BIOINFORMATICS

dnaplotlib: visualization of genetic constructs, libraries and associated data

Thomas E. Gorochowski[1,†], Bryan Der[1,†], Emerson Glassey[1,†], D. Benjamin Gordon[1] and Christopher A. Voigt[1,*]

[1]Department of Biological Engineering, Synthetic Biology Center, Massachusetts Institute of Technology, USA.
[†]These authors contributed equally to this work.

Associate Editor: XXXXXXX

ABSTRACT

Summary: dnaplotlib is a computational toolkit that enables highly customizable visualization of individual genetic constructs and libraries of design variants. Publication quality vector-based output is produced and all aspects of the rendering process can be easily customized or replaced by the user. dnaplotlib is capable of SBOL Visual compliant diagrams in addition to a format able to better illustrate the precise location and length of each genetic part. This alternative visualization method enables direct comparison with nucleotide-level information such as RNA-seq read depth. While it is envisaged that access will be predominantly via the programming interface, a web interface is also provided to support broader usage.

Availability: dnaplotlib is cross-platform and open-source software developed using Python and released under the OSI recognized NPOSL-3.0 licence. Source code, documentation and a web front-end are available at the project website: http://www.dnaplotlib.org.

Contact: cavoigt@gmail.com

Engineering disciplines rely on the use of standards to clearly communicate design information and allow for the construction of large and complex systems. As synthetic biology

Attempts have been made to visualize conceptual genetic designs To date limited attempts have been made to aid in the visualization of conceptual genetic designs (Bhatia & Densmore, 2013) and although many DNA design tools offer the ability to . Firstly, output is generally The biggest limitation of these approaches so far though is an inability to integrate with existing scientific computing visualization and analysis tools. This is essential as

To address these needs dnaplotlib has been built using matplotlib (Hunter, 2007), a highly portable Python-based 2D graphics library that allows for graphical output in the form of vector based PDFs or rasterized PNG or JPEG images. A major advantage oFurthermore, it enables the rendering of genetic designs to be easily integrated into existing analysis scripts. Python was chosen due to its increasing use for the analysis of biological data (Cock *et al.*, 2009) and its availability across all major operating systems.

A key consideration in the design of dnaplotlib was ensuring all aspects of the rendering process could be customized to user

requirements. As fields such as synthetic biology are still evolving, the the precise way in which such tools will be used and the discovery of new types of genetic part that need to be included, but may not yet have a standardized representation. Enabling custom elements to be easily added and refined over time will ensure such elements are captured and also contribute to the standardization process.

To generate a

In addition to direct library access through Python analysis scripts, we also provide two scripts to enable input in the form of text files

To ensure broadest application of these tools that included use by non-programmers, we also developed a web-based interface that allows text files describing part styles and DNA designs to be uploaded and processed. These files uploaded to the server, processed by the scripts described previously, and generated images are returned to the user. This ensures broadest use of the library and ensures that members across a lab can share the same styling of diagrams to ensure clear communication of their designs.

We see a key feature of ensuring adoption as

As synthetic biology moves ever closer to automated design procedures that harness the potential to construct huge libraries of design variants (Smanski *et al.*, 2014; Bilitchenko *et al.*, 2011), supporting visualization tools will become important to ensure clear communication of these large and diverse designs. The SBOL Visual initiative was started to aid in this task with the definition of agreed pictorial representations of genetic elements (Quinn *et al.*, 2013). However, so far it has seen limited uptake due to a lack a tools that are able to be easily integrated into existing analysis work-flows. dnaplotlib fills this gap by providing a computational toolkit that both adheres to this standard while also enabling highly customizable visualizations that are most appropriate for the specific needs of the laboratory. It is hoped that by providing researchers with this flexibility that they will be more willing to apply such standards to their work, improving general uptake of SBOL Visual across the field.

dnaplotlib is under continual development with a current focus on broadening the types of genetic element covered to include new synthetic biological parts. The project welcomes contributions from others within the community through the project website and public development repository: http://www.dnaplotlib.org.
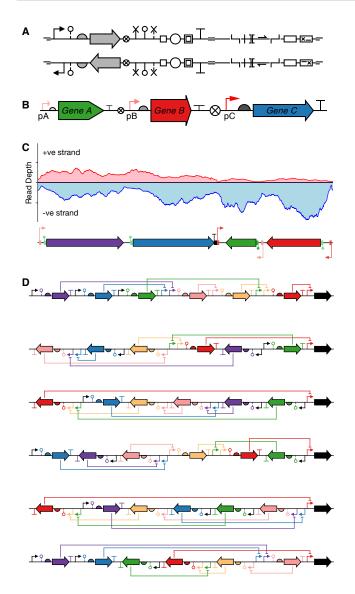
*to whom correspondence should be addressed

REFERENCES

Bhatia, S. and Densmore, D. (2013). Pigeon: A Design Visualizer for Synthetic Biology, *ACS Synth. Biol.*, **2**, 348-350.

Smanski, M.J., Swapnil, B., Park, YJ., Zhao, D., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Gordon, D.B., Densmore, D. and Voigt, C.A. (2014) Combinatorial design and assembly of refactored gene clusters, *Nat. Biotech.*, **?**, ???-???.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment, *Computing in Science & Engineering*, **9**, 90-95.

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, and de Hoon MJ. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422-1422.

Bilitchenko, L., Liu, A., Cheung, S., Weeding, E., Xia, B., Leguia, M., Anderson, J.C., Densmore, D. (2011) Eugene A Domain Specific Language for Specifying and Constraining Synthetic Biological Parts, Devices, and Systems. *PLoS ONE*, **6**, e18882.

Quinn, J., Beal, J., Bhatia, S., Cai, P., Chen, J., Clancy, K., Hillson, N., Galdzicki, M., Maheshwari, A.P., Umesh; P., Matthew; R.C.; Stan, G.-B., Endy, D. (2013) "Synthetic Biology Open Language Visual (SBOL Visual), version 1.0.0."

Fig. 1. Overview of the core dnaplotlib functionality. (**A**) The full set of SBOL Visual parts is available in both forward and reverse orientations. (**B**) The size, color and shape of all elements be easily customised by the user. (**C**) To allow for direct comparison between a genetic design and nucleotide-level data such as RNA-seq read depths, trace-based renderers are provided. These use indicators of actual part widths on the design backbone (CDS length and small filled rectangles) to enable a visual alignment of design information with trace data. (**D**) An example rendering of a library of genetic design variants implementing the same 3-input (black promoters), 1-output (black CDS) device. Colors have been used to link repressor genes and their cognate promoters.