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

Ibrahim Tanyalcin¹

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


