Comparative Genomes Viewer

2D visualization of alignments and annotations

Ivan Molineris, Gabriele Sales and Michele Caselle
Theoretical Physics Department, Università di Torino, Via Pietro Giuria 1, Torino, Italy
Email: molineri@to.infn.it, sales@to.infn.it

Motivation

The amount of information about genomes, both complete sequences and annotations, has been exponentially increasing in the last few years. As a result there is the need for tools providing a **graphical representation** of such informations being both **comprehensive and intuitive**.

Visual representation is especially important in the comparative genomics field since it should provide a combined view of informations belonging to different genomes.

We believe that existing tools are limited in this respect as they concentrate the attention on a single genome at a time (conservation histograms) or compress alignment representations in a single dimension. Therefore we took example from dot-plots (actually **using the full 2D space**) and improved them both in terms of rendering speed and information content.

Methods

We start by performing whole genome high sensitivity alignments using the BLAST software.

We treat each alignment as a segment in a bidimensional plane whose sides correspond to the two chromosomes being aligned (as in a dot-plot). We cluster bidimensionally these segments and use the resulting hierarchical structure to store them efficiently in a MySQL database; this form has the added benefit of making queries extremely fast (up to 40 times with respect to the naïve format).

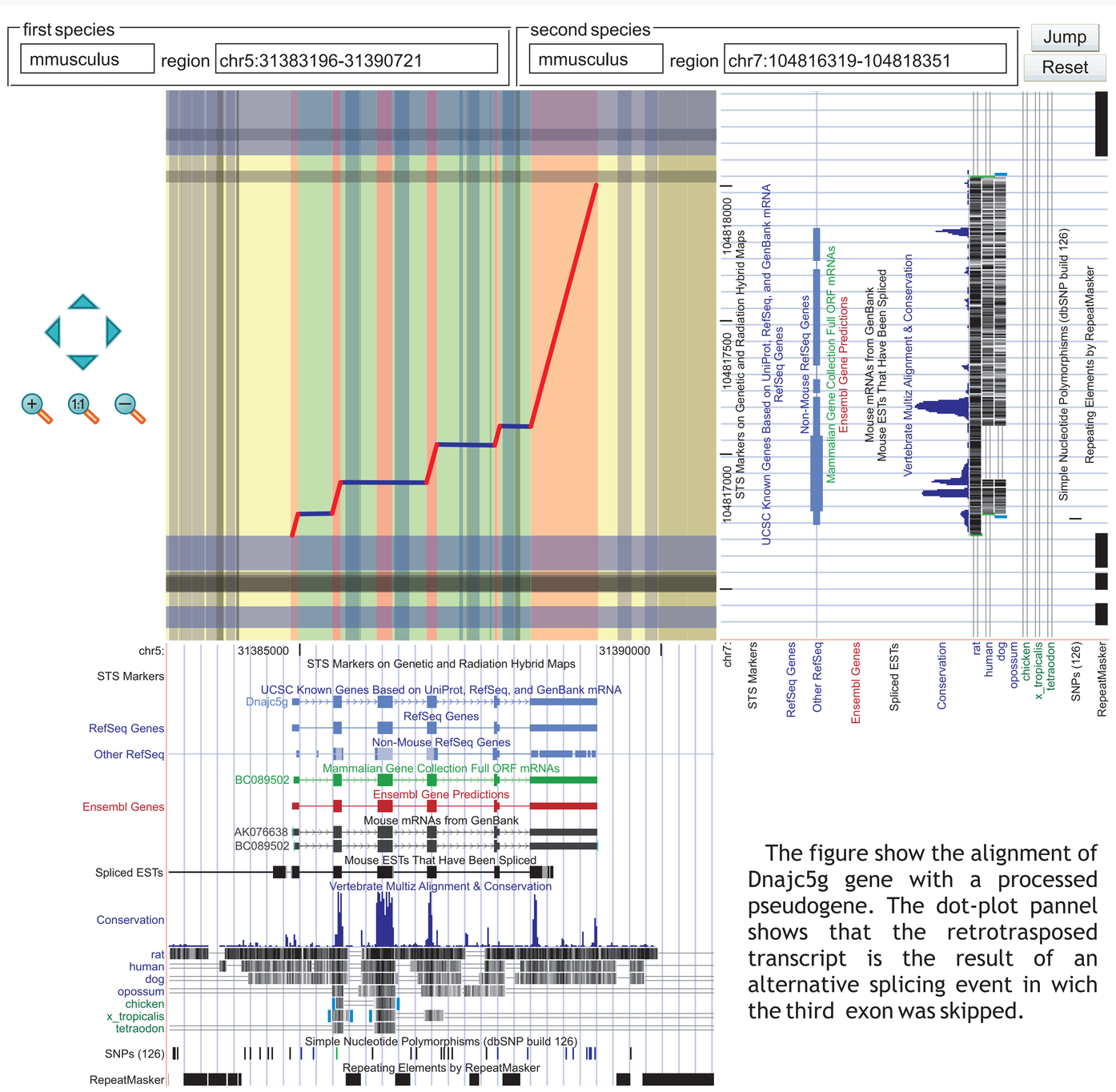
Given a pair of regions in chromosomal coordinates, our system retrives all the overlapping alignments (segments) and presents them graphically; the background of the image is coloured depending on the functional annotations.

Moreover we retirve gene annotations and other tracks from the UCSC genome browser and we show them at both sides of our plot.

Results

We developed a web based tool called Comparative Genomes Viewer (CGV): it **integrates the expressive power of dot-plots** (allowing an intuitive representation of alignments between two regions, both at small and big scales) **with the richness of informations typical of existing genome browsers**.

We give access to our system through a web based interface allowing the user to update the representation interactively and in **real time**, for example by easy to use (click and drag) shifting and zooming features. Our tool is also **embeddable** in external web pages by means of a straightforward **AJAX API**.



The figure show the alignment of Dnajc5g gene with a processed pseudogene. The dot-plot pannel shows that the retrotrasposed transcript is the result of an alternative splicing event in wich the third exon was skipped.

Interface

- species
- coordinates
- move buttons
- zoom
  - point and click (AJAX)
  - zoom in
  - adjust ratio (1:1)
  - zoom out

Dot-Plot Panel

- segments (represents 2D informations)
  - HSPs (alignments)
  - corruption gaps
  - deletion/insertion gaps
- bands (representing 1D informations)
  - coding regions
  - introns
  - UTRs
  - transposons
  - non trasposable repeats

Flanking Standard Genome Browsers

- for both regions
- real-time synchronization whith dot-plot panel
- multi-species conservation rates from other sources

Alignment datasets

We performed whole genome alignments of:

Homo sapiens ⟷ Homo sapiens

Mus musculus ⟷ Mus musculus

Homo sapiens ⟷ Mus musculus

using BLAST (Washington University implementation) with the following parameters:

- seed: 11
- scoring matrix: pam10
- score cutoff: 5580 (corresponding to 30 consecutive matches)

We were able to build each of these datasets in about three weeks using our infrastructure.

We can also integrate pre-made alignments or build them from scratch starting from custom sequences.

How to integrate in your web pages:

1) include a single js file to get the full AJAX API

2) make a single js call to the function `load_cgv` for each instance of CGV in your page

```
<script type="text/javascript" src="http://remoteserver.it/cgv.js">
<script  type="text/javascript">
    load_cgv(1307404, { x_species:hsapiens,
                        x_chr:10,
                        x_start:74597900,
                        x_stop:74665211,
                        y_species:hsapiens,
                        y_chr:15,
                        y_start:95100946,
                        y_stop:95102950})
</script>
```