

Visualizing syntenic relationships among the hemiascomycetes with the Yeast Gene Order Browser

Kevin P. Byrne* and Kenneth H. Wolfe

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

Received August 10, 2005; Revised and Accepted October 3, 2005

ABSTRACT

The Yeast Gene Order Browser (YGOB) is an online tool designed to facilitate the comparative genomic visualization and appraisal of synteny within and between the genomes of seven hemiascomycete yeast species. Three of these genomes are polyploid, and hence contain intra-genomic syntenic regions, the correct assembly of which is a particular success of YGOB. Designed to accurately assemble, display and score gene order relationships, YGOB is both an interactive tool for browsing genomic data, and a software engine now being used for evolutionary analyses on a whole-genome scale. Underlying the online interface is the YGOB database, which consists of homology assignments across the species, extensively curated based on sequence similarity and novelty, an appraisal of genomic context (synteny) in multiple genomes. Currently the YGOB database incorporates genome data from *Saccharomyces cerevisiae*, *Candida glabrata*, *Saccharomyces castellii*, *Ashbya gossypii*, *Kluyveromyces lactis*, *Kluyveromyces waltii* and *Saccharomyces kluyveri*, but the system is scaleable to accommodate additional genomes. This paper discusses the usage and utility of version 1.0 of YGOB, which is publicly available at <http://wolfe.gen.tcd.ie/ygob>.

INTRODUCTION

The Yeast Gene Order Browser (YGOB) has been developed to take advantage of the potential the hemiascomycete yeasts offer for studying many aspects of genome evolution, including in particular the whole-genome duplication (WGD) event we proposed in 1997 (1) as having occurred in an ancestor of *Saccharomyces cerevisiae*, and which was conclusively confirmed in 2004 (2–4).

YGOB version 1.0 features seven genomes: three from lineages that separated after the WGD [‘post-WGD’ species: *S.cerevisiae* (5), *Saccharomyces castellii* (6) and *Candida glabrata* (4)], and four from outgroups [‘pre-WGD’ species: *Ashbya gossypii* (3), *Kluyveromyces lactis* (4), *Kluyveromyces waltii* (2) and *Saccharomyces kluyveri* (6)]. The phylogenetic relationship among these species and the position of the WGD event are shown in Figure 1A. YGOB consists of a curated database of homology assignments across these genomes, an intuitive and interactive online interface and software that accurately assesses and displays gene order (synteny).

YGOB is publicly available online (wolfe.gen.tcd.ie/ygob) and is being used by our laboratory for a number of research projects. We believe it will be of similar utility to yeast researchers in general, and of interest to evolutionary biologists as a pioneering comparative genomics tool.

The algorithms involved in accurately assigning synteny and aligning chromosomal fragments, information on the curation of the genomic data and research carried out using YGOB are presented in detail elsewhere (7).

VISUAL USER INTERFACE

Figure 1B shows a YGOB screenshot with a window size of 6. The tracks are labeled at right and show from top to bottom: the A tracks of the three post-WGD species, the four single-track pre-WGD species and the B tracks of the post-WGD species. Genes are represented by boxes stating the gene name, chromosome or scaffold name, and species name (except for *S.cerevisiae*, where systematic and genetic names are both shown instead). Each genome has a color palette that is used to distinguish genes from different chromosomes or contigs; for example, the three differently colored *S.kluyveri* contigs in Figure 1B. Genes without synteny are colored gray, for example the *S.kluyveri* gene 950.1 in the column marked ‘b’ in Figure 1B. The in-focus gene, which was used to compose the entire display, is highlighted by an orange border.

