BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 22 2014, pages 3276–3278
doi:10.1093/bioinformatics/btu531

Phylogenetics

Advance Access publication August 5, 2014

AliView: a fast and lightweight alignment viewer and editor for large datasets

Anders Larsson

Systematic Biology, Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala 75236, Sweden

Associate Editor: David Posada

ABSTRACT

Summary: AliView is an alignment viewer and editor designed to meet the requirements of next-generation sequencing era phylogenetic datasets. AliView handles alignments of unlimited size in the formats most commonly used, i.e. FASTA, Phylip, Nexus, Clustal and MSF. The intuitive graphical interface makes it easy to inspect, sort, delete, merge and realign sequences as part of the manual filtering process of large datasets. AliView also works as an easy-to-use alignment editor for small as well as large datasets.

Availability and implementation: AliView is released as open-source software under the GNU General Public License, version 3.0 (GPLv3), and is available at GitHub (www.github.com/AliView). The program is cross-platform and extensively tested on Linux, Mac OS X and Windows systems. Downloads and help are available at http://ormbunkar.se/aliview

Contact: anders.larsson@ebc.uu.se

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on May 28, 2014; revised on July 30, 2014; accepted on August 1, 2014

1 INTRODUCTION

As DNA and protein datasets are getting larger, the demand for a refined and fast alignment editor increases. The need for an improved alignment editor and viewer, therefore, emerged in the 1000 plants project (1KP, www.onekp.com) while designing degenerate primers for a diverse set of ferns from transcriptome data (Rothfels *et al*., 2013).

What was lacking in the previous available programs was the combination of abilities to (i) get an overview of large nucleotide alignments, (ii) visually highlight various conserved regions, (iii) have a simple and intuitive way to align, rearrange, delete and merge sequences and (iv) find degenerate primers in selected semiconserved regions. Although some of these features are individually present in current alignment editors, the combination is not.

In addition to the core functionality meeting these specific needs, AliView (Fig. 1) is designed with a complete set of intuitive general functions meeting the most common demands for preparing a multiple sequence alignment.

Here, AliView is introduced as an alignment viewer and editor with a unique combination of features that allows the user to work with large datasets. The intuitive user interface provides easy visual overview and navigation and works with unlimited size alignments.

2 IMPLEMENTATION

AliView is cross-platform, built in Java and thoroughly tested on Linux, Mac OS X and Windows operating systems. It uses the Java Evolutionary Biology Library v2.0 (available at http://code.google.com/p/jebl2/) for parsing files in Nexus format.

3 FEATURES

3.1 Large alignments, speed and more

The key features of AliView include the ability to swiftly handle large alignments with low memory impact (see Table 1 for comparison with other popular free cross-platform alignment viewers). AliView loads large alignment files 2–14 times faster and demands less than half of the memory resources than comparable alignment editors (Table 1; Supplementary Table S1A–C). AliView will read unlimited size alignment files in FASTA, Phylip, Nexus, Clustal and MSF-format (Table 2). This works through an indexing process where the sequences in the file initially are indexed and only cached in memory when viewed. Aside from the built-in indexing of large files, the program also reads and saves Fasta index files (.fai) as implemented by Samtools (Li *et al*., 2009). The program either reads the whole alignment into memory or leaves parts on file, depending on memory resources available on the specific computer. This way any alignment file can be opened regardless of the memory resources of the computer.

Another important feature of AliView is the speed in rendering large alignments. The speed, together with the mouse wheel zoom feature, makes it possible to get a quick overview and easily navigate in large alignments.

AliView can merge overlapping sequences into a consensus sequence. This feature is useful when working with multiple read NGS-generated sequences. Sometimes the overlap of different sequences or contigs falls outside of the tolerance of assembly programs, and a manually merged sequence is needed.

AliView has unique functionality aimed at supporting the design of universal degenerate primers. It is possible to select an alignment region and have AliView calculate all possible primers (Kämpke *et al*., 2001). To make it easy to select which

Table 1. Time to open alignment file and memory usage. Comparison of AliView with popular free and cross-platform alignment editors

Alignment		Dimension (sequence × character)	Program				
Size	Format		AliView	JalView 2	SeaView	ClustalX	Mesquite
22.4 GB	FASTA	479 726 × 46 512	5–110 s (88 MB)[a,b]	Not supported	Not supported	Not supported	Not supported
22.4 GB	FASTA	479 726 × 46 512	0.6 s (88 MB)[a,c]	Not supported	Not supported	Not supported	Not supported
2.1 GB	FASTA	39 407 × 54 103	17 s (2.2 GB)	73 s (4.7 GB)	51 s (5.7 GB)	Memory error	>10 min
1.3 GB	FASTA	11 792 × 107 401	5.6 s (1.2 GB)	33 s (3.3 GB)	23 s (3.6 GB)	Memory error	>5 min
1.3 GB	PHYLIP	11 799 × 107 401	5.9 s (1.2 GB)	Not supported	17 s (2.7 GB)	Memory error	>5 min
1.3 GB	NEXUS	11 792 × 107 401	5.7 s (1.2 GB)	Not supported	18 s (3.5 GB)	Not supported	>5 min
317 MB	FASTA	361 874 × 4958	2.1 s (608 MB)	31 s (3.1 GB)	9.5 s (3.8 GB)	Memory error	>5 min
42.2 MB	FASTA	5441 × 7682	0.6 s (53 MB)	2.8 s (160 MB)	1.2 s (145 MB)	20 s (1GB)	>5 min

Note: Test results shown were performed on Linux Ubuntu 12.04, Intel Core i7 2700K 3.5 Ghz, 16 GB internal memory and Intel 520 SSD. Similar results were obtained on Mac OS X and Windows systems. For a more extensive comparison including the test methodology, see Supplementary Table S1A–C.
[a]The 22.4 GB FASTA file was not read completely into memory but instead accessed as an indexed file. In all other tests the files were read into memory.
[b]Times depending on how many sequences being indexed at once.
[c]With alignment file already indexed.

