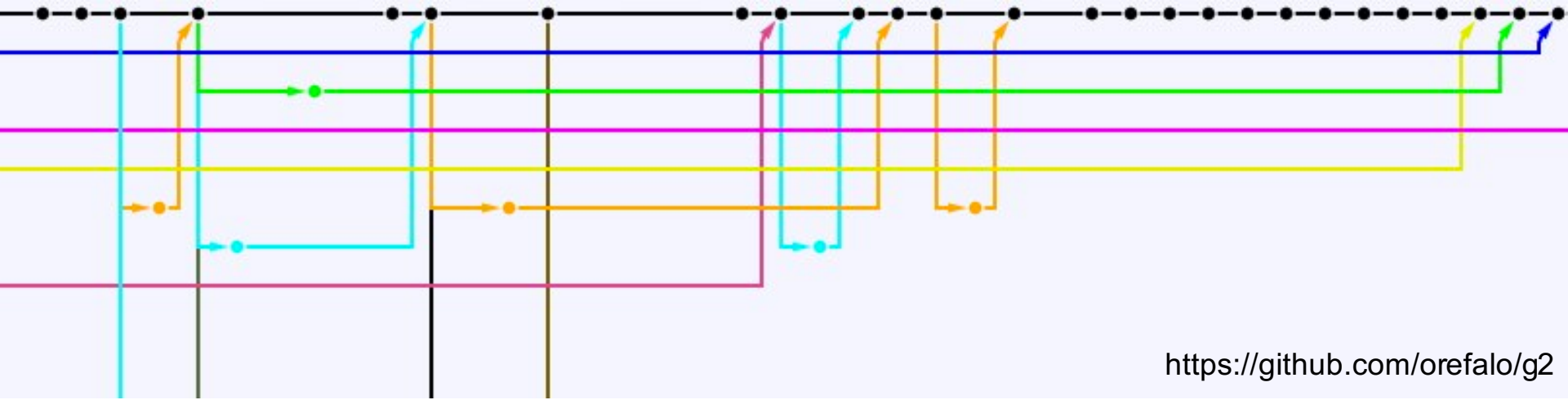


Version & Variant Control for Synthetic Biology

Bryan Bartley
bartleyba@sbolstandard.org
The Sauro Systems & Synthetic Biology Lab
University of Washington, Seattle, WA

Molecular Systems Biology 7; Article number 512; doi:10.1038/msb.2011.41
Citation: *Molecular Systems Biology* 7:512

www.molecularsystemsbiology.com

molecular systems biology

PERSPECTIVE

Retrieval, alignment, and clustering of computational models based on semantic annotations

Marvin Schulz[1], Falko Krause[1], Nicolas Le Novère[2], Edda Klipp[1,*] and Wolfram Liebermeister[1,3]

