



AARHUS
UNIVERSITET

INSTITUT FOR MOLEKYLÆRBIOLOGI OG GENETIK

FAUPE – Forbedring af Afgrødernes Udbytte og Produktionsmæssige Egenskaber

Arbejdspakke 2 – Genome browser

STØTTET AF
promilleafgiftsfonden
for landbrug

Genome browsers for the FAUPE project

A genome browser is a graphical interface to display information from biological databases in the context of genomic data. Genome browsers enable researchers to visualize and browse entire genomes with annotated data, including gene prediction, gene and protein architecture, structural variation, quantitative gene expression data, allele frequency and population data, etc.

Modern genome browsers, like GBrowse (Stein et al., 2002) and JBrowse (Skinner et al, 2009, Buels et al., 2016) implement a Web-based interface which is able to frame-up genome-scale sequences, megabases in length and can display an arbitrary set of features on nucleotide and protein sequences. Web-based genome browsers operate in a client-server mode, providing central databases and customizable interfaces that can be accessed from local computers via general Web browsers like Firefox or Internet Explorer. In addition to centrally provided genomic data, users can upload their own annotation data using standard file upload dialogs of the genome browser.

Uploaded user-specific features will persist between sessions. Although the uploaded user data will be located in a publicly accessible server, it is possible to restrict data access via secret keys stored in cookies on the end-user's computers. Sensitive private data stored exclusively on the end-user's machine can be visualized and analyzed in the server's genome browser too. On the other hand, users can also decide to share their uploaded data with peer users, peer user's groups or make certain data publicly available. For users not familiar with bioinformatics tools and pipelines, genome browsers are easy and convenient tools to download genomic data of interest (sections of genomic references sequences, transcript sequences, annotations, etc).

In the frame of the FAUPE project, according to the need of the project participant, we established genome browsers displaying the latest versions of annotated genome assemblies of perennial ryegrass, wheat and barley.

Configuring a virtual server on the GenomeDK cluster

As a first step for creating externally available Web services hosting genome browsers and genomic databases, a virtual machine was set up and configured on the GenomeDK cluster of Aarhus University. This virtual machine (tentatively named as vm8) presently has 16 GB memory, 8 GB storage space on its main node and 106 TB dedicated network storage space for databases. These resources have been sufficient to install all software and databases required by the present project, but to ensure a better performance and enough space for continuous software and database updates, upgrades (both in terms of memory and storage space) might be necessary in a short time. The virtual machine *vm8* is a self-containing computer, running CentOS Linux 6. Apache2 Web Server, MySQL database server, BLAST server and a wide range of UNIX tools and databases necessary for a versatile bioinformatics server are installed on it. It is operating independently from the GenomeDK host and is accessible from outside without an established connection to the Aarhus University network. The Web server of *vm8* is communicating to the outside world through the Port 5080 and ssh connection to *vm8* can be established via the Port 2022. For the FAUPE project an opening Web page with links to perennial ryegrass, wheat, barley and potato resources has been created. The address of this Web page is <http://185.45.23.197:5080/FAUPE.html>

Perennial ryegrass genome browser resources

The perennial ryegrass genome browser is available through a link on the FAUPE opening Web page (see above) or through a Web page originally created for the Perennial Ryegrass Genome Sequencing Project: <http://185.45.23.197:5080/ryegrassgenome>

The main perennial ryegrass genome browser is a GBrowse instance displaying genomic scaffolds and annotations described in the perennial genome paper, Byrne et al., 2015. An other ryegrass genome browser is also available. The later is a a JBrowse instance, showing detailed scaffolding information with read alignments in BAM assembly format.

Wheat genome browser resources

The wheat genome browser data are integrated into the perennial ryegrass GBrowse starting Wheat pseudo-

chromosomes and genomic scaffolds can be selected in the contig selector window, or by clicking on the links on one of the example chromosomes. The source of the current wheat genomic sequences and gene annotations is the IWGSC1.0(28) release of the *Triticum aestivum* genome, provided by Ensembl (<http://plants.ensembl.org>). To enable browsing of publicly available SNP data, we mapped flanking sequences and SNP positions of the Illumina iSelect 90K SNP bead chips (Wang et al. 2014) onto the three (A, B, D) homologous wheat genomes. These SNP data are displayed on separate tracks in the ryegrass/wheat genome browser.

