics*, **20**, 1812–1814.
- Kurtz,S. *et al.* (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.*, **29**, 4633–4642.
- Mackinlay,J.D. *et al.* (1991) The perspective wall: detail and context smoothly integrated. In *Proceedings of CHI'91*, New Orleans, LA, ACM Press, NY, pp. 173–179.
- Peeters,T. *et al.* (2004) Case study: visualization of annotated DNA sequences. In Deussen,O., Hansen,C.D., Keim,D.A. and Saupé,D. (eds), *Vissym*. Eurographics Association, pp. 109–114.
- Perlin,K. and Fox,D. (1993) Pad: an alternative approach to the computer interface. *Computer Graphics*, **27**, 57–72.
- Ren,X. *et al.* (2005) Local coexpression domains of two to four genes in the genome of *Arabidopsis*. *Plant Physiol.*, **138**, 923–934.
- Sonnhammer,E. and Durbin,R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–10.
- Spence,R. (2000) *Information Visualization*. Addison-Wesley, Harlow, UK.
- Stein,L. *et al.* (2002) The Generic Genome Browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.