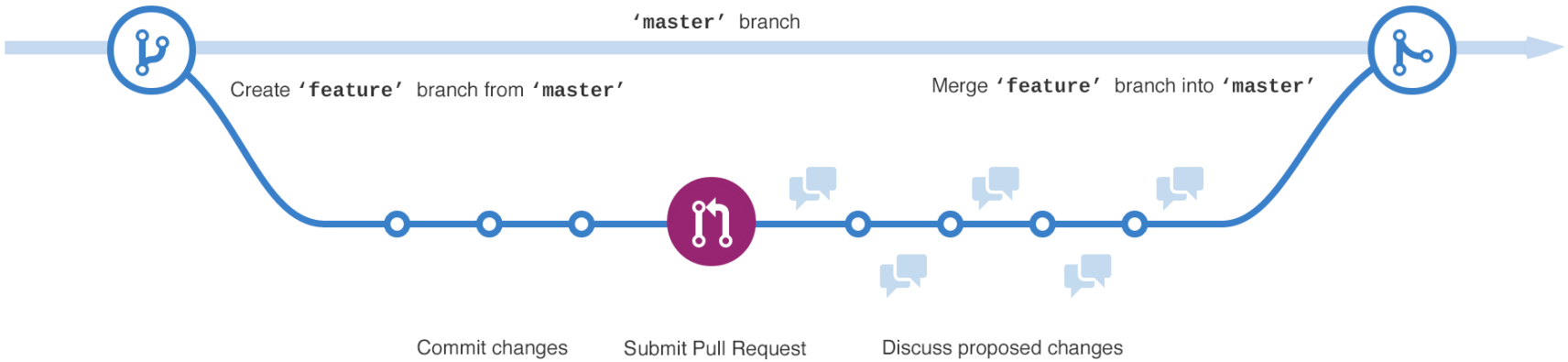
[1] Institut für Biologie, Theoretische Biophysik, Humboldt-Universität zu Berlin, Berlin, Germany,
[2] Computational Neurobiology, EMBL-EBI, Wellcome-Trust Genome Campus, Cambridge, UK and
[3] Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel
* Corresponding author. Institut für Biologie, Theoretische Biophysik, Humboldt-Universität zu Berlin, Invalidenstr. 42, Berlin 10115, Germany. Tel.: +49 302 093 9040; Fax: +49 302 093 8813; E-mail: edda.klipp@rz.hu-berlin.de

Received 21.9.10; accepted 31.5.11

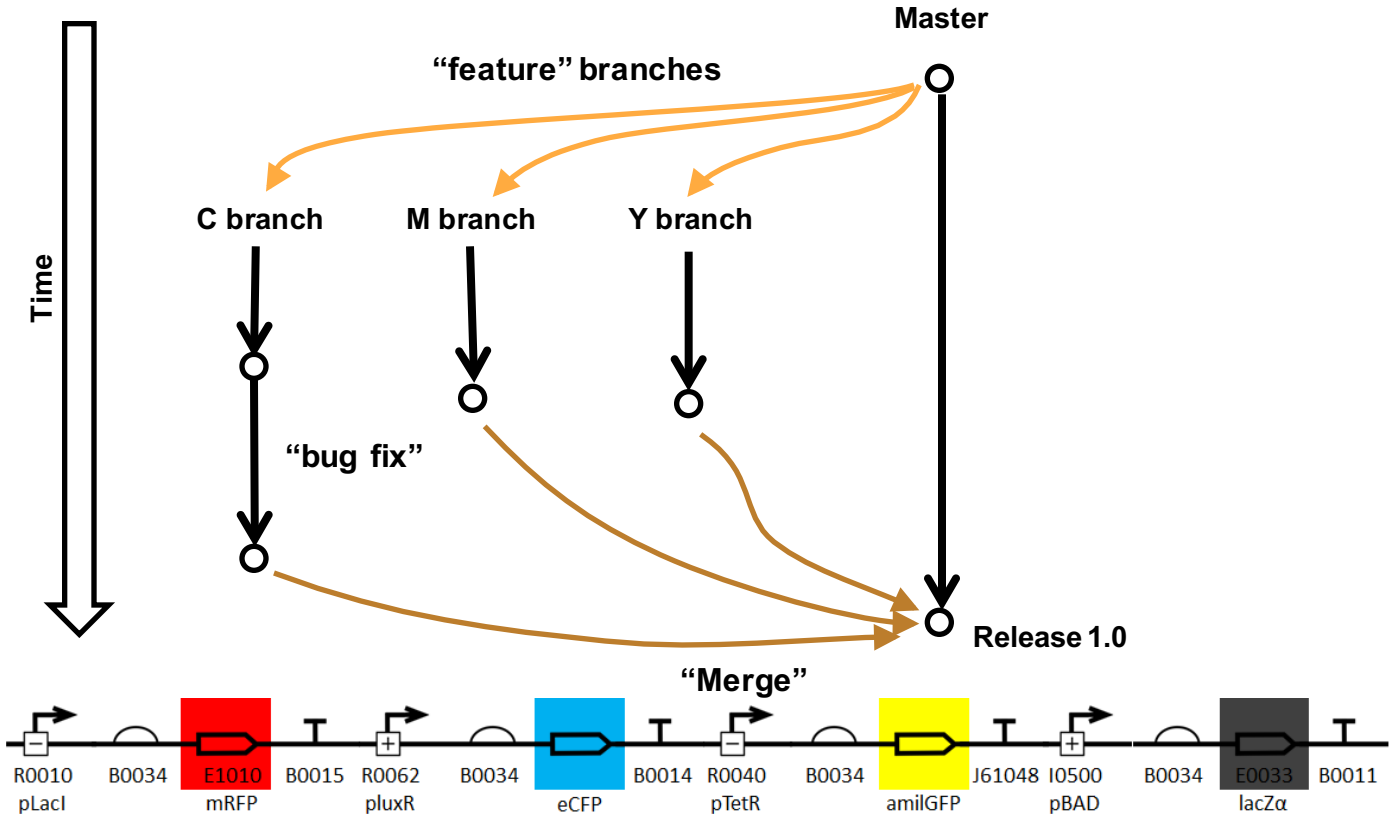
in mathematical models, which statically or dynamically describe the interconversion of biochemical compounds within reaction networks. A wealth of models, picturing various regions of the cellular networks, are available in public repositories like the BioModels Database (Le Novère *et al*, 2006) or JWS Online (Olivier and Snoep, 2004) in the machine-readable format Systems Biology Markup Language (SBML; Hucka *et al*, 2003). Meta-information on existing databases can be found on websites like PathGuide (Bader *et al*, 2006). The models in these repositories serve as information sources and they may be reused, refined, and combined for new research studies.