Barley genome browser resources

The FAUPE barley genome browser is a JBrowse instance showing genomic sequences and annotations of the v1.31 *Hordeum vulgare* genome provided by Ensembl. SNP data can be displayed as separate track along with reference sequences and gene annotations. Presently the iSelect 8K SNP bead chip data (Close et al., 2009) are mapped on 7 pseudo-chromosomes of barley. It is announced that new, 50K SNP chip data will be available by the end of 2016. We are intending to integrate these data into the recent barley genome upon availability.

Potato genome browser resources

On the FAUPE Web page we are mirroring the current version of the Potato Genome Browser of the Potato Genome Sequencing Consortium, maintained by Michigan State University (USA). This browser is based on the v4.03 pseudomolecule sequences from the *S. tuberosum* Group Phureja DM1-3 assembly. A wide range of gene models and transcript data can be selected to display on the pseudomolecules, including representative genes (a manually curated set of annotated genes, represented by the longest transcript isoforms).

References

- Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, et al (2002): The generic genome browser: a building block for a model organism system database. *Genome Res* 12:1599–610
- Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009): JBrowse a next-generation genome browser. *Genome Res* 19:1630–8
- Buels R, Yao E, Diesh CM, Hayes RD3,6, Munoz-Torres, Helt, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016): JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biology* 17:66
- Close TJ et al. (2009): Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10:582
- Byrne SL, Nagy I, Pfeifer M, Armstead I, Swain S, Studer B, Mayer M, Campbell JD, Czaban A, Hentrup S, Panitz F, Bendixen C, Hedegaard J, Caccamo M, Asp T (2015): A synteny-based draft genome sequence of the forage grass *Lolium perenne*. *Plant Journal* 84:816–826
- Wang et al. (2014): Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* 12: 787–796

