BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 6 2008, pages 861–862
doi:10.1093/bioinformatics/btm598

Genome analysis

GenomeVx: simple web-based creation of editable circular chromosome maps

Gavin C. Conant* and Kenneth H. Wolfe

Smurfit Institute of Genetics, Trinity College, University of Dublin, Dublin 2, Ireland

Received on October 17, 2007; revised on November 21, 2007; accepted on November 28, 2007

Advance Access publication January 28, 2008

Associate Editor: Martin Bishop

ABSTRACT

We describe GenomeVx, a web-based tool for making editable, publication-quality, maps of mitochondrial and chloroplast genomes and of large plasmids. These maps show the location of genes and chromosomal features as well as a position scale. The program takes as input either raw feature positions or GenBank records. In the latter case, features are automatically extracted and colored, an example of which is given. Output is in the Adobe Portable Document Format (PDF) and can be edited by programs such as Adobe Illustrator.

Availability: GenomeVx is available at http://wolfe.gen.tcd.ie/GenomeVx

Contact: conantg@tcd.ie

Since the invention of Sanger sequencing (Sanger *et al*., 1977), the cost of DNA sequencing has fallen continually, and the volume of sequence data in public databases has correspondingly risen (Collins *et al*., 2003). Thus, we estimate that the number of sequenced organellar genomes from eukaryotes has doubled roughly every two years since the first such sequences were published (data not shown; Ohyama *et al*., 1986; Shinozaki *et al*., 1986). Indeed, the number of sequenced angiosperm chloroplast genomes in GenBank tripled from 24 at the end of 2005 to at least 73 at the end of 2007, and new sequencing technologies are already pushing this number upward (Moore *et al*., 2006).

One of us (KHW) sequenced one chloroplast genome in 1992 and another in 2007. We were surprised to find that, although the annotation toolkits for organelle genomes had improved over this period (Wyman *et al*., 2004), there were still no easy-to-use, free web tools for making publication-quality genomic maps. Most papers reporting sequences of chloroplast or mitochondrial genomes include a circular map showing gene locations (Cai *et al*., 2006; Guo *et al*., 2007; Saski *et al*., 2005; Talla *et al*., 2005). However, by making enquiries, we discovered that these maps are still often laboriously hand-drawn, although there are at least three packages which can generate them (Gibson and Smith, 2003; Sato and Ehira, 2003; Stothard and Wishart, 2005).

With all bioinformatics software there are trade-offs between ease-of-use and the number of features. The three tools mentioned above are very feature-rich. However, as a result, all must be installed locally by the user, and all also have some combination of local library dependence, complex input formatting and output formats that are difficult to edit. We set out to create a simple Web-based tool (GenomeVx) for making circular maps. This program would not attempt to duplicate the features available with existing tools; rather it would aim to be as simple-to-use as possible, with a web-based interface and an output format which would be accessible on almost all hardware platforms.

Our initial Postscript maps are generated by a C++ program linked to the GNU plotutils package (http://www.gnu.org/software/plotutils/). This program is wrapped into a CGI front-end (http://wolfe.gen.tcd.ie/GenomeVx) which allows users to input data either as a list of features and coordinates or as a preexisting GenBank flatfile. Because the Postscript format is not universally supported, GenomeVx's output is an Adobe Portable Document Format file. Although the output from GenomeVx can be used directly for presentation (Fig. 1), we expect that most users will edit the resulting map with a program such as Adobe Illustrator.

Figure 1 illustrates the unedited output from GenomeVx, using the chloroplast genome sequence of the American sycamore *Platanus occidentalis* (Moore *et al*., 2006) as input. The image shown was created directly from the features in the GenBank record for this genome (NC_008335). GenomeVx does not attempt to correct for the occasional overlaps in gene names shown, but the labels can be moved manually by editing the PDF file. The colors in the Figure were produced automatically: genes are colored based on the first two letters of the gene name (except for ribosomal RNAs which are all given a single color).

Input into GenomeVx can either be done manually by pasting genes or nucleotide features (e.g. SNPs) from a program such as Microsoft Excel or by uploading a GenBank-format flatfile. When imported from a GenBank file, coordinates appear as an editable list in the web interface, giving the user the option of overriding the automatic coloring decisions or editing the annotation. The program can also include miscellaneous features on one or more inner scale rings (e.g. the inverted repeats in Fig. 1).

GenomeVx is intended to simplify the study of organelle genomes by allowing biologists to produce circular genome maps without resorting to ad-hoc solutions. The web-based

*To whom correspondence should be addressed.