Continued research aiming for improved and complex models, e.g., for biomedical purposes, makes it desirable to

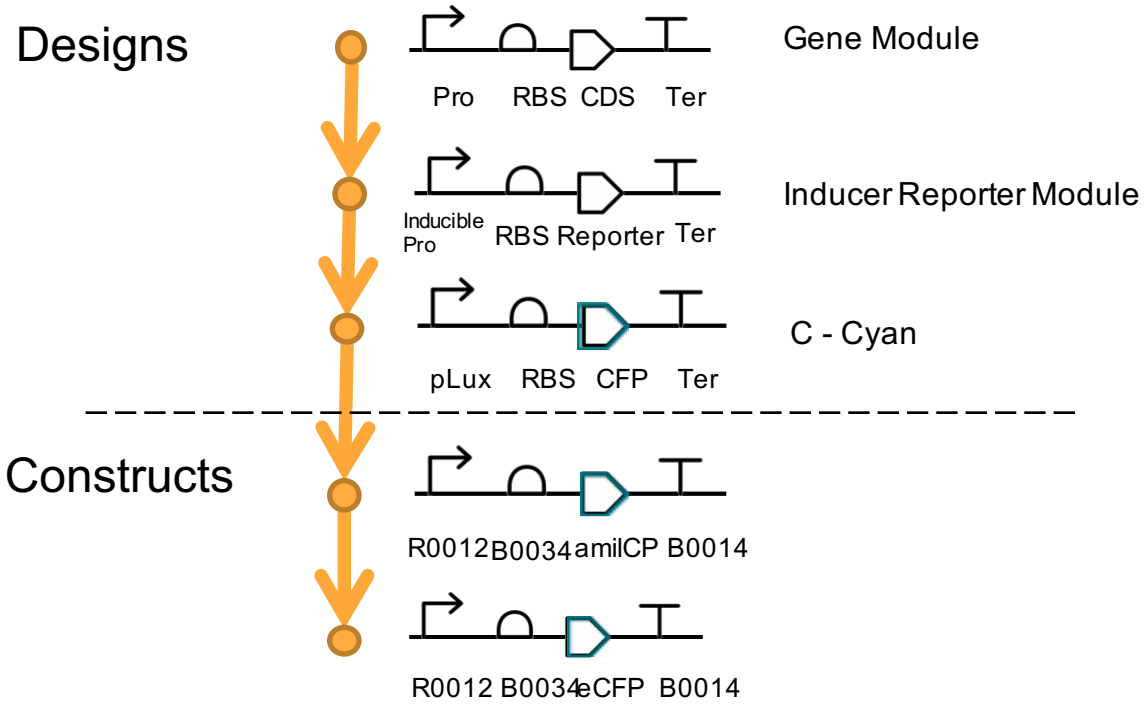
Version Control Helps the Programmer Keep Track of Software at Different Stages of Development



