



ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(3): 384-388
© 2022 TPI

www.thepharmajournal.com

Received: 20-01-2022

Accepted: 23-02-2022

Vivek Kumar

Research Scholar, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

Anu Sharma

Principal Scientist, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

SB Lal

Principal Scientist, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

KK Chaturvedi

Principal Scientist, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

Md. Samir Farooqi

Scientist, ICAR- Indian
Agricultural Statistics Research
Institute, New Delhi, India

DC Mishra

Senior Scientist, ICAR- Indian
Agricultural Statistics Research
Institute, New Delhi, India

Lakshmi Sonkusale

Research Scholar, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

Rajeev Kumar

Research Scholar, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

Corresponding Author

Anu Sharma

Principal Scientist, ICAR-Indian
Agricultural Statistics Research
Institute, New Delhi, India

Comparative analysis of genome browser for visualization of genetic variants

Vivek Kumar, Anu Sharma, SB Lal, KK Chaturvedi, Md. Samir Farooqi, DC Mishra, Lakshmi Sonkusale and Rajeev Kumar

Abstract

Genetic variations are used to describe the variations present in the DNA sequence of an individual's genome. Single Nucleotide Polymorphisms (SNPs) is the commonly used genetic variant for comparative genomics. Each SNP represents a difference in a single DNA base namely Adenine (A), Thymine (T), Cytosine (C) or Guanine (G) in DNA. Similarly, there are many other genetic variants namely Simple Sequence Repeat (SSR), Copy Number Variations (CNV), In Dels etc. At present, these genetic variants have been identified using various methods. These identified genetic variants are scattered at different locations in a genome and are also explored in isolation. A number of Genome browsers are used for the visualization of these variants on web. Therefore, it is required to compare the features of these browsers in order to select a best browser in a given situation. This paper describes the comparison of various features of important genome browsers and a web based prototype developed for visualization of SNPs using HTML, CSS, JavaScript, JSP and JBrowse. This tool will help users to access, addition of other genetic variants and its associated information through a single window interface.

Keywords: Genetic variants, SNPs, visualization, web

1. Introduction

In modern agriculture, genetics, plant breeding and government policies plays a significant role in increasing the agricultural products. Plant breeding is a science of changing the traits of plants for better characteristics and improving the qualities of products. In crop improvement, the application of genetics has yielded many successes leading to unprecedented growth in agricultural productivity. On other hand, there is a tremendous increase in computer technologies which brings students, researchers and scientist to a common platform in which they are all collaborating with each other in a convenient and efficient manner. By using the technologies and new methodologies one can accelerate their work. In present era, web platform plays a prestigious role in all sector.

Web interface play a significant role in field of bioinformatics in which one can visualize huge amount of genetic variants in a short effort and analyse these data to generate result. Genetic variants are used to describe the variations present in a DeoxyriboNucleic Acid (DNA) sequence of an individual's genome. Individuals belonging to distinct species has matching characteristics but they are barely exact, the difference between them is due to genetic variation. The identified genetic variants are scattered at different locations in a genome and are also explored in isolation. There is a need to integrate these genetic variants to bring at common platform where they can be easily explored and compared to identify the causes of such variations (Davidson, 2000) [3]. There are many genetic variants present in plants, some of them are Single Nucleotide Polymorphisms (SNPs), Simple Sequence repeats (SSR), Copy number variation (CNV) etc. These genetic variants are visualised through different genome browser.

Genome browser is a browser which provides a graphical view to end users to search, browse, retrieve, analyse sequence data and annotation track. Genome browser works on the principle of aggregation (Hubbard *et al.*, 2008) [4]. The genome browser helps researcher to visually, compare, analyse and correlate among several useful information from several different sources (Cline and Kent 2009) [2]. Mostly, genome browser is based on common gateway interface protocol (CGI). GBrowse uses CGI technique which imposes a page-based model of viewing and analysing the data in which each action performs and reloads the whole genome page, which results in significant delay and makes the user uncomfortable (Skinner *et al.*, 2009) [7]. Other major drawback of CGI technique is that the server perform most of the work such as showing genomic data to the end user. Hence, finally use of CGI technique result in significant increase

in computational cost and user's valuable waiting time for each new pages.