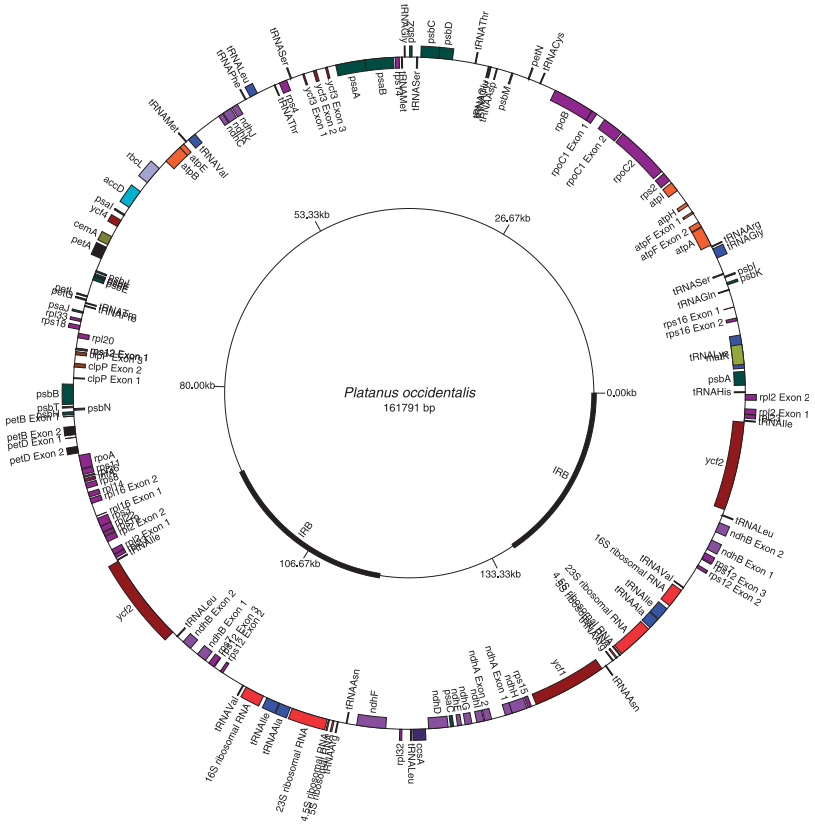

Fig. 1. A chloroplast genome map produced by GenomeVx. Genes are colored in groups based on the first two letters of the gene name. A user-specified number of evenly spaced scale indices are produced after rounding the genome size to the nearest two digits (thus, the distance between the last label and 0 kb may be greater than between the other markers). GenomeVx can orient the map either anticlockwise from 3 O'clock (the convention for chloroplast genomes) or clockwise from 12 O'clock (the convention for mitochondrial genomes).

interface allows quick access without the installation of local software. We hope that GenomeVx will enable researchers who study circular genomes to spend less time drawing and more time on topics of greater scientific interest.

ACKNOWLEDGEMENTS

We thank K. Diekmann, A.C. Frank and J. Mower for suggestions during the design and testing of GenomeVx. We also thank K. Byrne and K. Hokamp for technical assistance developing the web interface. This work was supported by Science Foundation Ireland.

Conflict of Interest: none declared.

REFERENCES

Cai,Z. *et al.* (2006) Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.*, **6**, 77.

Collins,F.S. *et al.* (2003) The human genome project: lessons from large-scale biology. *Science*, **300**, 286–290.

Gibson,R. and Smith,D.R. (2003) Genome visualization made fast and simple. *Bioinformatics*, **19**, 1449–1450.

Guo,X. *et al.* (2007) Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts. *BMC Genomics*, **8**, 228.

Moore,M.J. *et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.*, **6**, 17.

Ohyama,K. *et al.* (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, **322**, 572–574.

Sanger,F. *et al.* (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, **265**, 687–695.

Saski,C. *et al.* (2005) Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.*, **59**, 309–322.

Sato,N. and Ehira,S. (2003) GenoMap, a circular genome data viewer. *Bioinformatics*, **19**, 1583–1584.

Shinozaki,K. *et al.* (1986) The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.*, **5**, 2043–2049.

Stothard,P. and Wishart,D.S. (2005) Circular genome visualization and exploration using CGView. *Bioinformatics*, **21**, 537–539.

Talla,E. *et al.* (2005) The complete mitochondrial genome of the yeast *Kluyveromyces thermotolerans*. *FEBS Lett.*, **579**, 30–40.

Wyman,S.K. *et al.* (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.