BIOINFORMATICS APPLICATIONS NOTE Vol. 31 no. 4 2015, pages 596–598
doi:10.1093/bioinformatics/btu676

Sequence analysis

Advance Access publication October 15, 2014

miRseqViewer: multi-panel visualization of sequence, structure and expression for analysis of microRNA sequencing data

Insu Jang[1,†], Hyeshik Chang[2,†], Yukyung Jun[3], Seongjin Park[1], Jin Ok Yang[1], Byungwook Lee[1], Wankyu Kim[3], V. Narry Kim[2] and Sanghyuk Lee[3,*]

[1]Korean Bioinformation Center (KOBIC), KRIBB, Daejeon 305-806, Korea, [2]School of Biological Sciences, Seoul National University, Seoul 151-742, Korea and [3]Department of Life Science and Ewha Research Center for Systems Biology (ERCSB), Ewha Womans University, Seoul 120-750, Korea

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: Deep sequencing of small RNAs has become a routine process in recent years, but no dedicated viewer is as yet available to explore the sequence features simultaneously along with secondary structure and gene expression of microRNA (miRNA). We present a highly interactive application that visualizes the sequence alignment, secondary structure and normalized read counts in synchronous multi-panel windows. This helps users to easily examine the relationships between the structure of precursor and the sequences and abundance of final products and thereby will facilitate the studies on miRNA biogenesis and regulation. The project manager handles multiple samples of multiple groups. The read alignment is imported in BAM file format. Implemented features comprise sorting, zooming, highlighting, editing, filtering, saving, exporting, etc. Currently, miRseqViewer supports 84 organisms whose annotation is available at miRBase.

Availability and implementation: miRseqViewer, implemented in Java, is available at https://github.com/insoo078/mirseqviewer or at http://msv.kobic.re.kr.

Contact: sanghyuk@ewha.ac.kr

Received on May 28, 2014; revised on September 12, 2014; accepted on October 13, 2014

1 INTRODUCTION

MicroRNA sequencing (miR-seq) is a powerful method to study virtually all aspects of miRNA biology, including biogenesis, molecular function and expression of miRNAs. With the advances in deep sequencing technologies, more laboratories are adopting miRNA sequencing instead of microarrays or quantitative PCR for such analyses (Cho *et al*., 2013). Typical procedure of analyzing miR-seq data consists of preprocessing, mapping, quantification and identification of novel and/or differentially expressed miRNAs, as implemented in recent integrative platforms of MAGI (Kim *et al*., 2014) and CAP-miRSeq (Sun *et al*., 2014). Even though the sequence variations such as uridylation, iso-miRs and RNA editing are known to be important in biogenesis and function of miRNAs (Kim *et al*., 2009), we still lack a dedicated alignment viewer to examine the details of sequence features interactively and simultaneously with other features such as secondary structure and read abundance. To our knowledge, miRspring (Humphreys *et al*., 2013) is the only software to support analysis of sequence variations, but its visualization lacks interactive features.

Here, we present miRseqViewer, a Java-based standalone platform for visualizing miR-seq data. The read alignment is concurrently displayed with the secondary structure as well as the read count table that indicates miRNA expression in multiple samples of multiple groups. Diverse features are implemented to enhance user convenience.

2 IMPLEMENTATION

miRseqViewer is designed to be an OS-independent standalone program to support individual laboratories to analyze their own miR-seq data without uploading or transferring sequencing data to other computers. Main program was developed using Java Swing and JavaFX, thus requiring JDK 1.6 or higher version. The program requires minimal dependency on other software for easy installation. For example, no database system is required in spite of its extensive use of miRNA information extracted from the miRBase (Kozomara and Griffiths-Jones, 2014) since SQLJet (TMate Software), compatible with SQLite 3.6, is implemented as the backend database system. Basic GUI is based on docking windows framework under the General Public License (InfoNode Docking Windows from NNL Technology, 2009) that allows users to move, dock and undock windows dynamically. We have tested the program in multiple operating systems including Mac OS X, Windows and Linux.

Read alignment information is imported into miRseqViewer in BAM file format produced by other alignment programs. In principle, BAM files need to be pre-sorted and indexed using SAM tools (Li *et al*., 2009) or Picard (http://picard.sourceforge.net). If not, miRseqViewer will use Picard to create sorted and indexed BAM files during the importing step.

We support all miRNAs in the miRBase database from version 15 to version 21, covering 84 organisms in total. To avoid unnecessary dump of huge genome sequence data of 84 organisms to the user's computer, the genomic sequence of selected miRNAs to view are fetched in real time from our server via HTTP protocol. The computer should be thus on-line during