Connectors join nearby genes: solid connectors join adjacent genes, two small bars connect genes <5 genes apart, and

*To whom correspondence should be addressed. Tel: +353 1 608 1288; Fax: +353 1 679 8558; Email: kevin.byrne@tcd.ie

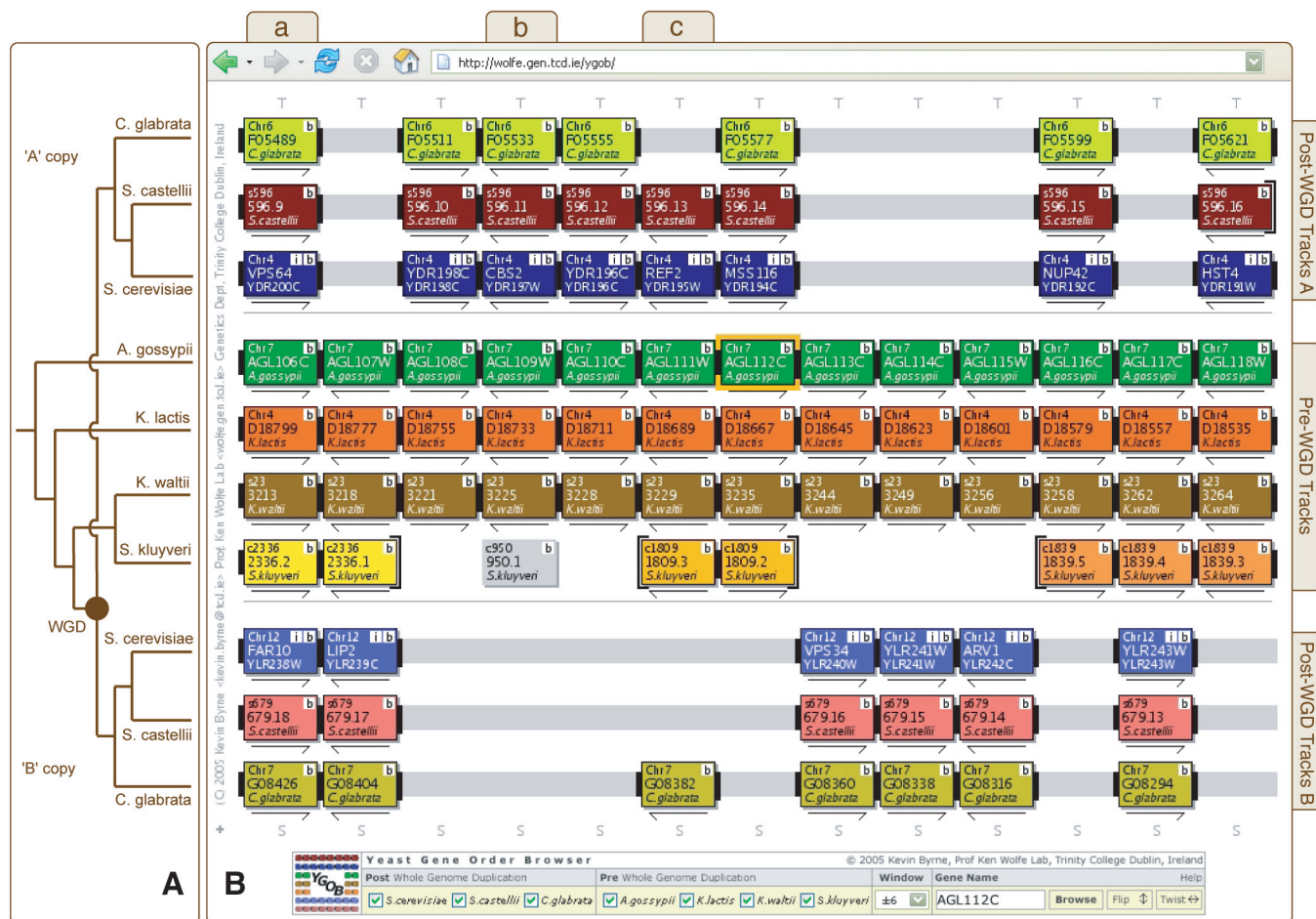


Figure 1. (A) Approximate phylogenetic relationship of the yeasts included in YGOB, with the WGD event marked as a closed circle. The tree is based on Kurtzman *et al.* (10). (B) YGOB screenshot focused on the *A. gossypii* gene AGL112C with a window size of 6. Tracks are labeled at right. At the bottom of the interface is the control console, which allows users to select the window size and the gene to focus on. Each box in the display represents a gene, and each color, a chromosome. The 'b' buttons open a window with BLASTP results against YGOB's database, 'S' buttons click through to protein sequences, 'T' buttons draw approximate phylogenetic trees and '+' buttons output YGOB data in a tabulated text format.

one small bar connects genes <20 genes apart. These connectors are usually colored black, but are highlighted in orange if they denote an inversion. When there is an intervening space between genes, for example in the post-WGD tracks in Figure 1B, then the connectors are extended in gray over that space. The end of a chromosome or contig is denoted by a brace (e.g. *S. kluyveri* contigs in Figure 1B). The arrows under a gene box denote its relative transcriptional orientation. Whether a rightwards arrow corresponds to a Watson or a Crick strand gene is arbitrary and can be changed by using the 'twist' option, which reverses the left-right sequence of the columns. For this reason, no orientation arrow is shown below genes without synteny (i.e. those colored gray).