Table 2. Comparison of AliView features with popular free and cross-platform alignment editors

Feature / Program	AliView	JalView 2	SeaView	ClustalX	Mesquite
Open alignments of unlimited size (read from disk)	Yes	–	–	–	–
Maximum number of sequences visible at once[a]	Unlimited	495 or overview window	106	120	68
Maximum sequence length visible at once[a]	Unlimited	1830 or overview window	305	345	1650
Merge sequences	Yes	–	–	–	–
Find degenerate primers in selected areas	Yes	–	–	–	–
Define exon boundaries and codon positions for translating nucleotides	Yes	–	–	–	Yes
Highlight difference from consensus or 'trace sequence'	Yes	–	–	Yes	–
Highlight consensus residues	Yes	Yes	Only protein	Only protein	Yes

Note: A more thorough comparison is included as Supplementary Table S2.
[a]Maximum number of sequences and maximum sequence length visible were tested at 1920 × 1200 screen resolution.

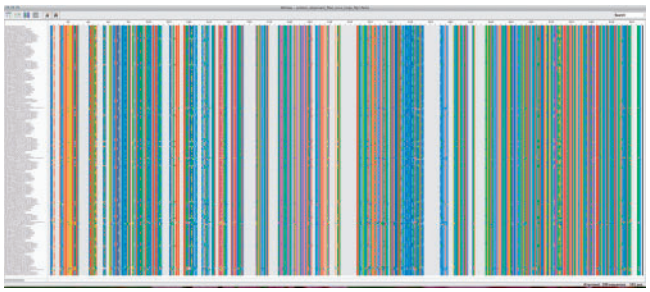

Fig. 1. Alignment zoomed out to give a complete overview of the regions

primer to use, they are presented as an ordered list sorted by the number of degenerate positions, self-binding values and melting temperature.

3.2 Other features

Apart from the key features, AliView also has several other alignment program functions. Alignment can be done by calling any external alignment program. AliView includes and has MUSCLE integrated as the default alignment program (Edgar, 2004), but the user can incorporate other programs if desired. Other features include, for example, manual editing capabilities to insert, delete, change, move or rename sequences in an alignment; undo/redo functionality; several visual cues to highlight consensus characters or characters deviating from the consensus; ClustalX conserved region color scheme (Larkin *et al*., 2007); search functionality that finds patterns across gaps and follows IUPAC codes; implementation of the Nexus specification of Codonpos, Charset and Excludes.

AliView is intended to be a simple easy-to-use alignment editor, and not a complete program for phylogenetic analyses. Instead, the 'external interface' function is aimed to ease the use of AliView as one program in a chain of software, making it possible to call other programs from within AliView with the current alignment or selected sequences as arguments. As a proof of concept, AliView comes with a preset code that adds a button for directing the alignment to FastTree (Price *et al*., 2010) that calculates a phylogenetic tree that is then automatically opened in FigTree (Rambaut, 2012).

For comparison of the key features of AliView with other free cross-platform editors such as Jalview 2 (Waterhouse *et al*., 2009), SeaView (Gouy *et al*., 2010), ClustalX (Larkin *et al*., 2007) and Mesquite (Maddison, and Maddison, 2011) see Table 2. For a more comprehensive comparison of features see Supplementary Table S2.

3.3 User interface and usability

Because an alignment editor is an everyday tool for many researchers, AliView was designed with extensive focus on usability and intuitive handling, implemented by following the logical standards of commonly used software such as text-editors, word processors, browsers and, of course, other alignment viewers.

ACKNOWLEDGEMENTS

Thanks to my colleagues at the Systematic Biology department for immense beta-testing, bug-reporting and mostly good suggestions. Thanks to Allison Perrigo, John Petterson and Martin Ryberg for comments on the manuscript. I also would like to thank the three reviewers and the associate editor David Posada, who gave me very valuable feedback on the previous versions of the manuscript.

Funding: This work was supported by grants from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas) to Petra Korall (2006-429 and 2010-585).

Conflict of interest: none declared.

REFERENCES

Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*., **32**, 1792–1797.

Gouy,M. *et al.* (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol*., **27**, 221–224.

Kämpke,T. *et al.* (2001) Efficient primer design algorithms. *Bioinformatics*, **17**, 214–225.

Larkin,M.A. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

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

Maddison,W.P. and Maddison,D.R. (2011) Mesquite: a modular system for evolutionary analysis. Version 2.75, http://mesquiteproject.org (18 August 2014, date last accessed).

Price,M.N. *et al.* (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.

Rambaut,A. (2012) Figtree 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/ (18 August 2014, date last accessed).

Rothfels,C.J. *et al.* (2013) Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS One*, **8**e76957.

Waterhouse,A.M. *et al.* (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.