Both of these problems are solved by using JBrowse. It is a fast genome browser built using JavaScript and HTML5. It is available under open source, freely downloadable, portable and JavaScript-based genome browser. JBrowse can be used to navigate genome data over the web for visualization. JBrowse works in such a way that it preserve the sensibility of users by shrinking gapping transitions, in lieu of offering smoothly animated traverse, zooming, steering, and track selection. By using new technique such as HTML5, JavaScript and AJAX features seems that it will work for almost all modern web browsers in present situation. As it work on client side, it significantly reduce the burden of server-side handle costs, making it easier to support large numbers of end-users and large amounts of data.

A lot of work has been done in developing genome browser for exploring different genetic variants.

Table 1.1: Comparison of Important Genome Browsers

Ensemble	UCSC	GBrowse	JBrowse
Ensemble is based on Page-based Browsing technique (CGI Based)	UCSC is also based on Page-Browsing technique (CGI Based)	Gbrowse is based on Map-like Browsing technique within a Limited region	JBrowse is based on Map-like browsing technique along whole genome (AJAX+CGI)
Use server side scripting language	Use server side scripting language	Use set of Perl script and library station on server side scripting language	Use client side scripting language (HTML and Java script)
It is an extensible software architecture which support powerful API.	It is a freely downloadable software package for local browser installation and configuration.	This type of genome browser is based on multiple genome annotation visualization technique, especially for model organism database projects	It is a JavaScript and HTML5 based genome browser, It providing a Google-map like new browsing experience.

Although these genetic variants have been identified using various methods but they are scattered at different locations in a genome and are also explored in isolation. Various genome browsers like Ensemble, UCSC genome browser and GBrowse are available but these browsers use server side scripting language and are based on either Common Gateway Interface (CGI) or Asynchronous JavaScript and XML (AJAX) browsing technique. In this article we have developed a web based prototype for visualization of SNPs using HTML, CSS, JavaScript, JSP and JBrowse.

2. Materials and Methods

2.1 Configure JBrowse

JBrowse has been used as a genome browser for developing web based tool for visualization on genetic variants. JBrowse has been created by using Generic Model Organism Database (GMOD) platform, by developing some reusable software components. GMOD is freely available, open source software for storing, managing visualising and disseminating genetic and genomic data. JBrowse is freely available and it is an open source software. JBrowse is a fast, embeddable genome browser built completely with JavaScript and HTML5, with optional run-once data formatting tools written in Perl. JBrowse uses different file formats for visualizing various types of annotation on reference genome like FASTA, GFF3 and BAM etc. For our purpose, JBrowse-1.12.3 has been installed on Debian GNU/Linux.

For installing JBrowse, download JBrowse and extract JBrowse under a file served by web server. For Debian or Ubuntu environment the extracted folder must be put in /var/www/ and for Fedora, SUSE or Red Hat, in /var/www/html path. After extracting the JBrowse file, for running JBrowse, ./setup.sh shell script is executed. After successful run, JBrowse will show volvox data set example. Its means JBrowse is working. Now we need to set JBrowse and

Stein *et al.* (2002) ^[9] has developed Generic Model Organism Database project. It is recyclable software which models organism system databases. Manske *et al.* (2009) ^[6] has stated that sequencing of genome to greater depth may be much explanatory about variability within single one or an aggregate. Skinner *et al.* (2009) ^[7] has developed an open available genomic browser JBrowse. Wang *et al.* (2011) ^[10] has developed a modern rice genomic browser named Rice-Map. It is developed utilizing next-generation web techniques and high-throughput data based on experiment. Kong *et al.* (2012) ^[5] has developed ABrowse. It is a customizable and based on new-generation genome browser framework which is based on latest technologies which gives connective browsing skill. Buels *et al.* (2016) ^[1] has developed JBrowse, a more dynamically web based platform for genomic visualization and study. Comparative analysis of different genome browser is shown in Table 1.1