The online interface is a Perl (www.perl.org) application with the visual output created using the GD package (www.boutell.com/gd). Some advanced aspects of the control console (discussed below), but none of the features in the main visual interface, require users to have JavaScript enabled in their web browser. The YGOB source code is available on request (kevin.byrne@tcd.ie).

GENOME BROWSING

The user can click on any other gene in the display (Figure 1B) to refocus the browser on it. This allows users to walk along chromosomes by repeatedly refocusing on genes at the edge of the display; these effectively serve as 'step right' and 'step left' buttons. Browsing a genomic region when focused on a pre-WGD genome often gives the best view of genes' syntenic context. This is because of the way the pre-WGD genomes act as a 'scaffold' for the sister genomic regions from post-WGD genomes (2,3,7). YGOB's display of homology information in the syntenic context of many genomes makes features such as annotation problems, fast evolving loci, species-specific genes, gene clusters, chromosomal rearrangements, ohnologs (paralogs arising from a WGD, e.g. column 'a' in Figure 1B), or cases of differential gene loss (e.g. column 'c' in Figure 1B) immediately apparent. A number of these features, in particular the patterns of gene loss, have been examined by us elsewhere (7).

The control console, which is at the bottom of YGOB's web interface (Figure 1B), allows a user to select several parameters: which genomes to display; the window size