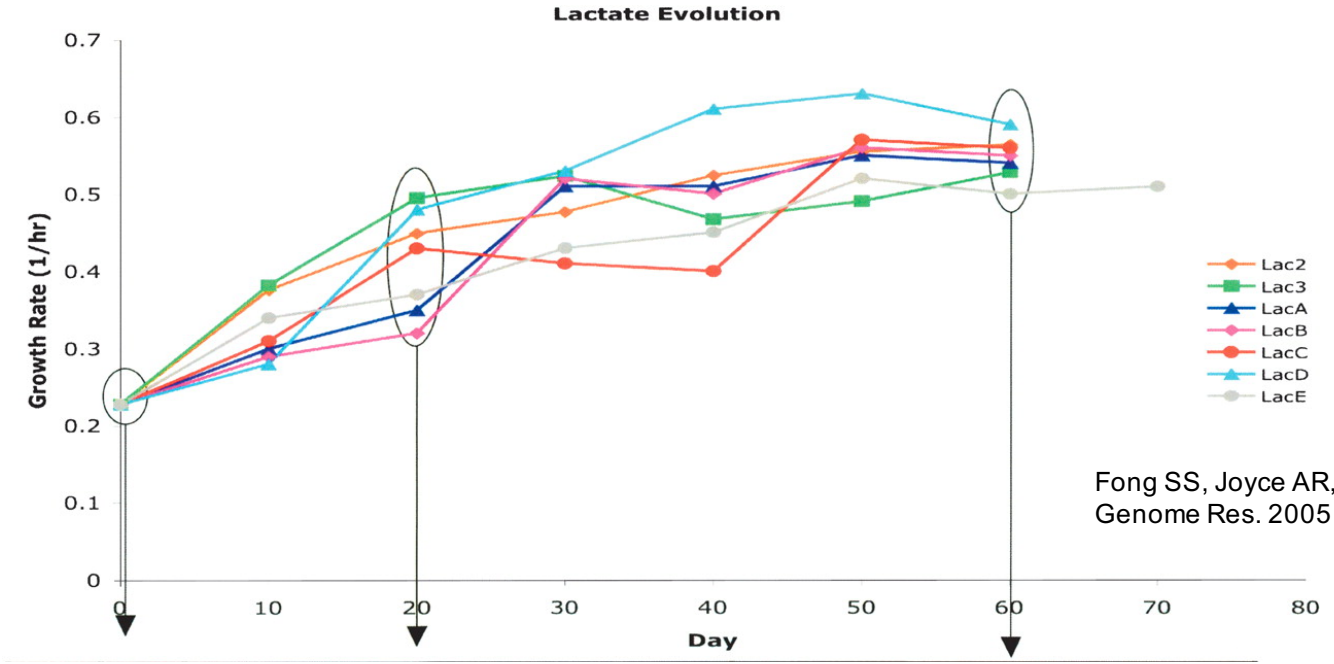
As Synthetic Biology Projects Grow in Complexity, Versioning Systems Could Prove Useful



