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

Summary

The multiple sequence alignment (MSA) is a common entity in comparative analysis of 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 & Higgins, 2014), MAFFT (Katoh & Standley, 2013), MUSCLE (Edgar, 2004), MACSE (Ranwez, Harispe, Delsuc, & Douzery, 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, Procter, Martin, Clamp, & Barton, 2009), SeaView (Gouy, Guindon, & Gascuel, 2009), AliView (Larsson, 2014), and MEGA (Kumar, Stecher, Li, Knyaz, & Tamura, 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 ("NCBI multiple sequence alignment viewer," n.d.), EBI's MView ("MView," n.d.), and Wasabi (Veidenberg, Medlar, & 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, L. (2018). Alv: A console-based alignment viewer. https://github.com/arvestad/alv.

Eddy, S. (2015). hmmalign: Align sequences to a profile HMM. Retrieved from hmmer.org

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), 1792–1797. doi:10.1186/1471-2105-5-113

Gouy, M., Guindon, S., & Gascuel, O. (2009). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, *27*(2), 221–224. doi:10.1093/molbev/msp259

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772–780. doi:10.1093/molbev/mst010

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA x: Molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), 1547–1549. doi:10.1093/molbev/msy096

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. doi:10.1093/bioinformatics/btu531

MView. (n.d.). EBI. Retrieved from https://www.ebi.ac.uk/Tools/msa/mview

NCBI multiple sequence alignment viewer. (n.d.). NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/projects/msaviewer/

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. (2011). MACSE: Multiple alignment of coding sequences accounting for frameshifts and stop codons. *PloS ONE*, *6*(9), e22594. doi:10.1371/journal.pone.0022594

Sievers, F., & Higgins, D. G. (2014). Clustal omega, accurate alignment of very large numbers of sequences. In *Multiple sequence alignment methods* (pp. 105–116). Springer. doi:10.1002/0471250953.bi0313s48

Veidenberg, A., Medlar, A., & Löytynoja, A. (2015). Wasabi: An integrated platform for evolutionary sequence analysis and data visualization. *Molecular biology and evolution*, *33*(4), 1126–1130. doi:10.1093/molbev/msv333

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. doi:10.1093/bioinformatics/btp033