EasyChair Preprint

№ 9044

Steps Towards Functional Synthetic Biology

Ibrahim Aldulijan, Jacob Beal, Sonja Billerbeck, Jeff Bouffard, Gaël Chambonnier, Nikolaos Delkis, Isaac Guerreiro, Martin Holub Martin Holub, Daisuke Kiga, Jacky Loo, Paul Ross, Vinoo Selvarajah, Noah Sprent, Gonzalo Vidal and Alejandro Vignoni

October 11, 2022

Steps Towards Functional Synthetic Biology

[1]Ibrahim Aldulijan [2,*]Jacob Beal [3]Sonja Billerbeck [4]Jeff Bouffard [5]Gaël Chambonnier [6]Nikolaos Delkis [7]Isaac Guerreiro [8]Martin Holub [9]Daisuke Kiga [10]Jacky Loo [11]Paul Ross [7]Vinoo Selvarajah [12]Noah Sprent [13]Gonzalo Vidal [14]Alejandro Vignoni

[1]Stevens Institute of Technology, [2]Raytheon BBN, [3]University of Groningen, [4]Concordia University, [5]Massachusetts Institute of Technology, [6]University of Thessaly, [7]iGEM Foundation, [8]Delft University of Technology, [9]Waseda University, [10]Aalto University, [11]BioStrat Marketing, [12]Imperial College London, [13]Newcastle University, [14]Universitat Politecnica de Valencia

*Corresponding author: jakebeal@ieee.org

1 INTRODUCTION

While synthetic biology has made great progress in methods for modular assembly of genetic sequences and in engineering biological systems with a wide variety of functions, current paradigms entangle sequence and functionality in a manner that makes abstraction difficult, reduces engineering flexibility, and impairs predictability and design reuse. Functional Synthetic Biology [1] proposes a roadmap to overcome these limits by focusing on behavior descriptions, predictability, flexibility, and risk reduction, so synthetic biologists can more effectively share successes and avoid failures.

The iGEM community, like other synthetic biology communities, faces challenges in effective sharing and reuse of biological devices. These are particularly acute for iGEM, since iGEM teams need to execute projects in only a few months and many team members have little prior experience. At the same time, barriers for adoption are lowered by the culture of openness, sharing, and reuse that is encouraged by iGEM. For these reasons, the iGEM Engineering Committee has been working to implement the early phases of the Functional Synthetic Biology roadmap in the context of iGEM's annual DNA distribution.

2 AGILE CURATION OF DNA DESIGN PACKAGES

As a first step, we have deployed an agile data curation workflow for community development of DNA design packages, leveraging distributed version control and continuous integration tooling. Each year, iGEM sends teams a distribution of DNA parts expected to be useful for their projects. For the 2022 season, iGEM developed an all new distribution, enlisting a larger community to aid in its design. To support the community design process, we built on work from the DARPA SD2 program [6] to deploy an agile data curation workflow on GitHub (Figure 1). With this workflow, contributors submit DNA design packages developed with spreadsheets and design files. These undergo community review and revision using the Gitflow workflow, while complementary automation tests packages for errors and collates package contents to produce a distribution plan and synthesis orders.

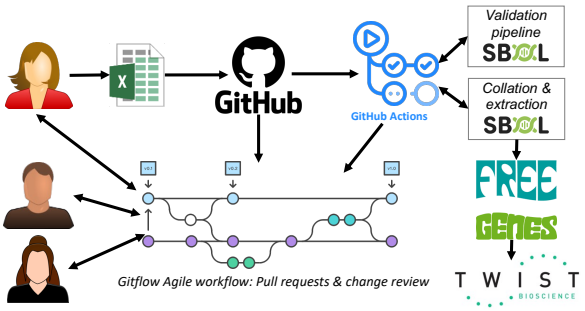

Figure 1: iGEM distribution agile data curation workflow.

Workflow automation is implemented using GitHub Actions, an integrated continuous integration and continuous delivery (CI/CD) framework. In our usage, continuous integration maps to checking specifications for coherence and correctness, while continuous deployment maps to compilation of all designs together into a complete plan for the distribution and the synthesis orders for building it.

Figure 2 shows details of this workflow. Excel templates provide a user-friendly interface to specify "packages" organizing groups of related parts (e.g., a collection of fluorescent reporters), and the build plans for how to combine part sequences into composites, flank them with prefixes and suffixes for BioBricks or Type IIS assembly, and insert them into plasmids for propagation and dissemination.