SBOL Currently Supports Versioning of Designs, but not Constructs



Versioning Needs to Track Provenance and Contextual Information



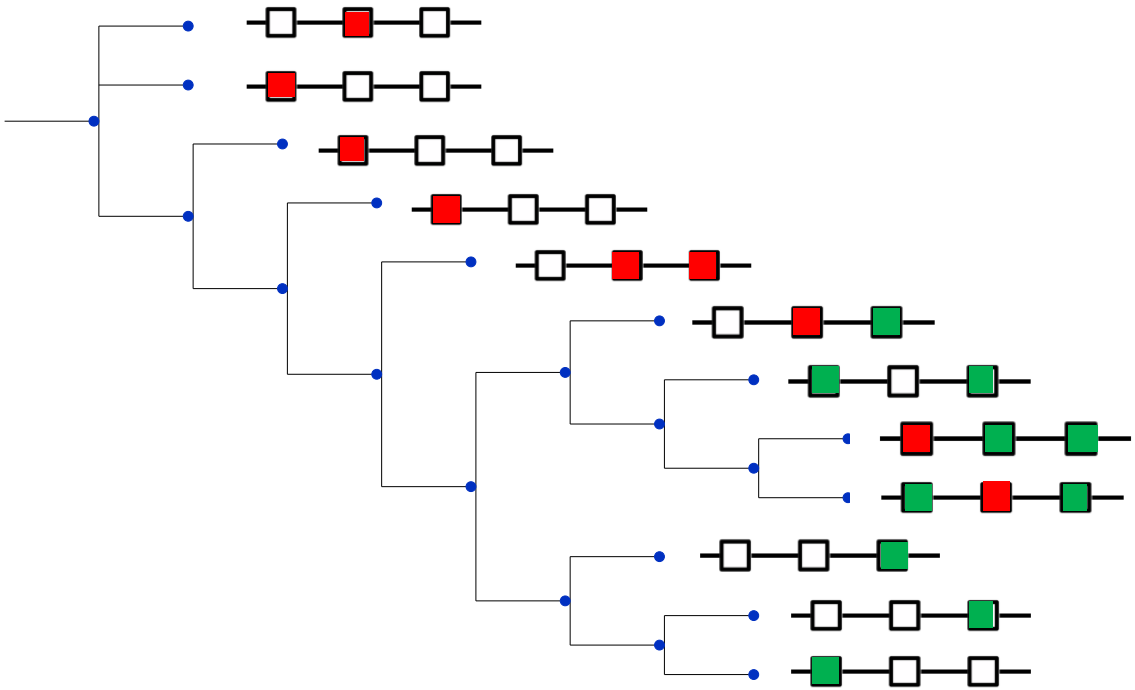
Fong SS, Joyce AR, Palsson BØ.
Genome Res. 2005 Oct;15(10):1365-72.

“Branching” is the Rule in Biology…

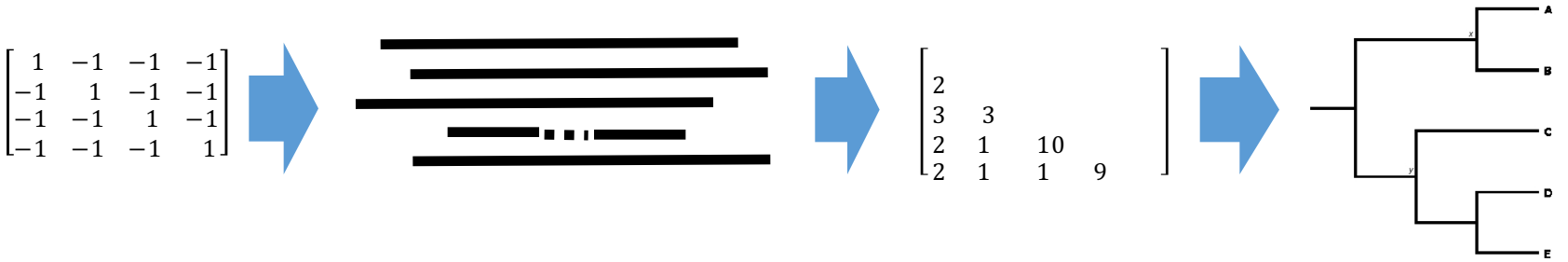
- 7 terminator variants: Sleight et al. *Designing and engineering evolutionary robust genetic circuits*. JBE 4:12
- 60 logic gate variants: Nielsen et al, *Genetic circuit design automation*, Science 351:6281.
- 217 combinatorial promoters: Cox et al. *Programming gene expression with combinatorial promoter*, MSB 3: 145.

