Aequatus: Visualising complex similarity relationships among species

Anil S. Thanki[1], Sarah Ayling[1], Javier Herrero[1, 2], Robert P. Davey[1]

[1] The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK

[2] Research Dept. of Cancer Biology, UCL Cancer Institute, 72 Huntley Street, London WC1E 6DD

Project Website: http://browser.tgac.ac.uk/aequatus
Source Code: https://github.com/tgac/aequatus-browser
User guide: http://browser.tgac.ac.uk/aequatus-user-guide/
License: GPL V3
Contact: anil.thanki@tgac.ac.uk

The phylogenetic information inferred from the study of homologous genes helps us to understand the evolution of gene families, which plays a vital role in finding ancestral gene duplication events as well as identifying regions those are under positive selection within species (1). Conservation of homologous loci results in syntenic blocks, and there are various tools available to visualise syntenic information between species, such as Ensembl Browser (2), Genomicus (3), SyMap (4), and MizBee (5). These tools are able to provide an overview of syntenic regions as a whole, reaching down to the gene level, but none provide any information about structural changes within genes such as the conservation of ancestral exon boundaries amongst multiple genomes. To this end, we present the Aequatus Browser, a web-based tool with novel rendering approaches to visualise homologous, gene structures among differing species or subtypes of a common species.

The Aequatus Browser utilises common open source web technologies to provide a fast and intuitive browsing experience over complex data, processing and visualising comparative genomics information directly from the Ensembl Compara and Ensembl Core schema databases. Precalculated genomic alignments, in the form of CIGAR strings, are held in Ensembl Compara and Aequatus cross-references these sequences to Ensembl Core databases for each species to gather genomic feature information. Aequatus then processes the comparative and feature data to provide a visual representation of the phylogenetic and structural relationships among the set of chosen species. Whilst applicable to species with high-quality gold-standard reference genomes such as human or mouse, the Aequatus Browser was designed with large fragmented genome references in mind, particularly hard-to-assemble polyploid plants. The ultimate goal of the Aequatus Browser is to provide a unique and informative way to render and explore complex relationships between genes from various species at a level that has so far been unrealised.

The latest version of the Aequatus Browser supports the Ensembl Compara schema v78 and later, as well as refactored code for faster data retrieval, improved visualisation algorithms, and a simplified and informative user interface. It also includes a new REST API for consistent access to genes of interest, making it easy to share information with collaborators via persistent URLs.

References

1. http://www.ncbi.nlm.nih.gov/pubmed/19029536
2. http://www.ensembl.org/
3. http://www.genomicus.biologie.ens.fr/genomicus-75.02/
4. http://www.agcol.arizona.edu/software/symap/
5. http://www.cs.utah.edu/~miriah/mizbee/Overview.html