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

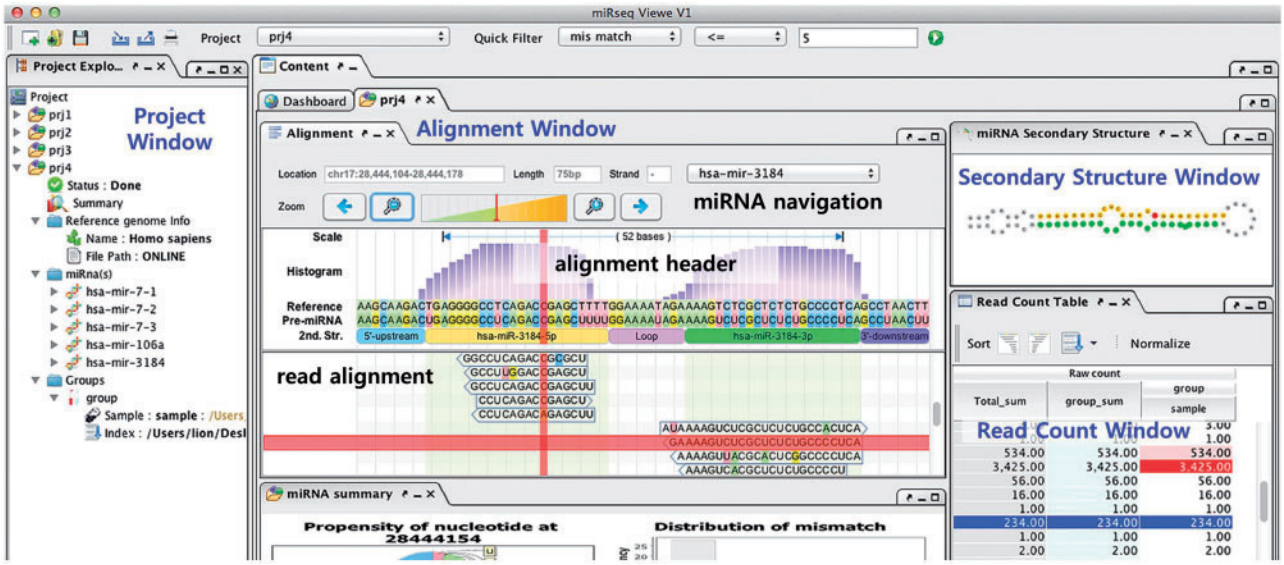

Fig. 1. Screenshot of miRseqViewer. Mouse actions are synchronized among multiple windows, including the nucleotide position (vertical bar within the alignment window and the red dot in secondary structure) and the read (horizontal bars within the alignment and read count windows) of interest. Note that the mismatched nucleotides are colored in the alignment. Most color schemes can be customized

the analysis. In addition, it is important to use the identical reference genome for read alignment and for visualization. We recommend using the genome assemblies as specified in the miRBase in the alignment process.

For novel miRNAs not included in the miRBase database, users should identify novel miRNAs in advance by running a program such as miRDeep2 (Friedlander *et al*., 2012), and prepare a BED file containing the genomic information on miRNAs of interest.

3 RESULTS

The GUI of miRseqViewer consists of project manager, alignment window, secondary structure window and read count table as illustrated in Figure 1.

Defining a project is the starting point of every analysis. We provide two wizards to define projects. Standard wizard allows users to specify organism, groups of samples for comparison and alignment BAM files, BED files for novel miRNAs of interest. Quick wizard is useful for simple examination of sample data for known miRNAs when no comparison between groups is necessary. BAM file should be pre-sorted and indexed in quick wizard mode.

The alignment window consists of navigation part, alignment header and read alignment parts. Users select a miRNA or genomic region to display. The depth profile of read alignment, reference genome sequence, pre-miRNA sequence from miRBase, and secondary structure segments are shown in the alignment header. Reads in the alignment window can be sorted in various ways, moved by drag-and-drop and edited.

Secondary structure of pre-miRNA is calculated using ViennaRNA package in real time (Lorenz *et al*., 2011). The color code and cursor position are coincident with those in the alignment window. This helps users to explore sequence features along with secondary structure information readily.

Read count table shows the normalized read counts of each read for samples and groups as defined in the project. The background color indicates relative abundance. We implemented four normalization methods (RPM, TMM, Lowess and quantile) as well as diverse sorting and display options. Highlighted row in the table is synchronized with the selected read in the alignment window.

Several other features are implemented as well for user convenience. Filtering by read count, read length or mismatches is supported. Exporting projects saves all alignment data and visualization parameters into a local file (**.msv*), which can be shared with colleagues or imported at a later time. The exported *msv* file contains all information for miRseqViewer function including read alignment, but the file size is much smaller because only mapped reads are stored. In fact, the original BAM file is no longer used in miRseqViewer once the project is created. We further provide a summary window that reports nucleotide propensity, distribution of mismatches, read length, etc.

In conclusion, miRseqViewer provides a unique visualization platform optimized for interactive analysis of miRNA deep sequencing data, supporting 84 organisms in the miRBase database.

ACKNOWLEDGEMENTS

We appreciate critical comments and usage feedback from anonymous reviewers and A.T.M. Golam Bari for the program development.

Funding: This work was supported by grants from the National Research Foundation of Korea (NRF-2012M3A9D1054744, NRF-2012M3A9B9036673 and NRF-2011-0019745), GIST

Systems Biology Infrastructure Establishment Grant through ERCSB, and Korea Research Institute of Bioscience and Biotechnology (KRIBB) Research Initiative Program.

Conflict of interest: none declared.

REFERENCES

Cho,S. *et al.* (2013) miRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.*, **41**, D252–D257.

Friedlander,M.R. *et al.* (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.

Humphreys,D.T. *et al.* (2013) miRspring: a compact standalone research tool for analyzing miRNA-seq data. *Nucleic Acids Res.*, **41**, e147.

Kim,J. *et al.* (2014) MAGI: a Node.js web service for fast MicroRNA-Seq Analysis in a GPU infrastructure. *Bioinformatics*, **30**, 2826–2827.

Kim,V.N. *et al.* (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.

Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.

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

Lorenz,R. *et al.* (2011) ViennaRNA Package 2.0. *Algorithm. Mol. Biol.*, **6**, 26.

Sun,Z. *et al.* (2014) CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics*, **15**, 423.