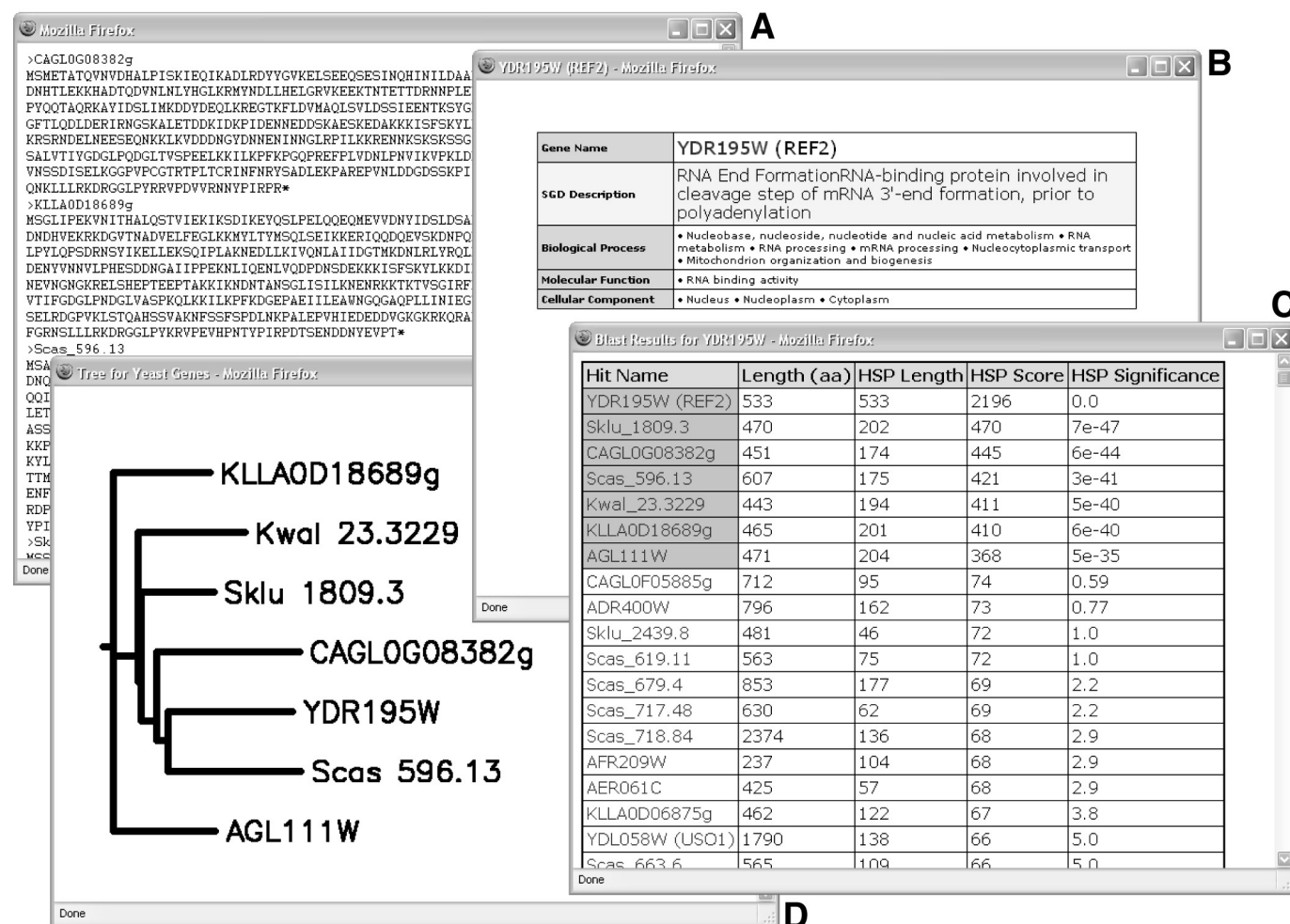


Figure 2. YGOB bioinformatics utilities screenshots for the homology pillar containing the *S.cerevisiae* gene *REF2*, and labeled 'c' in Figure 1B. Clockwise from top left: (A) protein sequences for the pillar; (B) SGD and Gene Ontology information for *REF2*; (C) BLASTP results for Ref2 against the YGOB database; and (D) approximate phylogenetic tree for genes in the pillar.

(between 4 and 50 columns can be shown on each side of the in-focus gene's column); and which gene to *focus* the display on. A new in-focus gene can be chosen by typing its name into the text box, as an alternative to clicking on one of the genes currently on screen. Gene naming conventions are explained in the online help, and synonyms are supported. The control console includes buttons that allow a user to invert the whole display output ('twist') or to swap the A and B tracks for the post-WGD species ('flip').

