ChromoWheel: a new spin on eukaryotic chromosome visualization

*Sven Ekdahl and Erik L. L. Sonnhammer**

Center for Genomics and Bioinformatics, Karolinska Institutet, S-17177 Stockholm, Sweden

ABSTRACT

Summary: ChromoWheel is an Internet browser application for generating whole-genome illustrations. It can be used to depict chromosomes, genes and relations between chromosomal loci. The circular layout of chromosomes is advantageous for showing relationships between different chromosomes, as the connecting line never crosses over a chromosome. All graphical image components are in the vector-based format Scalable Vector Graphics, which are highly scaleable and admit user interaction. ChromoWheel can either be run with user-provided data in the generic SFS format, or as a browser front-end for precompiled genomic data.

Availability: http://chromowheel.cgb.ki.se

Contact: Erik.Sonnhammer@cgb.ki.se

In the era of whole-genome analysis, it is often desired to get a global overview of features on the chromosomes in a genome. For instance, the genomic distribution of genes belonging to a family or a pathway may have striking appearance when plotted graphically. One may also wish to display and analyse how genes or other parts of the genome relate to each other, for instance by connecting similar repeats or homologous genes with lines.

Traditionally, chromosomes have been drawn as an array of straight bars. This is employed by several existing genome databases (e.g. Karolchik *et al.*, 2003). While the straight bar representation is popular for showing distributions or populations of objects on a chromosome (Harris *et al.*, 2003), it is not well-suited for showing inter-chromosome relationships. Connecting points between parallel straight bars involves crossing over intermediate chromosomes, resulting in a confusing and cluttered picture. This could be avoided by using a new strategy when composing the picture. By placing chromosomes on the edge of an imaginary circle, no lines representing relations between objects has to cross over a chromosome or hide a gene locus. With this kind of illustration, relations rather than chromosomes, will be focused on. The circular diagram has previously been used to display prokaryotic

genomes (Ogata *et al.*, 1999) and in some cases for eukaryotic chromosomes and for showing relationships between objects (Blanc *et al.*, 2003).

ChromoWheel is an illustration-generating process that applies a circular composing strategy. Unlike image generators for circular genomes, ChromoWheel allows any number of separate chromosomes to be processed into one resulting image. It is implemented for displaying relations between different genes and chromosomal loci. ChromoWheel generates pictures in the vector-based image format of Scalable Vector Graphics (SVG, http://www.w3.org/Graphics/SVG), which is a recommended Web standard, supported by W3C (http://www.w3.org). The processing is based on XML and XSLT and is implemented entirely on the server side using Xalan, an XSLT processor provided by the Apache project (http://www.apache.org). A number of Web browser plug-ins exists for displaying SVG files; Adobe's SVGViewer is probably the most widely used. This plug-in supports most platforms and is available for download, free of charge (http://www.adobe.com/svg). Due to changes in the plug-in API, however, Adobe's SVGViewer 3.0 is only compatible with version 6.1 of Netscape or higher. We expect SVG to be supported more generally in the near future by all popular browsers. The exact behaviour of, e.g. zooming is dependent on the browser plug-in.

Chromosomes are represented as thick coloured arcs along the circle edge. Across these arcs, lines are drawn at right angles, to represent chromosomal loci, typically genes. Between chromosomal loci, relations can be defined which are displayed by an arc, connecting one locus to the other. At present such arcs are simply black, but in the future they could be made colour-coded to highlight which chromosome or gene it belongs to, or to represent various qualities such as similarity score. When submitting data to ChromoWheel, it is sorted and organized before the image is created. The generated picture admits user interaction to some extent. At the top of the picture, there is an information box. Within this box, information about chromosomes and genes are displayed when pointed at with the mouse cursor. The SVG image format supports zooming and panning functions.

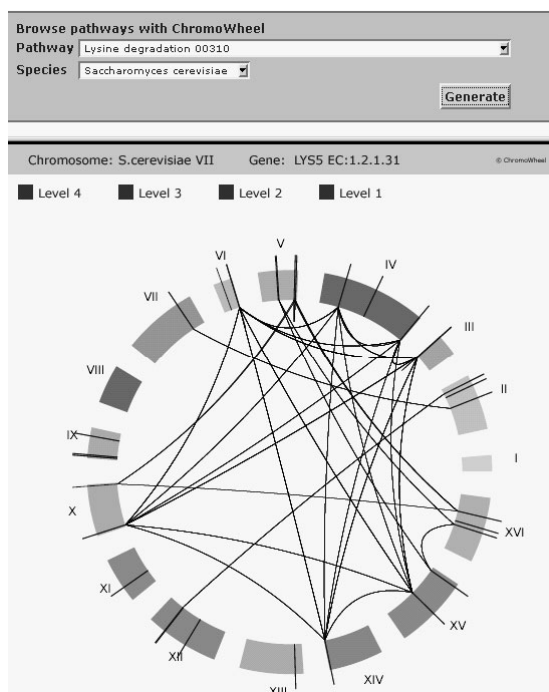*To whom correspondence should be addressed.

Fig. 1. Example of a ChromoWheel picture, showing Lysine degradation genes in *Saccharomyces cerevisiae*. In this example, genes are connected if they share the same E.C. number. The status bar shows the chromosome and gene that the mouse currently points to. Display of E.C. number connections can be restricted to different levels of granularity with the 'Level' buttons. For instance, connections between genes with only three E.C. levels known are turned on or off with the 'Level 3' button. Here all levels are shown.

These functions are accessed by right-clicking the image (Fig. 1).

ChromoWheel processing application is available as a Web service at Karolinska Institutet (http://chromowheel.cgb.ki.se). At this site, a ChromoWheel picture can be generated by submitting a file containing the data on which the picture should be based. The generated illustration can then be saved to a local computer by right-clicking the image. The format supported by ChromoWheel is SFS 1.0. This is a simple generic data definition format (Sonnhammer and Wootton, 2001) for sequence feature data. Only the SEG (segment) and HSP (pair) data types are currently recognized by ChromoWheel. To specify the full-length of the chromosomes, a magic tag 'Chromosome' is used as seqname in a SEG data element. An example on how to design SFS data used to generate ChromoWheel pictures is available at the Web site.

ChromoWheel is also well-suited to use as an user interface for a genomic data collection, e.g. a database of research results, since SVG is easily embedded in Web pages. The data requested for display is then automatically processed and an SVG image is displayed to the user. An example can be found under 'Pathway Browsing' on the ChromoWheel Web page, for genes in the KEGG metabolic pathways (Kanehisa *et al.*, 2002) that were mapped to genomic locations in five eukaryotes (Lee and Sonnhammer, 2003).

REFERENCES

Blanc,G., Hokamp,K. and Wolfe,K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.*, **13**, 137–144.

Harris,T.W., Lee,R., Schwarz,E., Bradnam,K., Lawson,D, Chen,W., Blasier,D., Kenny,E., Cunningham,F., Kishore,R. *et al.* (2003) WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res.*, **31**, 133–137.

Kanehisa,M., Goto,S., Kawashima,S. and Nakaya,A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.*, **30**, 42–46.

Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W., Thomas,D.J. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.

Lee,J.M. and Sonnhammer,E.L.L. (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res.*, **13**, 875–882.

Ogata,H., Goto,S., Sato,K., Fujibuchi,W., Bono,H. and Kanehisa,M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **27**, 29–34.

Sonnhammer,E.L.L and Wootton,J.C. (2001) Integrated graphical analysis of protein sequence features predicted from sequence composition. *Proteins*, **45**, 262–273.