BIOINFORMATICS APPLICATIONS NOTE

Vol. 27 no. 7 2011, pages 1015–1016
doi:10.1093/bioinformatics/btr056

Genome analysis

Advance Access publication February 3, 2011

Gobe: an interactive, web-based tool for comparative genomic visualization

Brent S. Pedersen*, Haibao Tang and Michael Freeling

Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

Associate Editor: Dmitrij Frishman

ABSTRACT

Summary: Gobe is a web-based tool for viewing comparative genomic data. It supports viewing multiple genomic regions simultaneously. Its simple text format and flash-based rendering make it an interactive, exploratory research tool. Gobe can be used without installation through our web service, or downloaded and customized with stylesheets and javascript callback functions.

Availability: Gobe is a flash application that runs in all modern web-browsers. The full source-code, including that for the online web application is available under the MIT license at: http://github.com/brentp/gobe. Sample applications are hosted at http://try-gobe.appspot.com/ and http://synteny.cnr.berkeley.edu/gobe-app/.

Contact: bpederse@berkeley.edu

Received on September 13, 2010; revised on January 7, 2011; accepted on January 24, 2011

1 INTRODUCTION

As more genomic data and data formats are created, there is an increasing demand for a tool that accepts a simple text-based format and renders an interactive view of the data that can then be used to develop hypotheses and convey information. A web-based application has the advantage of being more accessible and easily deployable, allowing users to quickly visualize and share data.

Existing genome visualization tools such as GBrowse (Stein *et al.*, 2002), GenomeGraphs (Durinck *et al.*, 2009) and genoPlotR (Guy *et al.*, 2010) generate useful static images but require installation, use in a specific programming language environment and do not provide interactivity. Browsers such as Anno-J (http://www.annoj.org) and the UCSC genome browser (Kent *et al.*, 2002) are interactive but more difficult to set up and do not easily support comparison of multiple genomic regions. Other stand-alone viewers such as GenomeView (http://genomeview.org/) and Tablet (Milne *et al.*, 2009) excel in displaying next-generation sequence data but are less customizable. While tools such as VISTA (Frazer *et al.*, 2004) do offer a comparative view, deletions, inversions and transpositions are not visible given that representation.

We introduce Gobe, a customizable, language-agnostic tool that can be run locally or used from an existing web-service. In either case, all data and styling are specified through a simple text format. Optional javascript callbacks allow interactivity for web developers.

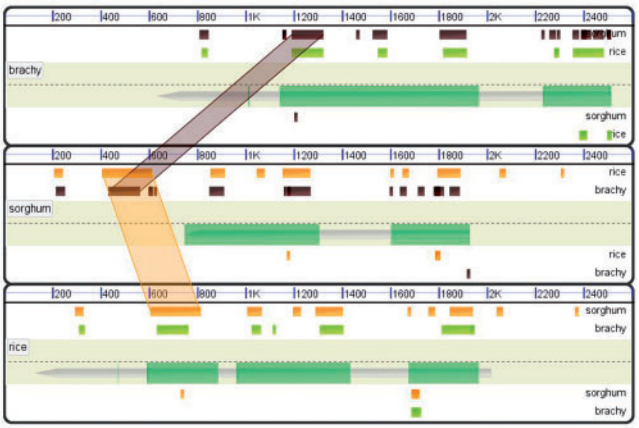

Fig. 1. Gobe view of orthologous genes in *Brachypodium distachyon* (brachy), *Sorghum bicolor* (sorghum) and *Oryza sativa* (rice). See text for description.

2 IMPLEMENTATION

2.1 Installation and customization

Though most users will be unaffected by implementation details, Gobe is implemented in HaXe (http://haxe.org), an open source programming language that compiles to the flash platform (http://www.adobe.com/software/flash/about/). Because flash is already installed in most web browsers, it will run for most users without installation. We have found that flash provides a balance of performance, interactivity and availability (Gobe requires flash version 9) that is hard to match with other technologies such as HTML5 and Java. We provide the HaXe source code for those wishing to modify the application itself.

Gobe represents multiple genomic regions as rows or tracks in the flash movie. A track can contain genomic features, text labels and line or histogram plots of numerical data. Gobe can render alignments between genomic regions as high-scoring sequence pairs (often called HSPs in BLAST-like alignment outputs) as wedges connecting the regions of sequence similarities. This helps to study the extents of sequence conservation or genomic rearrangements (see example in Fig. 1).