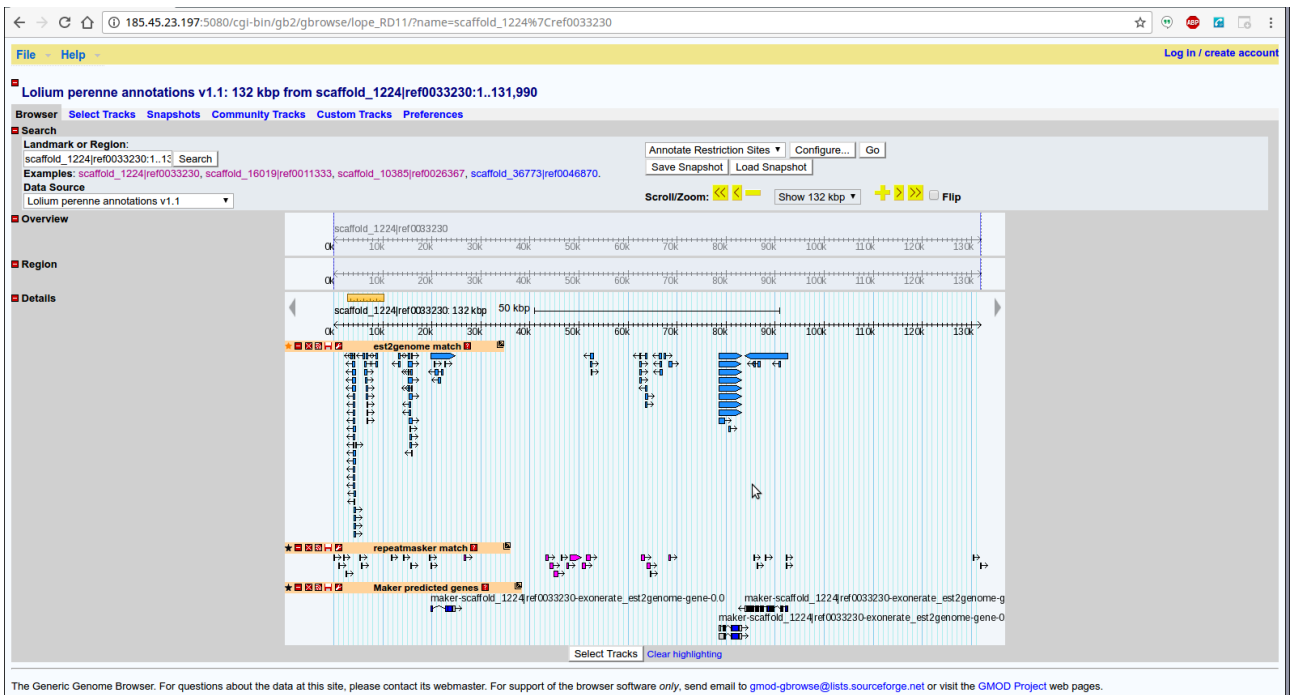


Figure 1 The perennial ryegrass genome browser (GBrowse)

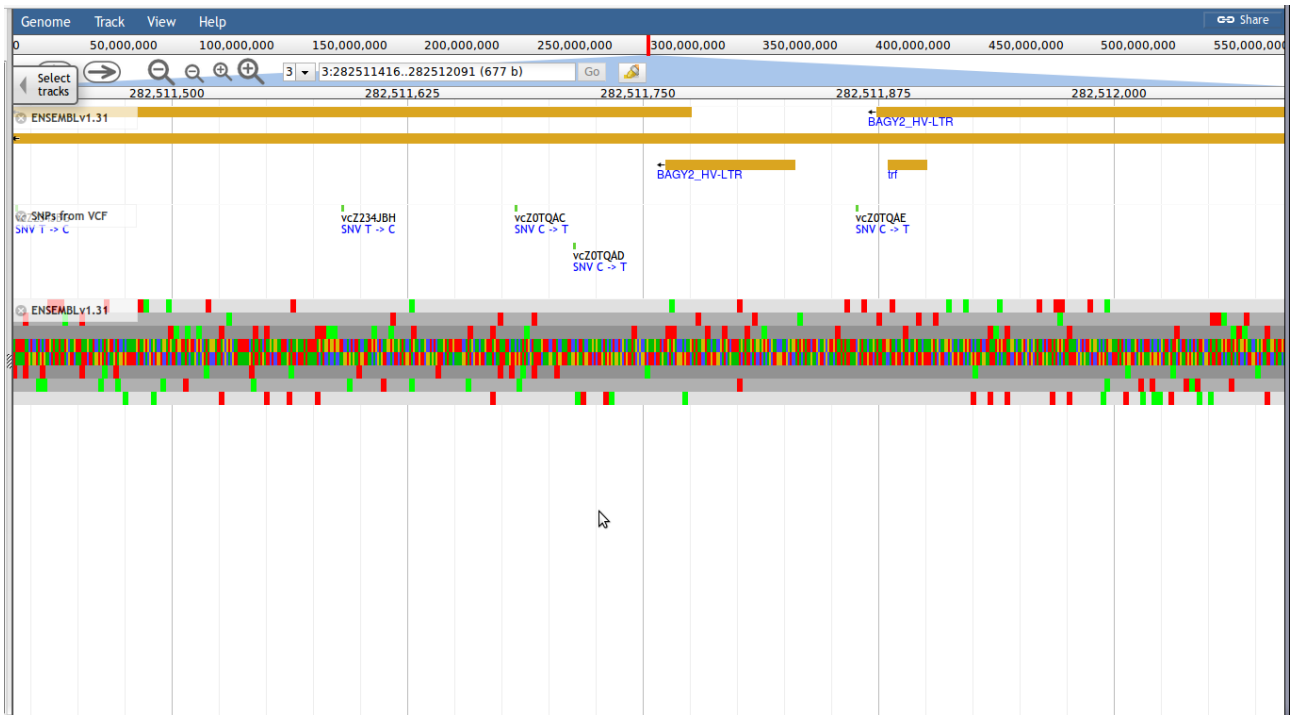


Figure 2 The barley genome browser (GBrowse)