BIOINFORMATICS APPLICATIONS NOTE Vol. 18 no. 10 2002
Pages 1398–1399

Genquire: genome annotation browser/editor

M. D. Wilkinson, D. Block† and W. L. Crosby*

Plant Biotechnology Institute, National Research Council Canada, Saskatoon, Saskatchwan, S7N 0W9, Canada

Received on February 18, 2002; revised on March 27, 2002

ABSTRACT

Summary: We present a software package, Genquire, that allows visualization, querying, hand editing, and de novo markup of complete or partially annotated genomes. The system is written in Perl/Tk and uses, where possible, existing BioPerl data models and methods for representation and manipulation of the sequence and annotation objects. An adaptor API is provided to allow Genquire to display a wide range of databases and flat files, and a plugins API provides an interface to other sequence analysis software.

Availability: Genquire v3.03 is open-source software. The code is available for download and/or contribution at http://www.bioinformatics.org/Genquire.

Contact: mwilkinson@gene.pbi.nrc.ca

Genome databases are most commonly visualized through CGI-generated web pages and 'clickable' images. These are generally slow, 'flat', and static in their user-interactions. Moreover, such interfaces are almost invariably read-only, and lack the flexibility that would encourage intensive browsing and analysis of the data by bench scientists.

Genquire is an open-source, platform-independent, stand-alone genome browser and hand annotation tool created with the goal of increasing the accessibility of genome annotation data to both annotators and bench scientists. Coded entirely in Perl/Tk, it uses, where possible, existing BioPerl data models and methods (BioPerl Project) for representation and manipulation of biological objects. A simple XML-based plugins interface allows results of external data analysis and manipulation tools to be displayed 'on the fly' in the context of existing information, and optionally imported into the underlying database. The Genquire system includes its own database schema, but is able to interface with a wide range of data sources, both databases and flat files, through its data-adaptor code layer. A well-coded adaptor to any read/write data source would allow most or all of Genquire's features to be functional. Both the adaptor layer and plugins API are continually under development to allow new features to be added to Genquire.

Multiple genomes from various data sources may be browsed simultaneously, with each genome being presented at three levels of resolution—Whole genome, Contig, and Nucleotide. Screen shots of these three visualization components are available at http://www.bioinformatics.org/Genquire/screenshots.html.
The interface at each visualization level provides distinct exploratory possibilities:

Whole Genome: At the whole genome level, chromosomes or chromosome fragments (for incomplete genomes) are displayed as segmented, proportionately sized vertical bars, with each segment representing a 'contig', or database equivalent. A query interface is provided that highlights the genomic locations of search results, allowing rapid global assessment of genomic content and feature distribution. Such searches may be based on annotation keyword, sequence homology, or other methodologies via specific plugins. Search results can be saved as genome-level annotations and retrieved in later sessions. At any time, one or more contigs may be selected for visualization at higher resolutions.

Contig: The 'Contig' visualization presents a familiar zoomable, scrollable display with sequence features represented as bars offset from a horizontal or vertical axis representing the underlying nucleotide sequence. It is generated using the BioTkPerl widget set (Helt, 1996) and is modeled after the Genotator browser (Harris, 1997). The contig display is split into two halves, with one half displaying all available sequence features, and the other displaying only genes and their transcripts. Query results from the Whole Genome display are passed on to this level of visualization, with 'hit' features being highlighted, and homology results mapped as new feature bars. This display allows de novo gene and transcript creation from new or existing features via an intuitive drag-and-drop interface. Features may also be annotated en masse by a simple tag/value system, or individually more deeply using Genquire's browsable interface to the Gene Ontology database (Gene Ontology Consortium,

*To whom correspondence should be addressed.
† Present address: Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA

2000). Selected features may be viewed at a higher resolution (Nucleotide display), or sent to other analysis tools with Genquire querying and displaying the results.

Nucleotide: Sequence regions selected in the Contig view may be displayed as raw DNA sequence. Any selected features on the Contig view that fall into this nucleotide region are highlighted in the same color as their Contig-level feature bar to maintain contextual information. Coding features are translated, with the amino acid sequence displayed in an adjacent panel. Feature boundaries are editable at the single nucleotide level, and common sequence features such as donor/acceptor splice sites, start, and stop codons are indicated. Features may be created and annotated de novo by mouse selection of the sequence of interest. All changes and additions are updated both in the translation panel, and the Contig display.

We believe Genquire has four major advantages over other similar genome browser systems: (a) Read/Write ability: Genquire is a viewer, an intuitive graphical editor, and a de novo annotation and markup tool for genome databases. (b) Flexibility: Genquire makes no pre-suppositions regarding the types of genomic features that may be created, and thus provides researchers a 'sandbox' in which to discover and annotate any region of DNA that is of interest to them. Included is the ability to bulk-annotate groups of hand-selected features, and to annotate and view features at the whole-genome level. (c) Compatibility: Genquire's database adaptor layer and API allows Genquire to 'connect to' and display a wide range of both databases and flat files, independent of the underlying schema. Thus, with a small amount of coding, Genquire can be used as a hand-annotation tool for almost any genome database. (d) Extensibility: Genquire offers a plugins API, allowing researchers to perform their own analysis of genomic regions from any data source using any analysis tool at their disposal, with results being overlaid (or optionally imported) into the underlying database. This allows direct comparison of newly generated data to existing local or remote data in the same display, and moreover does not require that the new sequence analysis algorithm be part of an existing pipeline.

ACKNOWLEDGEMENTS

This research was supported by the National Research Council Genomics and Health Initiative via a Plant Genomics Program grant to WLC. This publication is NRCC publication No. 43812. Thanks to The Arabidopsis Information Resource, in paticular Sue Rhee and Danny Yoo, for their support, assistance and and testing throughout code development. Thanks also to Suzi Lewis and the Apollo team for their enthusiastic moral support, to Nomi Harris for making the Genotator software freely available, and to Greg Helt for creating the original bioTkperl widget set.

REFERENCES

BioPerl Project, (http://www.bioperl.org) (manuscript in preparation).

Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genet.*, **25**, 25–29.

Harris,N.L. (1997) Genotator: a workbench for sequence annotation. *Genome Res.*, **7**, 754–762.

Helt,G. (1996) bioTkperl: graphics widgets for genomics. http://fruitfly.berkeley.edu/BDGP/informatics/bioTkperl.html.