BIOINFORMATICS UTILITIES

Every homology pillar and gene box in the YGOB visual interface (Figure 1B) features buttons offering ready access to standard bioinformatics tools (Figure 2). The 'S' (Sequence) button below a column clicks through to its protein sequences (Figure 2A). *S.cerevisiae* gene boxes have an 'i' (Information) button that clicks through to the *Saccharomyces* Genome Database (SGD) description (8) of the protein and its Gene Ontology terms (Figure 2B). Each gene box has a 'b' (BLAST) button, which opens a window with BLASTP results

for that gene's protein against all the proteins in YGOB's database. In the pop-up window of BLASTP results (Figure 2C), gene names are highlighted in red (seen as dark gray highlighting in Figure 2C) if they are in the same column as the query gene; blue if they appear on screen but not in its column, and orange if they are a tandem copy of the query gene. Normally colored white, the 'b' button is orange when the gene has a tandem repeat on screen nearby. Finally, where there are ≥ 3 genes in a column, the 'T' (Tree) button draws an approximate phylogenetic tree (Figure 2D), generated on the fly using a T-Coffee alignment (9) and Phylip's Neighbor-Joining program, rooted with pre-WGD species if possible. Rolling the computer mouse over a gene's box will also display brief information for that gene, for example from SGD title lines in the case of *S.cerevisiae*, and GenBank annotation tags for other genomes.

DATA RETRIEVAL

Every visualized YGOB page (Figure 1B) has a '+' button in its bottom left-hand corner. Clicking the button displays the

same YGOB data in a tabulated text format, showing each track as a column and each pillar as a row. This allows a user to easily retrieve the raw syntenic context information around a gene or region of interest. The YGOB website also hosts interactive lists of loci classed by syntenic locus type, allowing researchers to quickly acquire lists of genes/loci from particular classes or alternatively to browse through these lists online (e.g. a list of all loci that are single copy in *S.cerevisiae* but retained in duplicate in *C.glabrata* and *S.castellii* can be retrieved). In addition to this the website features lists of genes from locus classes further subdivided by Gene Ontology (www.geneontology.org) terms whereas interactive diagrams show the physical distribution of all locus classes over the three pre-WGD species.

CONCLUSION

The YGOB database consists of a set of extensively curated homology assignments across the seven yeast genomes currently included. The use of syntenic information in assigning homology means these pillars are virtually free from the demonstrated limitations of BLAST. The YGOB online interface (wolfe.gen.tcd.ie/ygob) provides browsing access to this database via an interface that is intuitive to use, with ready access to various bioinformatics utilities. YGOB has been primarily designed to facilitate the use of gene order in the study of polyploid genomes and post-polyploid evolution, but by showing genes and homology in an accurate syntenic context, YGOB makes the characteristics of genomic regions or loci of interest immediately apparent. The success of this contextual syntenic approach in polyploid yeasts encourages us to extend it to other groups of species for which there is sufficient genomic data.

ACKNOWLEDGEMENTS

We thank D. R. Scannell, J. L. Gordon and S. Wong for assistance with the development of YGOB. This study was

supported by Science Foundation Ireland. Funding to pay the Open Access publication charges for this article was provided by Science Foundation Ireland.

Conflict of interest statement. None declared.

REFERENCES

1. Wolfe, K.H. and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, **387**, 708–713.
2. Kellis, M., Birren, B.W. and Lander, E.S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, **428**, 617–624.
3. Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S. *et al.* (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science*, **304**, 304–307.
4. Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E. *et al.* (2004) Genome evolution in yeasts. *Nature*, **430**, 35–44.
5. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M. *et al.* (1996) Life with 6000 genes. *Science*, **274**, 546, 563–547.
6. Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A. and Johnston, M. (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science*, **301**, 71–76.
7. Byrne, K.P. and Wolfe, K.H. (2005) The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.*, **15**, 1456–1461.
8. Christie, K.R., Weng, S., Balakrishnan, R., Costanzo, M.C., Dolinski, K., Dwight, S.S., Engel, S.R., Feierbach, B., Fisk, D.G., Hirschman, J.E. *et al.* (2004) *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.*, **32**, D311–D314.
9. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.
10. Kurtzman, C.P. and Robnett, C.J. (2003) Phylogenetic relationships among yeasts of the ‘*Saccharomyces* complex’ determined from multigene sequence analyses. *FEMS Yeast Res.*, **3**, 417–432.