Current Methods for Managing Branches May be Insufficient for Synthetic Biology Projects….

Introducing Variant Control, a New Paradigm for Versioning of Genetic Constructs!



Variant Control is Based on Phylogenetic Analysis of DNA Sequences



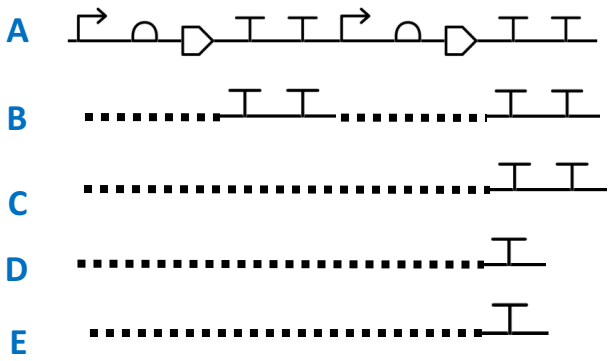
Scoring Matrix
n x n matrix
n: letters in alphabet

Multiple Sequence Alignment

Pariwise Distance Matrix
m x m matrix
m: sequences in alignment

Phylogenetic Tree

Variant Control Encodes the Composition of Genetic Circuits as Sequences



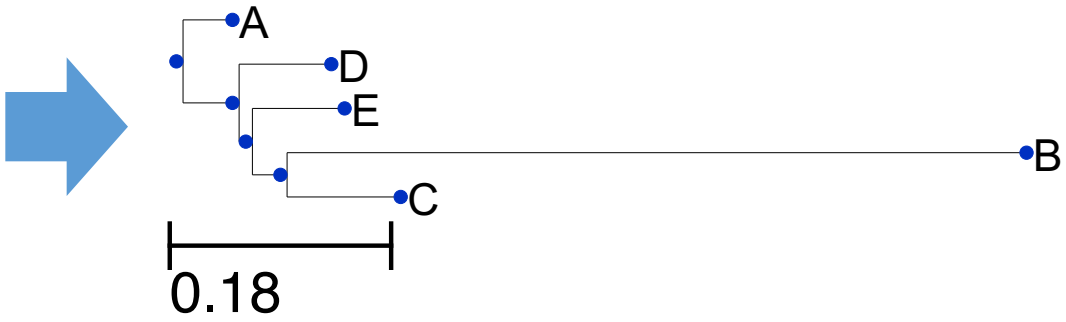
Sleight et al. *Designing and engineering evolutionary robust genetic circuits.* JBE 4:12

A	R0040	B0034	C0062	B0010	B0012	R0062	B0032	E0040	B0010	B0012
B				B0012	B0010				B0012	B0010
C									B0012	B0011
D									J61048	
E									B0011	

A	A	B	C	D	E	F	G	H	D	E
B				E	D				E	D
C									E	I
D									J	
E									K	

Parts-based Phylogenetic Analysis

```
A   ABCDEFGHDE
    |||   |||
B   ABCEDFGHED
    |||   ||||
C   ABCDEFGHEI
    ||||||||
D   ABCDEFGH-J
    ||||||||
E   ABCDEFGHK-
```

