BamView: viewing mapped read alignment data in the context of the reference sequence

Tim Carver∗, Ulrike Böhme, Thomas D. Otto, Julian Parkhill and Matthew Berriman
Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

ABSTRACT

Summary: BamView is an interactive Java application for visualizing the large amounts of data stored for sequence reads which are aligned against a reference genome sequence. It supports the BAM (Binary Alignment/Map) format. It can be used in a number of contexts including SNP calling and structural annotation. BamView has also been integrated into Artemis so that the reads can be viewed in the context of the nucleotide sequence and genomic features.

Availability: BamView and Artemis are freely available (under a GPL licence) for download (for MacOSX, UNIX and Windows) at: http://bamview.sourceforge.net/

Contact: artemis@sanger.ac.uk

Received on November 9, 2009; revised on January 6, 2010; accepted on January 7, 2010

1 INTRODUCTION

Second-generation sequencing produces large volumes of short-read sequence data. In common applications of the technology, such as resequencing or transcriptome sequencing, reads are mapped against a reference genome. In many cases, bases in a reference are covered with alignment depths varying by orders of magnitude and therefore present a challenge for visualization.

SAM (Sequence Alignment/Map) and BAM (Binary Alignment/Map) formats are emerging as a standard representation for read alignments. It is therefore important to have visualization software for this format. BAM files contain the same information as SAM. As BAM format is compressed it provides an efficient means to store the data and enables fast retrieval of regions and so this format has been adopted here.

SAMTools (Li et al., 2009) includes a very simple text alignment viewer using the GNU ncurses library giving a detailed view at the nucleotide resolution level. Lookseq (Manske and Kwiatkowski, 2009) is a perl-cgi application used to display, in a web browser, reads mapped against a reference. It can read the data either directly from BAM files or from a relational database to display reads and plots paired read positions against their inferred size.

Alignment tools can be used to produce BAM format files. For instance, SSAHA (Ning et al., 2001) now supports SAM format (Long et al., 2009) and Maq output (Li et al., 2008) can be converted to SAM/BAM using SAMTools.

BamView can be used as a stand-alone Java application or displayed in Artemis (Carver et al., 2008; Rutherford et al., 2000, Fig. 1) in conjunction with the reference sequence and annotation.

To whom correspondence should be addressed.

2 IMPLEMENTATION

The BAM file needs to be sorted and indexed using the SAMTools command tool. This creates the BAM index file that the viewer uses to access the region to display in a fast way. BamView uses picard (picard.sourceforge.net), which is a Java API to read from the BAM file the reads in the region of sequence being displayed.
When the user points the mouse cursor over a read, a tool-tip field is used in the pop-up menu. These are red vertical lines on the reads and indicate the start and end of a read. When forward and reverse reads are displayed as red nucleotides when zoomed in.

## 3 DISCUSSION

Using the different views in BamView and as an integrated window in the Artemis tool means that it has a range of uses. For example, BamView can be used to inspect the confidence of a deep alignment of short reads, by highlighting base discrepancies. As it is integrated into Artemis, the underlying reference consensus sequence can be edited directly based on manually inspecting data from transcriptome sequencing experiments (RNAseq, Otto et al., 2009; Wang et al., 2009). An annotator can zoom into intron–exon boundaries, identified from coverage plots, and see the quality of evidence supporting a prediction or manually adjust exons coordinates to fit the evidence. Simply viewing individually aligned reads cannot resolve alternate splicing patterns, but by clicking through read-pairs, an annotator can reconstruct the phase of exons in different isoforms.

## ACKNOWLEDGEMENTS

We would like to thank Gary Dillon, Jacqui McQuillan, Anna Proctor for their suggestions in the development of this application.

**Funding**
Wellcome Trust through their funding of the Pathogen Genomics group at the Wellcome Trust Sanger Institute (WT 085775/Z/08/Z).

**Conflict of Interest**
None declared.

## REFERENCES

Carver, T. et al. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics, **24**, 2672–2676.

Li,H. et al. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res., **18**, 1651–1658.

Li,H. et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics, **25**, 2078–2079.

Long,Q. et al. (2009) HI: haplotype improver using paired-end short reads. Bioinformatics, **15**, 2436–2437.

Manske,H.M. and Kwiatkowski,D.P. (2009) Lookseq: a browser-based viewer for deep sequencing data. Genome Res., **19**, 2125–2132.

Ning,Z. et al. (2001) SSAHA: a fast search method for large DNA databases. Genome Res., **11**, 1725–1729.

Ott,T.D. et al. (2009) New insights into the blood stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol. Microbiol., in press.

Rutherford,K. et al. (2000) Artemis: sequence visualisation and annotation. Bioinformatics, **16**, 944–945.

Wang,Z. et al. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet., **10**, 57–63.