Mouse actions, such as clicks, on genomic features in the flash movie will send callbacks to a javascript function. This allows further interactivity and customization. For example, a developer may run their own installation of Gobe that would override the javascript callback to retrieve and display detailed information for

*To whom correspondence should be addressed.

a feature when it is clicked as is done on the http://synteny.cnr.berkeley.edu/gobe-app/site.

Each genomic feature that is displayed in Gobe has an associated feature type (exon, intron, mRNA, transposon, etc.). The feature type is used to look up the rendering style in a stylesheet. For example, the type of 'mRNA' may have a height of 0.6, a color of green and an arrow shape to illustrate its transcriptional orientation. In addition to the type-based specification of style, one is also able to have per-feature color in the input file, which allows highlighting a specific feature of interest. A user may provide her own stylesheet to customize the look of features in a particular application.

2.2 Data formats

Gobe uses a comma-separated format with five required columns: *arbitrary id*, *track-id*, *start*, *end* and *feature-type*. Further optional columns may indicate strand, name and special color (see full specification at http://github.com/brentp/gobe). Consecutive lines with type 'HSP' are linked so that when clicked, a wedge is drawn between them. This format has the advantage of consistency and brevity while allowing simple specification of pair relationships between features. We include scripts to convert from other popular formats such as BED (http://genome.ucsc.edu/FAQ/FAQformat#format1), GFF3 (http://www.sequenceontology.org/gff3.shtml) and tab-delimited BLAST (Altschul *et al*., 1997) to the Gobe input format.

3 USAGE

When viewing orthologous regions, it is useful to have an interactive view that shows regions of similarity represented as HSPs. Figure 1 displays a comparative view of orthologous regions in three grass genomes—brachypodium, sorghum and rice. Each track (title on the left) contains annotated genes (light-gray) and exons (green) drawn in the center of each panel. HSPs are drawn either above or below the annotations, depending on their orientation. In Figure 1, orange, green and dark-brown HSP boxes indicate rice–sorghum, sorghum–brachypodium and rice–brachypodium pairs, respectively. When an HSP is clicked in the viewer, a semi-transparent wedge is drawn to connect the pairs.

In this example, the HSPs occur in the same order among homologs, suggesting that gene structure is well conserved since these three organisms diverged. The connector wedges in Figure 1 highlight a region in sorghum where there is likely an unannotated exon, as evidenced by regions of similarity covering exons in both rice and brachypodium. Other non-genic regions of similarity are conserved non-coding sequences—candidate regulatory sites.

4 WEB SERVICE

As an example usage of Gobe, we provide a fully functional, open-source web service backed by python scripts and hosted on Google App-Engine. This service is available at http://try-gobe.appspot.com/ and allows users to paste in their own genomic data, view the resulting interactive flash movie, make changes to past views and provide a link to that specific set of annotations. All results are retained on the server and accessible through a unique URL.

The site also contains a panel with links to example datasets exemplifying Gobe use in rendering comparative datasets, a dataset with next-generation sequencing data, and one showing the use of line and histogram data for displaying Guanine-Cytosine content.

We have also created an additional example application that implements dynamic panning, zooming and reverse-complementation here: http://synteny.cnr.berkeley.edu/gobe-app/. This application also shows the sequence alignment for each HSP as it is clicked.

5 CONCLUSIONS

We have created a visualization tool that relies on text input to specify the way the features are drawn and how they are related. It provides a means of comparing up to (a practical limit of) about 10 related genomic regions and visualizing associated numerical data. Gobe can be customized to offer further interactivity. This tool provides the basis for custom applications that utilize orthology and numerical data to reveal more about a region of interest. We have created two web sites that demonstrate some of the possibilities of using Gobe to create custom applications.

ACKNOWLEDGEMENTS

The authors would like to thank James Schnable and Eric Lyons for helpful comments.

Funding: National Science Foundation Plants Genome Research Program (DBI0337083).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Durinck,S. *et al*. (2009) GenomeGraphs: integrated genomic data visualization with R. *BMC Bioinformatics*, **10**, 2.

Frazer,K.A. *et al.* (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res.*, **32**, W273–W279.

Guy,L. *et al*. (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics*, **26**, 2334–2335.

Kent,W.J. *et al*. (2002) The Human Genome Browser at UCSC. *Genome Res.*, **6**, 996–1006.

Milne,I. *et al*. (2009) Tablet—next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.

Stein,L.D. *et al*. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.