

Geno viewer, a SAM/BAM viewer tool

Miklós Laczik^{1,2}, Edit Tukacs^{1,2}, Béla Uzonyi¹, Bálint Domokos¹, Zsolt Doma¹, Máté Kiss¹, Attila Horváth^{1,2}, Zoltán Batta^{1,2}, Zsuzsanna Maros-Szabó^{1,2}, Zsolt Török^{1,2}

¹Astrid Research Inc., 4029 Debrecen, Csapó street 42., Hungary, ²Bioinformatics Research Group, University of Debrecen, Faculty of Informatics, 4010 Debrecen, POB 12, Hungary; Edit Tukacs – Email: edit.tukacs@astridbio.com; *Corresponding author

Received December 21, 2011; Accepted December 7, 2011; Published January 20, 2012

Abstract:

The ever evolving Next Generation Sequencing technology is calling for new and innovative ways of data processing and visualization. Following a detailed survey of the current needs of researchers and service providers, the authors have developed GenoViewer: a highly user-friendly, easy-to-operate SAM/BAM viewer and aligner tool. GenoViewer enables fast and efficient NGS assembly browsing, analysis and read mapping. It is highly customized, making it suitable for a wide range of NGS related tasks. Due to its relatively simple architecture, it is easy to add specialised visualization functionalities, facilitating further customised data analysis. The software's source code is freely available; it is open for project and task-specific modifications.

Availability: <http://www.genoviewer.com/>

Background:

Next Generation Sequencing (NGS) emerged with the promise of revolutionizing genomics. Although recently it has been continuously evolving and spreading, it has its share of limitations and challenges [1]. One of the essential elements of the NGS workflow is visualization, allowing researchers to check sequencing information and result of previous data analyses. Several NGS viewers, browsers and other visualizing tools have been designed to address this problem. Developers of Genevieve [2], while getting frequent NGS related projects, constantly felt the need for a standard viewer for customized data presentation and a uniform working platform. While online solutions are in general easy to share and support, they usually raise security and maintenance issues. Available offline tools on the other hand fell short of our expectations regarding easy operation and task-specificity. Therefore authors had decided to develop a new proprietary SAM/BAM viewer/browser tool for NGS data presentation. Genoviewer's source code is freely available [3]. GenoViewer, unlike other genome browser tools like IGV [4] or BamView [5], can display alignments in color space. This functionality is especially useful for the development of ABI color space alignment algorithms. GenoViewer works with the concept of projects -

different alignment and annotation files can be grouped together and saved for later use. This makes it easy to switch between different data analysis tasks and work on different datasets simultaneously. Another interesting functionality is its configurability, it is easy to define a new theme for the look of the display and change the visual markers for the different features.

Methodology:

Software engineers used Java (1.6) [6] as a uniform language during the development process, enabling it to run on all major operation systems. Another major benefit is the support of reusable codes owing to Java API, for accelerated development. There is no installation required to run Genevieve, files just need to be copied to the user's computer. OS specific starter files were created as well. A clearly structured workspace system was developed for handling input and output files. Users can generate a consensus sequence and a table collecting all mutations. Whenever an annotation file is available, an annotation table is automatically created. A dynamic loading system contributes to a very efficient memory usage, the whole input file is partitioned into small pieces and when a section is displayed, only the

corresponding file part is loaded into the memory. The internet community has been involved extensively in the development process, several test versions were downloaded and it generated an active forum thread at a NGS forum site [7]. Geneview underwent a three level testing system: 1) testing algorithms written specifically for this project (checking the basic functionality of the written part in itself and in context of the whole integrated system); 2) testing by an outsider (usually someone from different team); 3) testing by a competent person, a potential user, life science expert of an outsider academic partner. A flowchart depicting the relations between the software's components and algorithms is shown in **Figure 1**. In order to further improve GeneViewer's functionalities and features, we opened its code to the wider publicity.

