Lexicon-Mono-Seq, DOM Text Based Async MSA Viewer

Ibrahim Tanyalcin[1]

1 Vrije Universiteit Brussel

Summary

In Bioinformatics, one of the most common problems is depiction of long amino-acid / DNA sequences and their comparison (Li et al., 2018). Routinely, these sequences are aligned to a reference and the data is stored in common formats such as clustal (Larkin et al., 2007). These alignments are called Multiple Sequence Alignments (MSA) and can be viewed with a variety of different tools.

Lexicon-Mono-Seq is an MSA viewer specially designed for browsers that use native DOM Text + SVG (Chojnacki, Cowley, Lee, Foix, & Lopez, 2017; Jaschob, Davis, & Riffle, 2014) instead of HTML5 canvas (Robinson et al., 2011; Yachdav et al., 2016) or CLI (Beraldi, 2017).

Several MSA viewers rely on rendering characters as an image using the *canvas* element. Although reliable and convenient, this can result in problems such as higher memory usage and blurred text (increasing resolution increases memory usage).

Lexicon-Mono-Seq aims to approach the alignment viewing problem from the DOM (Document Object Model) perspective, each character is a combination of a textNode and an SVG Graphics Element. The script recycles these nodes and paints only regions that are necessary (Tanyalcin et al., 2018). Comparison of sequences with different character widths is also a possibility, this allows one to compare amino acids and DNA regions within the same applet (i.e. one amino acid every three nucleotides in the DNA). *Lexicon-Mono-Seq* is agnostic about the character width of sequences, a dataset can have multiple sequences with different character widths.

At its core *Lexicon-Mono-Seq* is a monospace font ASCII character animation framework and comes with its own pre-baked easing functions, much like CSS Cubic Bezier functions, which provide customizable and cached timing points for animations. Other fields such as creative coding or data visualization can use *Lexicon-Mono-Seq* for transitioning between ASCII maps or large heatmaps.

With minimal technical debt (no dependencies), the library uses requestAnimationFrame (rAF) to asynchronously render blocks of sequences with the desired animation, duration and node limit. The library is about 50kB (as of version 0.19.0) and does not need any transpiling to use in older browsers that do not support ES6 (Ecma International, 2018). The amount of sequences is limited to your browser's capacity, the recommended limit is 1000 sequences with 100000 letters each for commodity laptops/mobile devices.

Lexicon-Mono-Seq is suitable for developers who only want the functionality of a minimal MSA Viewer but desire to design their own interface around it. The script provides several utility functions that can be hooked to custom buttons as desired (GitHub, 2018; Npm, 2018).

References

Beraldi, D. (2017). ASCIIGenome: a command line genome browser for console terminals. *Bioinformatics*, *33*(10), 1568–1569. doi:10.1093/bioinformatics/btx007

Chojnacki, S., Cowley, A., Lee, J., Foix, A., & Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Research*, *45*(W1), W550—W553. doi:10.1093/nar/gkx273

Ecma International. (2018). Ecma tc39. Retrieved February 19, 2018, from https://github.com/tc39

GitHub. (2018). GitHub.com. Retrieved February 19, 2018, from https://github.com

Jaschob, D., Davis, T. N., & Riffle, M. (2014). SnipViz: A compact and lightweight web site widget for display and dissemination of multiple versions of gene and protein sequences. *BMC Research Notes*, *7*, 468. doi:10.1186/1756-0500-7-468

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. doi:10.1093/bioinformatics/btm404

Li, M., Long, H., Wang, H., Fu, H., Xu, D., Shen, Y., Yao, Y., et al. (2018). Disease sequences high-accuracy alignment based on the precision medicine. *Biomedical Research International*, *2018*, 1718046. doi:10.1155/2018/1718046

Npm. (2018). Npmjs.com. Retrieved February 19, 2018, from https://www.npmjs.com/

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26. doi:10.1038/nbt.1754

Tanyalcin, I., Al Assaf, C., Ferté, J., Ancien, F., Khan, T., Smits, G., Rooman, M., et al. (2018). Lexicon visualization library and JavaScript for scientific data visualization. *IEEE Computing in Science and Engineering*, *20*(1), 50–65. doi:10.1109/MCSE.2018.011111125

Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., Lewis, S. E., et al. (2016). MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics*, *32*(22), 3501–3503. doi:10.1093/bioinformatics/btw474