BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 9 2011, pages 1307–1308
doi:10.1093/bioinformatics/btr111

Genome analysis

Advance Access publication March 3, 2011

Comparative visualization of genetic and physical maps with Strudel

Micha Bayer[1,*], Iain Milne[1], Gordon Stephen[1], Paul Shaw[1], Linda Cardle[1], Frank Wright[2] and David Marshall[1]

[1]Genetics Programme, Scottish Crop Research Institute and [2]Biomathematics and Statistics Scotland, Invergowrie, Dundee, DD2 5DA, UK

Associate Editor: John Quackenbush

ABSTRACT

Summary: Data visualization can play a key role in comparative genomics, for example, underpinning the investigation of conserved synteny patterns. Strudel is a desktop application that allows users to easily compare both genetic and physical maps interactively and efficiently. It can handle large datasets from several genomes simultaneously, and allows all-by-all comparisons between these.

Availability and implementation: Installers for Strudel are available for Windows, Linux, Solaris and Mac OS X at http://bioinf.scri.ac.uk/strudel/.

Contact: strudel@scri.ac.uk; micha.bayer@scri.ac.uk

Received on December 21, 2010; revised on February 11, 2011; accepted on February 24, 2011

1 INTRODUCTION

Crop genetics is still dominated by species for which fully sequenced and well-annotated genomes are unavailable. Comparative genomics is an important means of annotating unfinished genomes, and requires powerful visualization tools that elucidate the relationships with already annotated genomes.

There are a number of tools in this area, which range from web-based applications with database back-ends to standalone desktop applications (Fang *et al.*, 2003; Lewis *et al.*, 2002; Meyer *et al.*, 2009; Mueller *et al.*, 2008; Pan *et al.*, 2005; Sawkins *et al.*, 2004). The challenges faced by any comparative visualization application are the increasing volume of data, fast delivery of these to users, efficient on-screen rendering of a large amount of information and layout constraints.

Here, we present Strudel, a standalone Java desktop application that aims to combine ease of installation with ease of use, and allows the simultaneous multi-way comparison of several genomes. Usability has been a major design criterion for Strudel, and in early acceptance testing users were able to start generating insights into their data within minutes of downloading the application, without having to first consult the manual. Strudel's graphical interface has been designed to reduce visual clutter as much as possible, and a critical condition for this is that homologies between two chromosomes are never drawn across other genomes.

*To whom correspondence should be addressed.

2 IMPLEMENTATION

Strudel ships in easy-to-use installers bundled with Java Runtime Environments (JREs), so there is no requirement to install additional software, and no Java version issues. It is available for Windows, Linux, Solaris and Apple Mac OS X at http://bioinf.scri.ac.uk/strudel/. The installers feature an auto-update facility which alerts users to new releases and provides the option of upgrading the software.

In its current implementation, data input into Strudel is by means of flat text files only. This provides the advantage of users being able to generate their own datasets easily, for example, in spreadsheet software, without relying on complex database back-ends. Strudel uses its own simple data format as standard formats for comparative data that have not been developed so far. The Strudel file format is row based, with both features and homologs in a single plain-text file. The format is documented in the online manual, where an example file is also provided.

Example datasets are provided both on the Strudel web site and with the application itself. The example dataset that is distributed with the application consists of three cereal genomes with a high degree of conserved synteny: barley (*Hordeum vulgare*), rice (*Oryza sativa*) and the model grass species *Brachypodium distachyon*. The latter two species have complete physical maps, while barley is supplied as a consensus single nucleotide polymorphism (SNP) map (Close *et al.*, 2009). A worked example is provided (http://bioinf.scri.ac.uk/strudel/#useCases) of how Strudel was used to investigate the barley Int-C mutant (Ramsay *et al.*, 2011). This involved using the high density barley SNP map, and exploring syntenous regions of the rice and Brachypodium genome in order to identify potential candidate genes through the links to the rice and Brachypodium genome browsers.

The graphical interface of Strudel is shown in Figure 1. Genomes are arranged in columns, with chromosomes represented by vertical bars. Features on chromosomes—for example, SNPs or genes—are rendered as horizontal lines, and pairs of homologs are represented by lines between the features involved. Rendering features and homology lines is the computationally most costly part of the canvas drawing operation, but at lower zoom levels this is accelerated by avoiding duplicate drawing operations for features and links that occupy the same on-screen (pixel) coordinates when zoomed out only. This allows Strudel to display feature-dense genomes with tens of thousands of features without noticeable impact on rendering speed.