The workflow first exports Excel into two formats: CSV for git diff review, and SBOL3 [4] that specifies the parts (SBOL Components) and combinatorial build plans (SBOL CombinatorialDerviations). Parts are either fetched from public data sources by their identifiers (e.g., NCBI accession, BioBrick part number) or imported from files in the same directory as the sheet. The build plan is then validated to ensure it is coherent and fully specified. After validation, build plans are compiled to a full specification for each package. Each CombinatorialDerivation is expanded into a list of specific composite parts to produce, sequences are calculated for each construct, and a human-readable README file is generated summarizing the package and its contents. Finally, all packages are collated to produce the complete distribution,

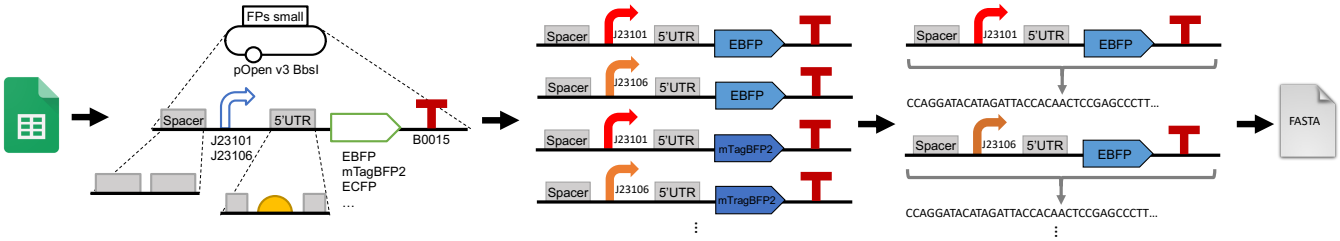

Figure 2: Production of synthesis orders from DNA package plans, as used by the iGEM Distribution repository: packages are specified in Excel sheets, from which are extracted SBOL3 documents specifying libraries of plasmid build plans. Each build plan is expanded into a list of all of the specific composite parts to produce. A sequence is then calculated for each construct, and the portion of each construct to be synthesized is exported to FASTA for placing a synthesis order.

and the SBOL is exported to GenBank for compatibility with other design tools, and to FASTA for ordering the plasmid inserts that are to be synthesized.

This workflow was able to be used effectively by the iGEM Engineering Committee in developing the iGEM 2022 distribution (available at https://github.com/iGEM-Engineering/iGEM-distribution), supporting a rapid pace of development and review by a large group of contributors. During the main development period of the distribution, from January 1st to February 16th, 2022, 15 contributors at 11 institutions in 8 different countries produced 571 commits, which were reviewed and merged in 87 pull requests, an average of nearly 2 contributions per day. The resulting distribution contains 16 packages organizing several hundred parts into thematic collections such as "CRISPR-Cas", "Fluorescent Reporters", "Small Molecule Inducers", and "Plant Parts." Critically, the learning curve also proved reasonable: most contributors were not programmers, and many had never used git before.

3 WORK IN PROGRESS

We are continuing to work towards the Functional Synthetic Biology vision, building on the lessons from the 2022 distribution. First, automated validation is being extended to include biological considerations, using pydna [5] to check assembly compatibility and using synthesis company APIs to check synthesizability. We have also been improving biologist-focused documentation for package development and use of git-based workflows, to support adoption of these methods by iGEM teams and the larger synthetic biology community.

Next, we are implementing a dependency management system for DNA design packages based on SBOL Enhancement Proposal (SEP) 054 (available at https://github.com/SynBioDex/SEPs). Analogous to software package management systems, this will allow DNA design packages to be broken out into their own repositories and maintained separately, then imported for use in the distribution or other packages. This is required for scalability to a large community and to minimize duplication and forking of materials.

Finally, we are running interlaboratory studies to develop reliable transcriptional toolkits. Prior work shows transcriptional regulators can be effectively insulated from genetic context (e.g., [2, 3]), but these results are not readily accessible or joined with predictive models. The committee is thus running studies to produce models quantifying insulated systems in replicable ERF/cell units. The first targets are constitutive promoters (for consistent expression levels) and fluorescent reporters (for debugging and quantification), to be followed by inducible promoters (for adjustable regulation and sensing). If successful, these will be collected in packages for distribution, making it simple for iGEM teams and other users to test new devices with known-reliable sensors, adjustable inputs, and reporters.

4 ACKNOWLEDGEMENTS

In addition to the listed authors, many other members of the iGEM Engineering Committee have made contributions that have supported the results described. This work was partially supported by AFRL/DARPA contract FA8750-17-C-0184, Concordia and Newcastle Universities, ERC Advanced Grant 883684, BBSRC award BB/M011178/1, Grant MINECO/AEI, EU DPI2017-82896-C2-1-R and MCIN/AEI/10.13039/501100011033 grant PID2020-117271RB-C21. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations. Views, opinions, and/or findings are should not be interpreted as representing the official views or policies of any of the funders.

REFERENCES

[1] Aldulijan, I., et al. Functional synthetic biology. *arXiv preprint arXiv:2207.00538* (2022).

[2] Carr, S. B., Beal, J., and Densmore, D. M. Reducing dna context dependence in bacterial promoters. *PloS one 12*, 4 (2017), e0176013.

[3] Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature biotechnology 30*, 11 (2012), 1137–1142.

[4] McLaughlin, J., et al. The synthetic biology open language (SBOL) version 3: simplified data exchange for bioengineering. *Frontiers in Bioengineering and Biotechnology* (2020), 1009.

[5] Pereira, F., Azevedo, F., Carvalho, Â., Ribeiro, G. F., Budde, M. W., and Johansson, B. Pydna: a simulation and documentation tool for dna assembly strategies using python. *BMC bioinf. 16*, 1 (2015), 1–10.

[6] Roehner, N., et al. Data representation in the DARPA SD2 program. In *IWBDA* (2021).