

DotPlotly: Dot plot for Sequence Similarity Visualization

Aadam
Ghulam Ishaq Khan Institute
Topi, Pakistan
Email: aadimotor@gmail.com

Obaidullah
Ghulam Ishaq Khan Institute
Topi, Pakistan
Email: obaidul024@gmail.com

Abstract—Graphical representation of data has a great impact on the comprehension of ideas and notions presented. They can be used to effectively communicate correlation between multiple sets of data. DotPlotly is a standalone cross-platform application that can generate interactive dot plots to show the similarity between two sequences. DotPlotly is an easy-to-use open-source software package developed in Python. The source code is available at <https://github.com/aadimotor/bioinformatics-project>.

I. INTRODUCTION

Dotplots, first described five decades ago [1], still are a great tool for visualizing complex or repetitive sequences and identifying the similarities between sequences. They are widely used in the field of Bioinformatics because of their ease-of-use and understanding in presenting insertions, deletions, inversions, or repeats in a given pair of sequences.

In recent years, a number of libraries and tools have been introduced to easily create visualizations, especially dot plots, to better understand the similarity between two or more sequences. **FlexiDot** [2] is an open-source cross-platform suite that could generate self, pairwise, and all-against-all visualizations and can also handle the ambiguity as well. **D-Genesis** [3] and **iDotter** [4] are some other open-source tools that are used to easily create interactive dot plot visualizations.

In this report we present a generic sequence alignment algorithm to calculate the dotplot representation between two sequences of text, thus, revealing primary similarities between the two sequences.

II. IMPLEMENTATION

DotPlotly is a cross-platform tool that can handle pairwise visual sequence comparisons and highlights the similarity, if any, between two input sequences.

Each character in one the sequences is looked up in the paired sequence. If any instance is found, its index number is noted and saved in a list of matches for that specific character. The steps are repeated for every character in the sequences and the index number if any is noted and saved as mentioned above. The list is nested within a single larger list of matched characters and is augmented with any other characters matched. Each character is then matched against its occurrences in the second sequence and its corresponding position plotted in the dotplot graphing tool, as can be seen

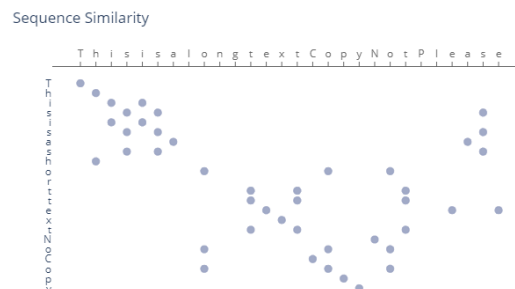


Fig. 1: Dot plot for a string of text

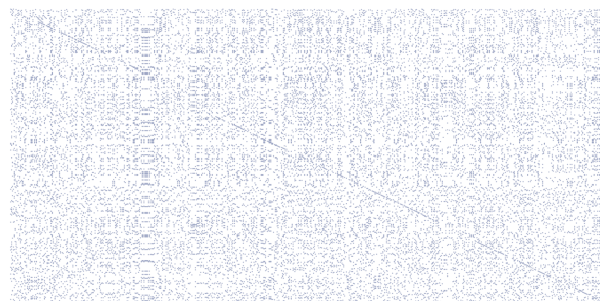


Fig. 2: Dot plots for ENSP00000342858 and ENSP00000339950 protein sequences

in Fig. 2. This process is repeated for every character in the first sequence against its matched occurrence.

The plotting library used for this step is Plotly [5]. The only package dependency hence for this tool is the aforementioned.

III. CONCLUSION

This tool reveals rudimentary but pertinent similarities between two sequences of text and represents the likelihood of them being same in a visual fashion, thus providing the users with visual confirmation of salient features of the given sequences.

ACKNOWLEDGMENT

The authors would like to thank Google Scholar, Wikipedia, creators of Python, Plotly and Jupyter projects, and the general scientific and development community.

REFERENCES

- [1] A. J. Gibbs and G. A. McIntyre, "The Diagram, a Method for Comparing Sequences: Its Use with Amino Acid and Nucleotide Sequences," *European Journal of Biochemistry*, vol. 16, no. 1, pp. 1–11, 1970.
- [2] K. M. Seibt, T. Schmidt, and T. Heitkam, "FlexiDot: Highly customizable, ambiguity-aware dotplots for visual sequence analyses," *Bioinformatics*, 2018.
- [3] F. Cabanettes and C. Klopp, "D-GENIES: Dot plot large genomes in an interactive, efficient and simple way," *PeerJ*, vol. 2018, no. 6, p. e4958, 6 2018.
- [4] D. Gerighausen, A. Hausdorf, S. Zänker, and D. Zeckzer, "iDotter - An interactive dot plot viewer," *Computer Science Research Notes*, vol. 2701, no. May, pp. 117–124, 2017.
- [5] "Plotly: The front-end for ML and data science models." [Online]. Available: <https://plotly.com/>