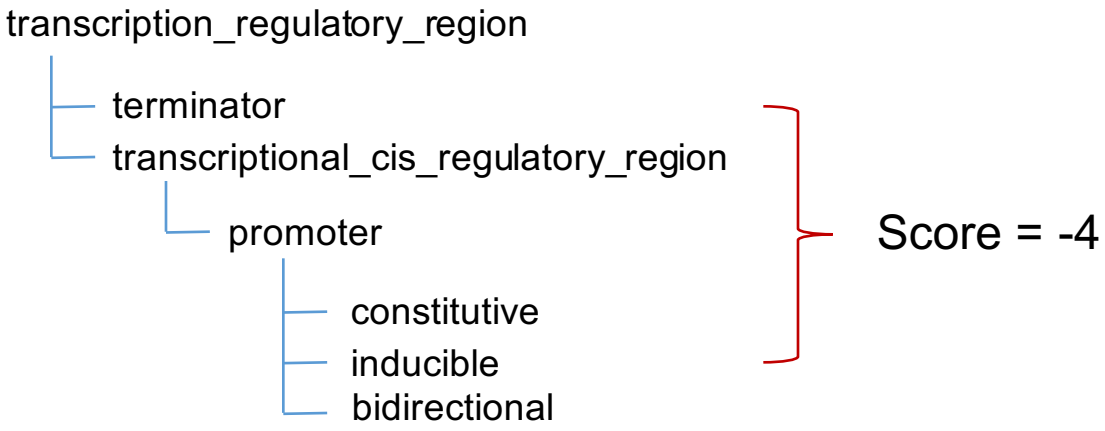


https://www.seqan.de/

ETE Toolkit

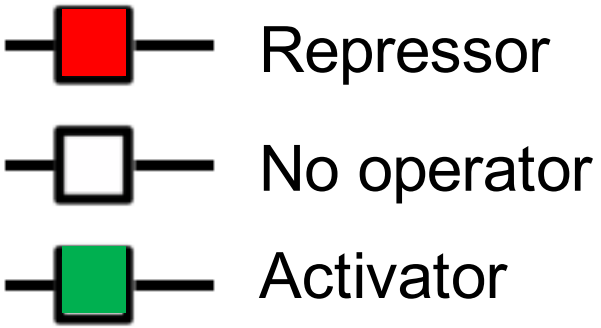
http://etetoolkit.org/

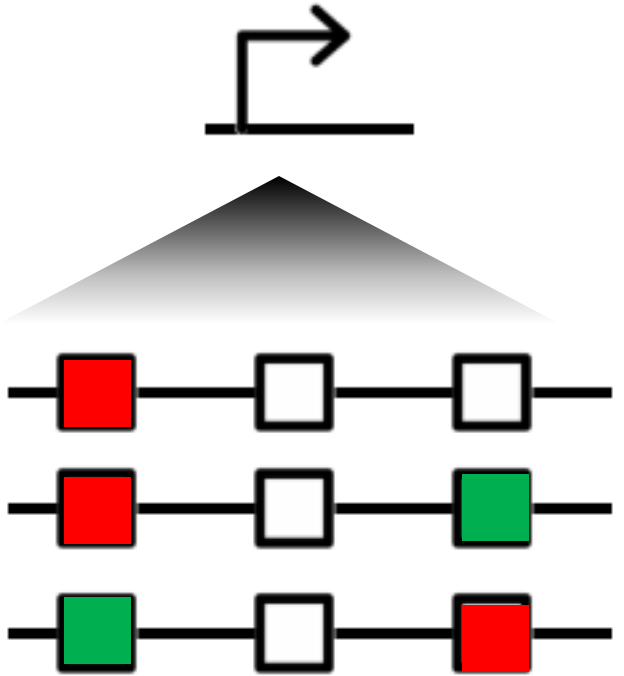
Variant Control Uses Semantic Annotations to Score Alignments