for this we need to prepare sequence alignment in FASTA file format. Before loading annotation data, we need to format our reference sequences for JBrowse using bin/prepare-refseqs.pl script. It will create data folder in JBrowse file. We need to put FASTA file in this data folder. And run the following script and open JBrowse on web; FASTA file will appear on the JBrowse.

```
bin/prepare-refseqs.pl --fasta
docs/tutorial/data_files/moong.fa
```

Now, we need to prepare annotated file in Generic Feature Format (GFF3). JBrowse can import feature data from flat files Format file by using bin/flatfile-to-json.pl script. Put Gff3 file into Data folder and run following script

```
bin/flatfile-to-json.pl --gff path/to/my.gff3 --trackType
CanvasFeatures --trackLabel mygff
```

Next, there is need to prepare BAM file. JBrowse can display alignments directly from BAM files, with no pre-processing necessary. Simply add a stanza with the relative URL of the file to your data/tracks.conf file, of the form:

```
# make a section for the new track called tracks. (unique name)
```

```
[tracks.my-bam-track]
# settings for what data is shown in the track
storeClass = JBrowse/Store/SeqFeature/BAM
urlTemplate = ../raw/volvox/volvox-sorted.bam
baiUrlTemplate = ../raw/volvox/volvox-sorted.bam.bai
# settings for how the track looks
category = NGS #< category for this track
type = JBrowse/View/Track/Alignments2
key = BAM alignments from sample XYZ
```

Refresh the genome browser BAM file will appear. BAM files

are required to be sorted and indexed, i.e., have a corresponding .bai file. JBrowse has two main track types that are designed especially for use with BAM data: Alignments2 and SNP Coverage. Alignment2 Shows individual alignments from the BAM file, as well as insertions, deletions, skipped regions, and SNPs encoded in the BAM's MD or CIGAR fields while SNP Coverage Shows a coverage histogram plot, with coloured bars showing the locations of base-level mismatches and possible SNPs in the reads.

2.2 Database Design

Database is created using MySQL for storing information of user's login and credentials and SNP information. Three tables namely crop, crop variety and SNP data table is used for storing SNP information.

2.3 Software Architecture

The proposed software has three layers namely; client side layer, server side layer and database layer as shown in Figure 1. Client Side Interface Layer (CSIL) is implemented using Hyper Text Markup language (HTML5), CSS and JavaScript. The CSIL consists of forms for accepting information from the user and its validation using JavaScript. It is also called as presentation layer which contains UI part of our application. This layer is used for the design purpose where data is presented to the user or input is taken from the user. In Server Side Application Layer (SSAL) all business logic written like validation of data, calculations, data insertion etc. This acts as an interface between CSIL and DBL. This layer is also called the intermediary layer helps to make communication faster between client and data layer. SSAL is implemented using JSP. DBL contains methods to connect with database and to perform insert, update, delete, get data from database based on our input data. Data base layer is implemented by using MySQL.

