Synthetic Biology Open Language

Chris J. Myers

University of Utah

COMBINE 2018

October 8, 2018

Reproducibility Crisis

Professor David Donoho
Stanford University

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.

Essential information for synthetic DNA sequences

To the Editor:

Following a discussion by the workgroup for Data Standards in Synthetic Biology, which met in June 2010 during the Second Workshop on Biodesign Automation in Anaheim, California, we wish to highlight a problem relating to the reproducibility of the synthetic biology literature. In particular, we have noted the very small number of articles reporting synthetic gene networks that disclose the complete sequence of all the constructs they describe.

To our knowledge, there are only a few examples where full sequences have been released. In 2005, a patent application[1] disclosed the sequences of the toggle switches published four years earlier in a paper by Gardner *et al.*[2]. The same year, Basu *et al.*[3] deposited their construct sequences for programmed pattern formation into GenBank[3]. Examples of synthetic DNA sequences derived from standardized parts that have been made available in GenBank include the refactored genome of the bacteriophage

gaps between key components are almost never reported, presumably because they are not considered crucial to the report. Yet, synthetic biology relies on the premise that synthetic DNA can be engineered with base-level precision.

Missing sequence information in papers hurts reproducibility, limits reuse of past work and incorrectly assumes that we know fully which sequence segments are important. For example, many synthetic biologists are currently realizing that translation initiation rates are dependent on more than the Shine-Dalgarno sequence[8]. Sequences upstream of the start codon are crucial for translation rates, yet are underreported. Similarly, it has been demonstrated that intron length can affect the dynamics of genetic oscillators[9]. Many more such examples are likely to emerge.

Because full sequence disclosure is critical, we wonder why the common requirement by many journals to provide GenBank entries for genomes and natural sequences has and welcome contributions from the greater community.

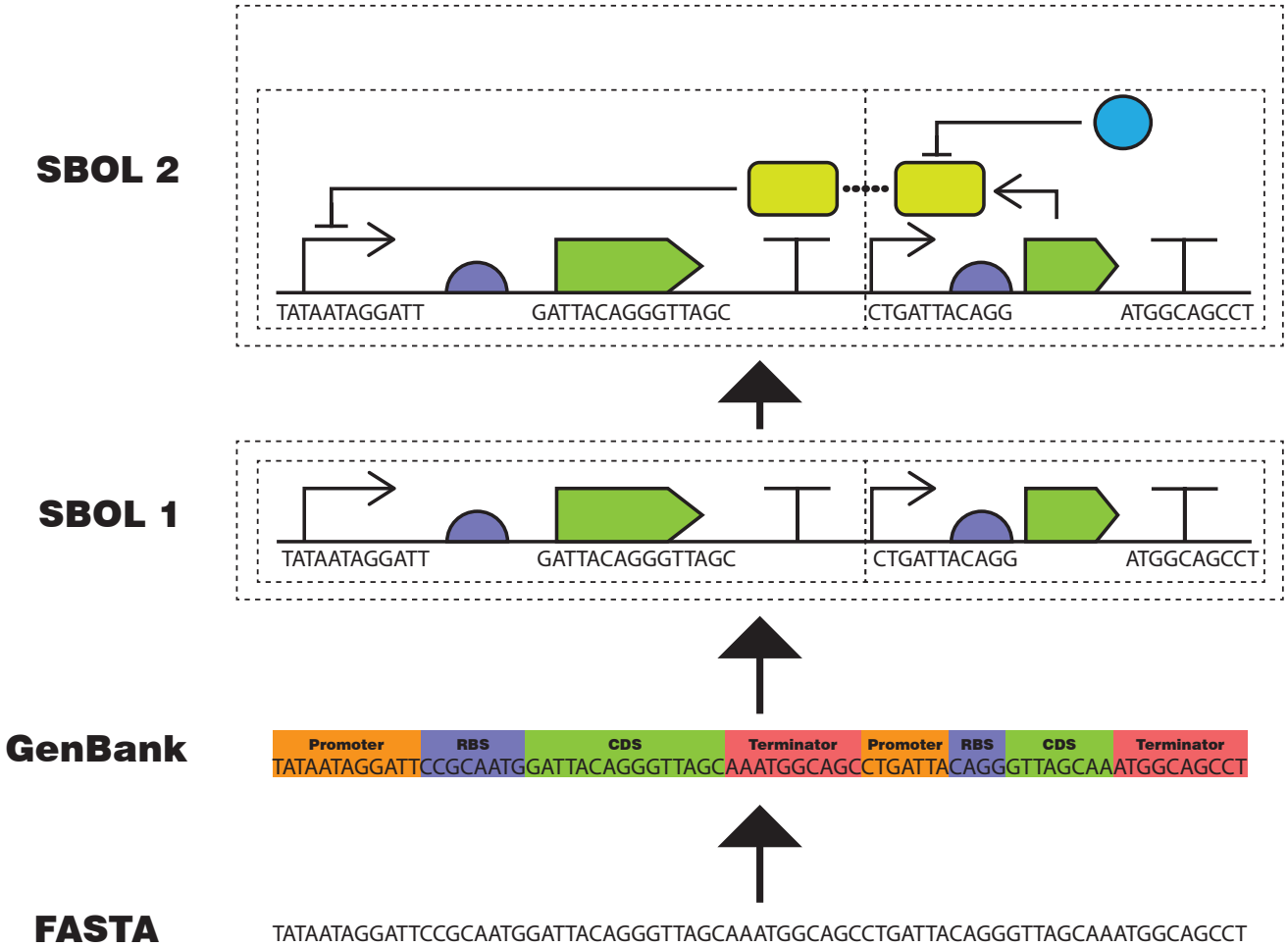
COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Jean Peccoud[1], J Christopher Anderson[2], Deepak Chandran[3], Douglas Densmore[4], Michal Galdzicki[5], Matthew W Lux[1], Cesar A Rodriguez[6], Guy-Bart Stan[7] & Herbert M Sauro[3]