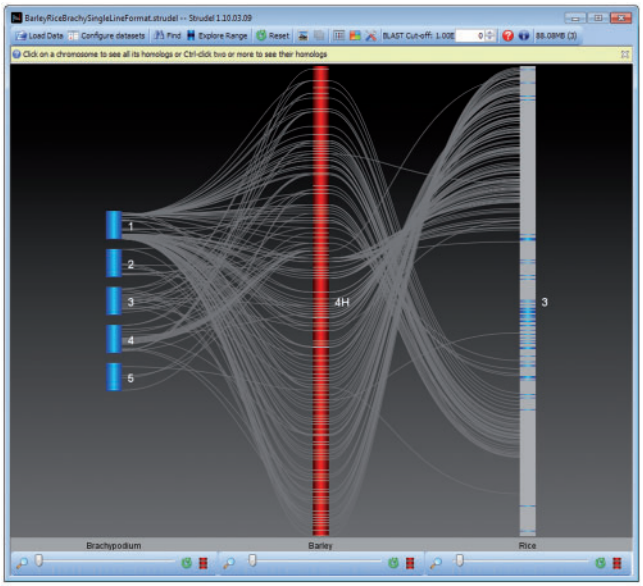

Fig. 1. Strudel's graphical interface, showing the example dataset provided with the application. Chromosome 4H of the barley genome has been expanded to fill the screen, showing homologies (gray lines) with the *B.distachyon* genome (left) and chromosome 3 of rice (right).

It is assumed that homology data for Strudel datasets are generated by BLAST (Altschul *et al*., 1990; not part of Strudel's functionality) or similar tools, and therefore a facility is provided that allows the user to filter the visible homologies (links) by, for example, their BLAST *e*-value to generate a more stringent view of the data as required. Other numerical variables can be used for this instead of *e*-values if required.

The number of genomes that can be compared is theoretically unlimited and only constrained by the available screen space. Additional graphical instances of genomes can be added without duplicating data (hence, conserving memory), to allow multi-way comparisons. Users are able to choose the number and position of the additional genome instances.

Quantitative trait loci (QTL) or any other regions of interest can be explored by defining an interval on a given chromosome. A table with the features contained within that region is then displayed. The table contains names of features and their homologs, along with their positions and any annotation information available. Feature names are clickable and the associated links point at user-defined URLs that provide annotation for the feature in question. Searching for a feature by name is also possible and results in a table being displayed such as that described above.

Zooming individual genomes is possible by means of individual zoom sliders at the bottom of each genome column, or by a click-and-drag motion that allows for a region to be highlighted and then zoomed into when the mouse button is released. High-level (zoomed out) views allow users to establish patterns of conserved synteny between genomes.

Chromosomes can be inverted to help to disentangle crossed-over links to regions that have undergone chromosomal inversion events. Users can customize numerous display features, such as color schemes, the shape of links (straight, curved or angled), enabling/disabling antialiased drawing and displaying distance labels.

A separate overview window shows all maps as laid out on the main canvas and allows easy orientation when the user has zoomed into one or more genomes. It also allows quick navigation by means of a highlighted area that shows the region currently visible on the main canvas in the context of the whole genome.

We have also developed close integration with the Germinate 2 (http://bioinf.scri.ac.uk/public/?page_id=159) database warehouse system to allow additional information on markers and genetic maps to be displayed to the user. This is performed by allowing the seamless movement between Strudel and the Germinate 2 web application and vice versa. In addition with the integration of Germinate 2 with our graphical genotype visualization tool, Flapjack (Milne *et al*., 2010), we have created an interactive and extendable software environment. Genetic maps held in Germinate 2 can be easily exported in Strudel format. Similarly, any other data source could in theory be adapted to interact with Strudel in this way, both in terms of data export to Strudel, and in terms of providing annotation URLs that can be accessed through the application (see above).

An online help manual is available, which includes a quickstart tutorial. There is also a hint panel built into the application that provides context-dependent advice on what actions are available in a given situation. This allows users to start using the application without constant referral to the manual.

ACKNOWLEDGEMENTS

We would like to thank our colleagues within the Genetics Programme at SCRI for their input to this project.

Funding: The work on Strudel was supported by the Scottish Government Rural and Environment Research and Analysis Directorate (RERAD), and the Scottish Funding Council and Scottish Enterprise through the Scottish Bioinformatics Research Network (SBRN) project.

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al*. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215**, 403–410.

Close,T.J. *et al*. (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics*, **10**, 582.

Fang,Z. *et al*. (2003) cMap: the comparative genetic map viewer. *Bioinformatics*, **19**, 416–417.

Lewis,S.E. *et al*. (2002) Apollo: a sequence annotation editor. *Genome Biol.*, **3**, research0082.

Meyer,M. *et al*. (2009) MizBee: a multiscale synteny browser. *IEEE T. Vis. Comput. Gr.*, **15**, 897–904.

Milne,I. *et al*. (2010) Flapjack – Graphical Genotype Visualization. *Bioinformatics*, [Epub ahead of print; doi:10.1093/bioinformatics/btq580, last accessed date October 18, 2010].

Mueller,L.A. *et al*. (2008) The SGN comparative map viewer. *Bioinformatics*, **24**, 422–423.

Pan,X. *et al*. (2005) SynBrowse: a synteny browser for comparative sequence analysis. *Bioinformatics*, **21**, 3461–3468.

Ramsay,L. *et al*. (2011) INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED. *Nat. Genet*., **43**, 169–172.

Sawkins,M.C. *et al*. (2004) Comparative Map and Trait Viewer (CMTV): an integrated bioinformatic tool to construct consensus maps and compare QTL and functional genomics data across genomes and experiments. *Plant Mol. Biol.*, **56**, 465–480.