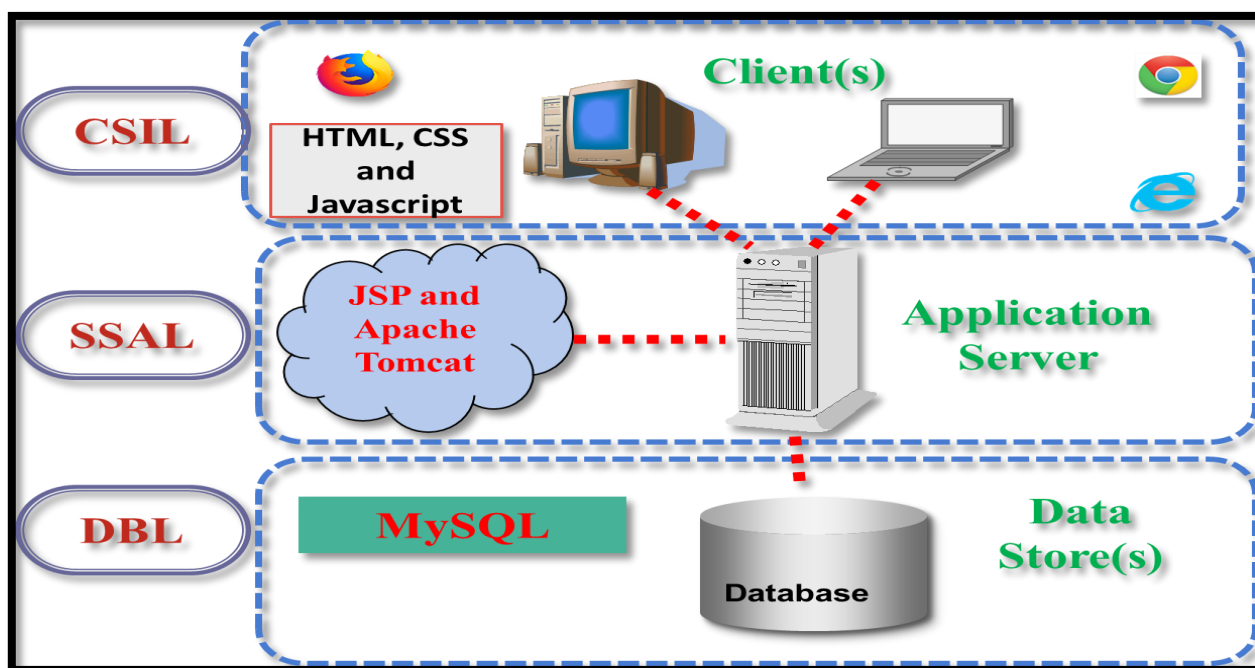


Figure 1: Three tier architecture of purposed software

3. Result

The developed tool is helpful to the users to access and study genetic variants and its associated information through single window. This tool has different menu options such as home page, about page, File Management, Processing, Visualization, Help, and Contact Us page. This tool is accessible to users through web browser at no cost. User authentication is needed

to ensure security. This tool is accessible only after entering valid user name and password. The Home page (Figure 2) of the software presents the user with a brief welcoming note on the software. It serves as the starting point of website. It is the default webpage that loads when any user visit the site. Users may create new account or may be logged in using the existing account.