[1]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA. [2]Department of Bioengineering, QB3: California Institute for Quantitative Biological Research, University of California, Berkeley, California, USA. [3]Department of Bioengineering, University of Washington, Seattle, Washington, USA. [4]Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts, USA. [5]Biomedical and Health Informatics, University of Washington, Seattle, Washington, USA. [6]BIOFAB, Emeryville, California, USA. [7]Department of Bioengineering and Centre for Synthetic Biology and Innovation, Imperial College London, London, UK.
e-mail: peccoud@vt.edu

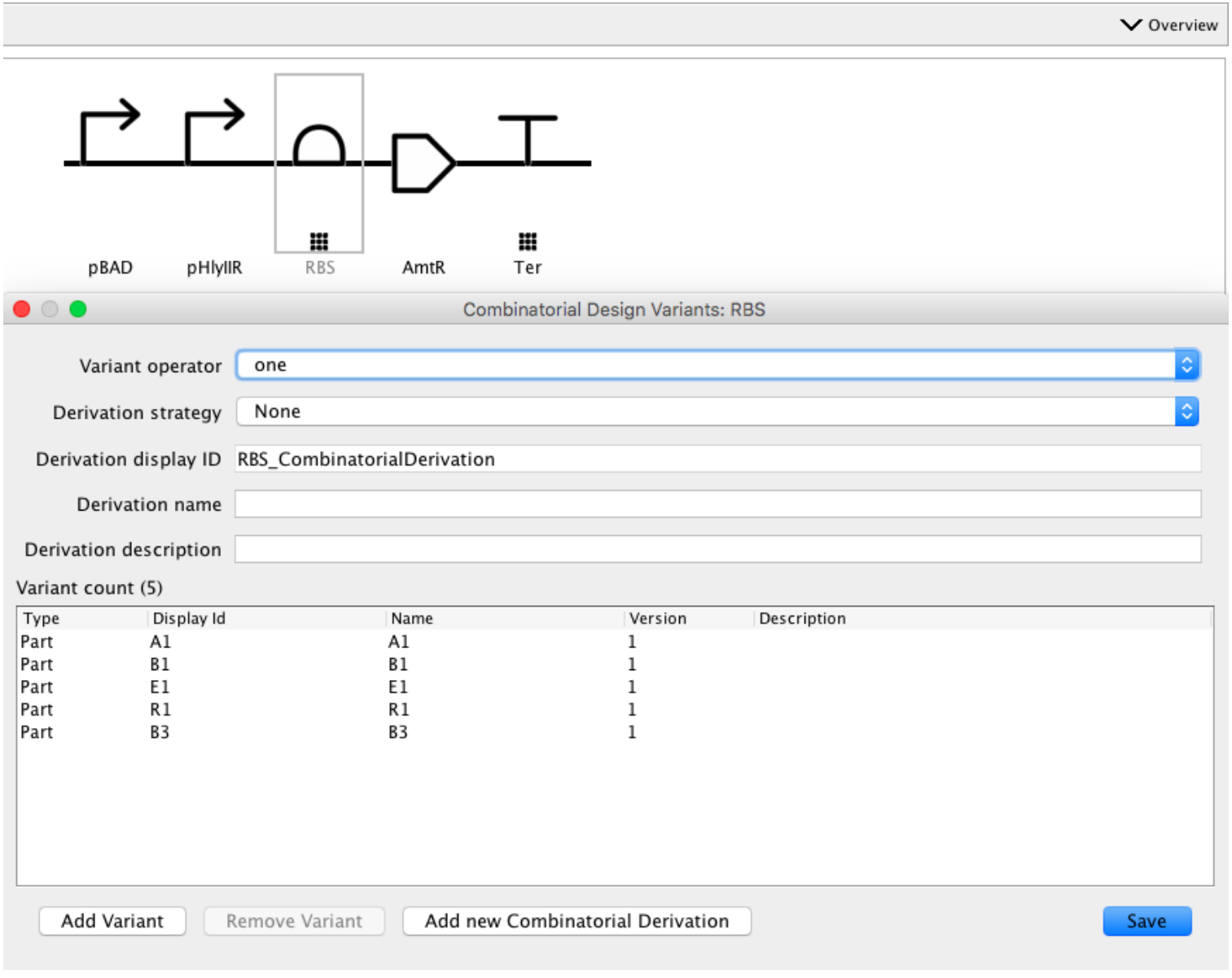
1. Gardner, T.S. & Collins, J.J. US patent 6,841,376 (2005).
2. Gardner, T.S., Cantor, C.R. & Collins, J.J. *Nature* **403**, 339–342 (2000).
3. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. *Nature* **434**, 1130–1134 (2005).

Synthetic Biology Open Language (SBOL)