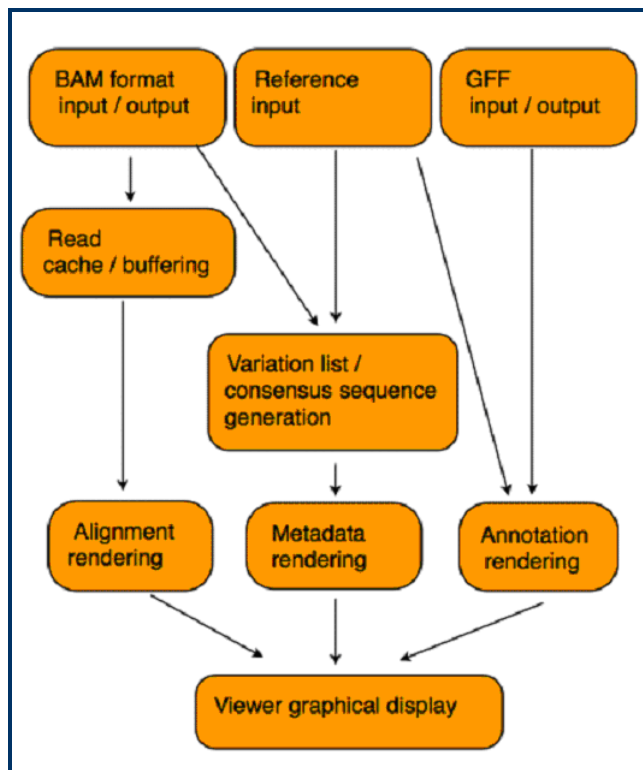


Figure 1: The flowchart illustrates the relations between the software's components.

Features:

Figure 2 shows the general structure of the GeneViewer window. The page contains four distinct areas: there is the menu at the top with functions and a toolbar for quick access to frequently used features; the left side panel with the workspace and project folder system; the central area for the alignment procedures; the toolbar at the bottom for additional functions and information. The software supports the following file formats: SAM/BAM [8], FASTA files for reference sequences and GFF [9] for annotations. The output is a mutation table in CSV format and the consensus sequence can be saved as a FASTA file. Read info is displayed in a pop-up window by clicking on a specific read. It shows details about the selected read, including exact position, the alignment to the reference (if a reference is available), mutations, mapping quality and various flags in the file marking different qualities.



Figure 2: The viewer has a traditional layout with menus and toolbar on the top, feedback tray at the bottom, and workspace/project folder tray, aligned left. Beneath the main visualization area there is a navigation panel, which makes it easy to jump through the mutations. Other functions are accessible through pop-up windows (Mutations table, Go to position, Edit profiles).

The main visualization area can be divided into five rows, showing the reference and consensus sequences, GFF annotation row, and coverage graph and mapped reads. The scrollable mutation table enables users to study altered nucleotides by type (SNP, MNP, insertion, deletion) and coverage. The table also lists the following related information: type, start and end position, length, reference and read sequence. GFF files can be loaded for annotation purposes. The groups and features in a GFF file are shown in the annotation field as lines and bands, with names written above and start/stop signs ([/]) discriminating overlapping annotations. Unique display modes ensure advanced customizability: reads are displayed as one continuous column or several evenly distributed columns, nucleotide versus colour coded reads, symboling turned on or off, etc.

Discussion and future directions:

GeneViewer has been launched as a freely modifiable and personalizable SAM/BAM viewer that can be used for a variety of NGS related tasks. Developers do not have the intention to morph it into complex software; it will remain open for further development and improvement by the wider scientific society. Potential future features include extended file handling (e.g. .bed [10]), extended NGS platform support, support of alternative alignment formats (ACE, NEXUS, ALN, PHYLIP etc.), more complex annotation, improved contig management, sequence and uncovered region search, multiple alignment of consensus sequences and comparative statistics tables.

References:

- [1] Gollery M. *Bioinformatics* 2011 **5**: 367 [PMID: 21383901].
- [2] <http://www.genoviewer.com>
- [3] <https://github.com/astrid/GenoViewer>
- [4] Robinson JT *et al. Nat Biotechnol.* 2011 **29**: 24 [PMID: 21221095]
- [5] Carver T. *et al. Bioinformatics.* 2010 **26**: 676 [PMID: 20071372]
- [6] <http://download.oracle.com/javase/7/specs/jls/JLS-JavaSE7.pdf>
- [7] <http://seqanswers.com/forums/showthread.php?t=6557>
- [8] <http://samtools.sourceforge.net/SAM1.pdf>
- [9] <http://www.sanger.ac.uk/resources/software/gff>
- [10] <http://www.ensembl.org/info/website/upload/bed.html>

Edited by P Kanguane

Citation: Laczik *et al.* *Bioinformatics* 8(2): 107-109 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.