The Journal of Open Source Software

alv: a console-based viewer for molecular sequence alignments

Lars Arvestad[1, 2, 3]

1 Department of Mathematics, Stockholm University, Sweden **2** Science for Life Laboratory, Solna, Sweden **3** Swedish e-science Research Centre

DOI: 10.21105/joss.00932

Software

- Review
- Repository
- Archive

Submitted: 06 September 2018
Published: 07 September 2018

Summary

The multiple sequence alignment (MSA) is a common entity in comparative analysis of molecular sequences representing molecules such as DNA, RNA, and proteins. An MSA lines up the sequence building blocks (letters representing nucleotides for DNA/RNA and amino acids for proteins) to form the basis for a hypothesis of how the molecules have evolved, and is computed using, for example, software like Clustal Omega (Sievers and Higgins 2014), MAFFT (Katoh and Standley 2013), MUSCLE (Edgar 2004), MACSE (Ranwez et al. 2011), and hmmalign (Eddy 2015). MSAs have many applications, from advanced analyses such as inferring evolutionary trees (phylogenies) or identifying function in subsequences, to basic use like visual inspection of data. We have written a tool named `alv` to support quick and basic viewing of MSAs (Arvestad 2018).

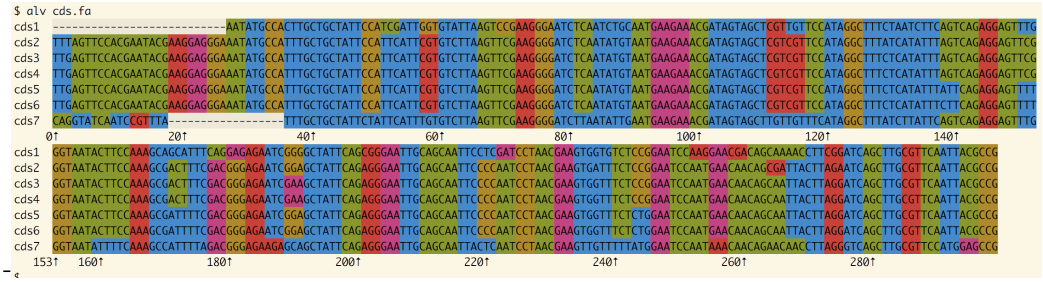
There are a number of MSA viewers available; JalView (Waterhouse et al. 2009), SeaView (Gouy, Guindon, and Gascuel 2009), AliView (Larsson 2014), and MEGA (Kumar et al. 2018) are popular applications with many features, including built-in analysis tools. However, due to their graphical user-interfaces, these programs do not always work well in a command-line based workflow. Web-based MSA viewers are also used, for example NCBI's MSA Viewer (**???**), EBI's MView (**???**), and Wasabi (Veidenberg, Medlar, and Löytynoja 2015). While offering the advantage of not needing local software installation, yet providing analysis features, online tools are inconvenient when working on the command line.

Much simpler tools suffice for quick browsing of MSAs. In fact, alignment formats like PHYLIP and Stockholm are designed to be easily read by both computers and humans, and are easily inspected with common command-line tools (e.g., `less`) or text editors. However, as pure text formats they lack color, which many feel improve visual interpretation of an alignment, and suffer from a fixed layout, which translates to suboptimal use of screen estate.

The `alv` software is an MSA viewer designed to work well in a command-line based environment and the typical invocation is simply `alv msa.fa`. Intended use cases for `alv` includes immediate inspection of a new alignment and quick, scriptable, browsing of many alignments. The viewer is invoked with a straightforward command and has a number of options available. Several MSA formats are recognized automatically (FASTA, Clustal, PHYLIP, Stockholm) and the input sequence type (DNA, RNA, AA, or coding DNA) is guessed by default, but can also be decided when invoking `alv`. The output is written to stdout, with a layout adapted to the size of the current terminal and colored to highlight similarity. For coding DNA, codons are colored according to their amino acid translation (and several genetic codes are supported). Stop codons and frameshifts are

easily identified thanks to a highlighting color scheme. Additional options are available to adapt the MSA output to the user's needs.

We recommend installing `alv` using PyPi: `pip install alv`. Note that `alv` requries Python v3.2 or later.



References

Arvestad, Lars. 2018. "Alv: A Console-Based Alignment Viewer." https://github.com/arvestad/alv

Eddy, Sean. 2015. "Hmmalign: Align Sequences to a Profile Hmm." 2015. hmmer.org.

Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5). Oxford University Press:1792–7. https://doi.org/10.1186/1471-2105-5-113.

Gouy, Manolo, Stéphane Guindon, and Olivier Gascuel. 2009. "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." *Molecular Biology and Evolution* 27 (2). Oxford University Press:221–24. https://doi.org/10.1093/molbev/msp259.

Katoh, Kazutaka, and Daron M Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4). Society for Molecular Biology; Evolution:772–80. https://doi.org/10.1093/molbev/mst010.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis Across Computing Platforms." *Molecular Biology and Evolution* 35 (6). Oxford University Press:1547–9. https://doi.org/10.1093/molbev/msy096.

Larsson, Anders. 2014. "AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets." *Bioinformatics* 30 (22). Oxford University Press:3276–8. https://doi.org/10.1093/bioinformatics/btu531.

Ranwez, Vincent, Sébastien Harispe, Frédéric Delsuc, and Emmanuel JP Douzery. 2011. "MACSE: Multiple Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons." *PloS ONE* 6 (9). Public Library of Science:e22594. https://doi.org/10.1371/journal.pone.0022594.

Sievers, Fabian, and Desmond G Higgins. 2014. "Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences." In *Multiple Sequence Alignment Methods*, 105–16. Springer. https://doi.org/10.1002/0471250953.bi0313s48.

Veidenberg, Andres, Alan Medlar, and Ari Löytynoja. 2015. "Wasabi: An Integrated Platform for Evolutionary Sequence Analysis and Data Visualization." *Molecular Biology and Evolution* 33 (4). Oxford University Press:1126–30. https://doi.org/10.1093/molbev/msv333.

Waterhouse, Andrew M, James B Procter, David MA Martin, Michèle Clamp, and Geoffrey J Barton. 2009. "Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench." *Bioinformatics* 25 (9). Oxford University Press:1189–91. https://doi.org/10.1093/bioinformatics/btp033.