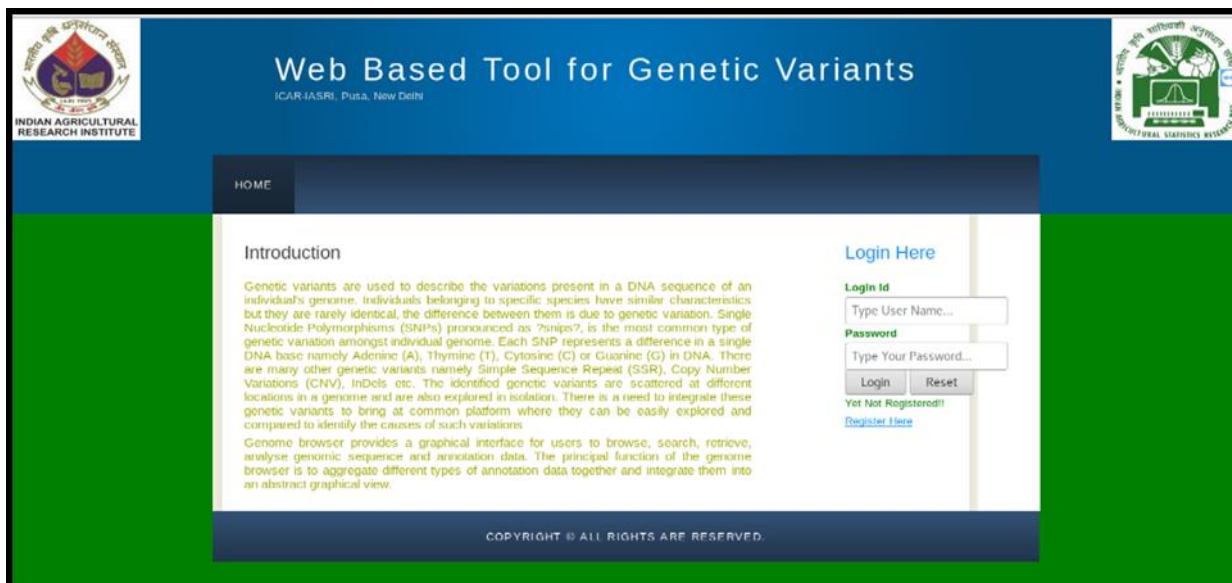


Fig 2: Home Page of Tool

After login, user has the rights to access all the contents of the developed tool. File management module consist of list of options such as upload, download, delete and view uploaded files. User can upload genome browser compatible file formats which is helpful in file management. Processing menu automates the various tasks involved in visualization. The steps involved during the processing of SNP file are preparation of reference file, preparation of annotated file (Genetic Variants), create a GFF3 file for genetic variants by fetching the data from

database through JSP for compatibility with JBrowse and JSP programs to automate various configuration parameter for visualization of GFF3 files into the JBrowse. Figure 3 shows the automatic GFF3 file created by fetching the data from SNPs table. After processing user can click on Visualization option to browse the genetic variants. When a user click on this tab, page navigates to genome browser, which shows the various variants data. It supports various data formats such as GFF3, BAM and FASTA etc. as shown in Figure 4.

```

1| nccb snp 10617 10619 .-.snp
2| nccb snp 10648 10650 .-.snp
3| nccb snp 10920 10922 .-.snp
4| nccb snp 10944 10946 .-.snp
5| nccb snp 10983 10985 .-.snp
6| nccb snp 11013 11015 .-.snp
7| nccb snp 11027 11029 .-.snp
8| nccb snp 11052 11054 .+.snp
9| nccb snp 11080 11082 .-.snp
10| nccb snp 11159 11161 .-.snp
11| nccb snp 11170 11172 .-.snp
12| nccb snp 11196 11198 .-.snp
13| nccb snp 11247 11249 .+.snp
14| nccb snp 11400 11402 .-.snp
15| nccb snp 11426 11428 .-.snp
16| nccb snp 1380 1382 .+.snp
17| nccb snp 1444 1446 .+.snp
18| nccb snp 1457 1459 .+.snp
19| nccb snp 1525 1527 .+.snp
20| nccb snp 1608 1610 .+.snp
21| nccb snp 1614 1616 .+.snp
22| nccb snp 1641 1643 .+.snp
23| nccb snp 1649 1651 .+.snp
24| nccb snp 1686 1688 .+.snp
25| nccb snp 1852 1854 .+.snp
26| nccb snp 1876 1878 .+.snp
27| nccb snp 1877 1879 .+.snp
28| nccb snp 1898 1900 .+.snp
29| nccb snp 1899 1901 .+.snp
30| nccb snp 2093 2095 .-.snp
31| nccb snp 2123 2125 .-.snp
32| nccb snp 2231 2233 .-.snp
33| nccb snp 2261 2263 .-.snp
34| nccb snp 2288 2290 .+.snp
35| nccb snp 2295 2297 .+.snp
36| nccb snp 2368 2370 .+.snp
37| nccb snp 2919 2921 .-.snp
38| nccb snp 3066 3068 .+.snp
    
```