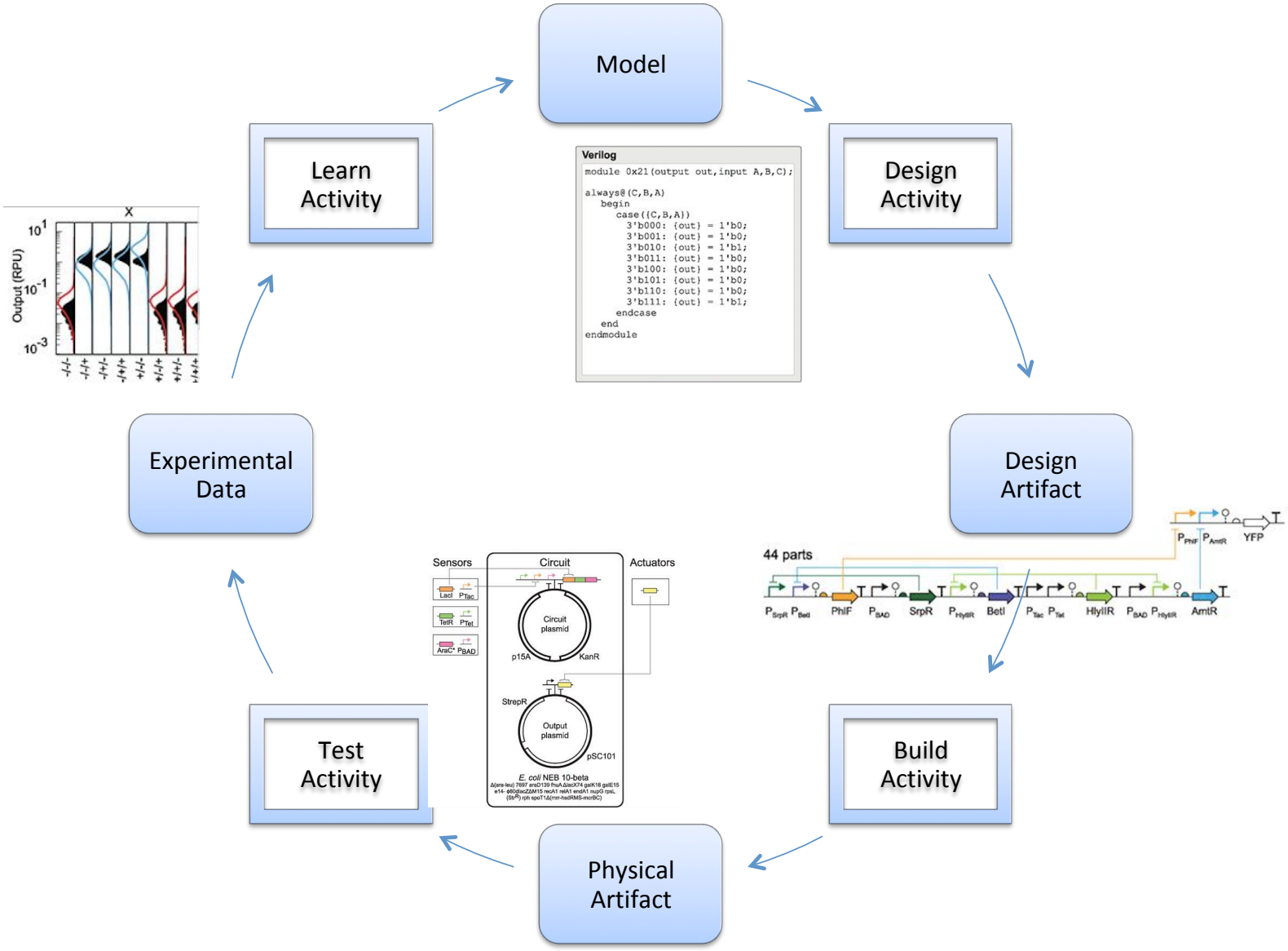


Galdzicki et al., *Nature Biotechnology* (2014)
Quinn et al., *PLoS Biology* (2015)
Roehner et al., *ACS Synthetic Biology* (2016)

SBOL Version 2.2: Combinatorial Derivations



SBOL Version 2.2: Design-Build-Test-Learn (Prov-O)



Updates from Harmony

- Completion of SBOL 2.2.1 - small updates and revised validation rules.
- SBOL Visual 2.1 discussions:
  - SEP V013 - Multi-source / multi-sink arrows
  - SEP V014 - Modules and MapsTo
- SBOL 2.3 discussions:
  - SEP 013 - Sequence insertion and deletion
  - SEP 021 - Experiments and Experimental Data
  - SEP 024 - Best practices for host context / ontologies for experiments
  - SEP 026 - Add a link from Location to Sequence
  - SEP 027 - Adding a type field to Activity class
- SBOL 3.0 discussions begun

Goals for COMBINE

- Tuesday 10:30 - noon : general community discussion (outreach, industrial releations, plans for SBOL 2.3, SBOLv 2.1, and SBOL3, etc.)
- Tuesday 1:30 - 3pm : Specifying parameters in SBOL
- Tuesday 3:30 - 5pm : Representing strains with SBOL
- Remainder of breakouts dedicated to work on SBOL 2.3 and SBOLv 2.1 specification, various papers, and grant proposals.