Figure 3: Automatic GFF3 File Created from Database

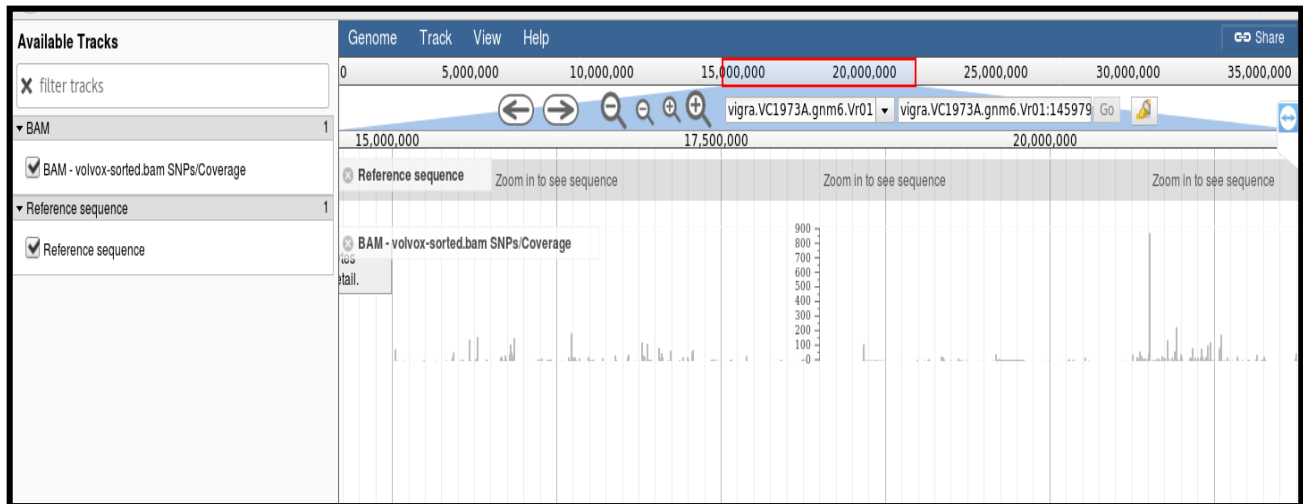


Figure 4: Visualization of Variants Data

4. Conclusion

Different genome browsers were studied and it was observed that JBrowse is a most widely used genome browser. The major advantage of this browser is that it work on client side rather than delegating every task to server side which minimize the significant amount of time. The developed tool visualizes GFF3 format of SNP data using JBrowse on client side and does all computation work on server side. The tool is helpful to the users to access and study genetic variants and its associated information through single window whereas CGI based genome browsers reload whole pages. This visualization tool can be further extended by including the many other file formats like BED, Tabix, GTF and VCF for JBrowse. Further, facility may be provided for comparison of genetic variants for other crops as well.

Rice-Map: a new-generation rice genome browser. *BMC genomics*. 2011;12(1):165.

References

1. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G *et al*. JBrowse: A dynamic web platform for genome visualization and analysis. *Genome biology*. 2016;17(1):66.
2. Cline MS, Kent WJ. Understanding genome browsing. *Nature Biotechnology*. 2009;27(2):153-155.
3. Davidson S. Research suggests importance of haplotypes over SNPs. *Nature Biotechnology*. 2000;18(11):1134.
4. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E *et al*. Ensembl 2009. *Nucleic acids research*. 2008;37(1):D690-D697.
5. Kong L, Wang J, Zhao S, Gu X, Luo J, Gao G. ABrowse- a customizable next-generation genome browser framework. *BMC bioinformatics*. 2012;13(1):2.
6. Manske HM, Kwiatkowski DP. LookSeq: a browser-based viewer for deep sequencing data. *Genome research*. 2009, gr-093443.
7. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. J Browse: A next-generation genome browser. *Genome research*. 2009;19(9):1630-1638.
8. Stalker J, Gibbins B, Meidl P, Smith J, Spooner W, Hotz HR *et al*. The Ensembl Web site: mechanics of a genome browser. *Genome research*. 2004;14(5):951-955.
9. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A *et al*. The generic genome browser: a building block for a model organism system database. *Genome research*. 2002;12(10):1599-1610.
10. Wang J, Kong L, Zhao S, Zhang H, Tang L, Li Z *et al*.