Sequence Ontology Classification Tree

Use Case: Combinatorial Promoter Variants





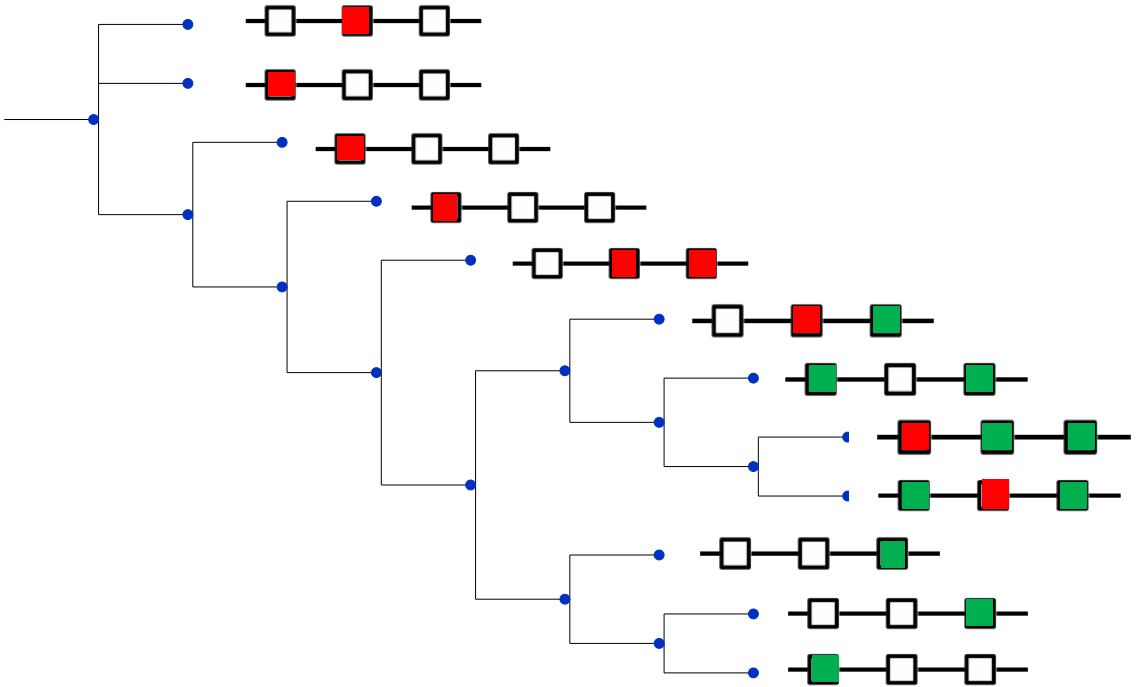
Cox et al. *Programming gene expression with combinatorial promoter*, MSB 3: 145.

Scoring Matrix Based on Semantic Annotations

	Repressor	No operator	Activator
Repressor	0	−2	−4
No operator	−2	0	−2
Activator	−4	−2	0

Scoring matrix

Variant Control Clusters Similar Designs by Both Sequence and Functional Similarity



Significance

- Variant control can determine the most parsimonious construction strategy (fewest intermediate constructs) for a library of variants given a starting template
- Variant control may be used to map and explore design space and make it easier to discover design rules
- Variant control may be used to systematically name variant libraries

Versioning Data for Genetic Constructs May Be Captured Using libSBOL Extensions

```
#define EXTENSION_CLASS "Host”
#define EXTENSION_PREFIX "host_context"
#define EXTENSION_NS "sys-bio.org/HostContext#"
class Host : public ModuleDefinition
{
   Host(sbol_type type, std::string uri) :
      ModuleDefinition(type, uri),
      modules(EXTENSION_NS "modules", this),
      parents(EXTENSION_NS "parents", this),
      children(EXTENSION_NS "children", this),
      generation(EXTENSION_NS "generation", this, 1),
      medium(EXTENSION_NS "medium", this, "www.ebi.ac.uk/efo/EFO_0000579"),
      vendorId(EXTENSION_NS "vendorId", this, "sigmaaldrich.com/L2542")
   {
      register_extension < Host > (EXTENSION_PREFIX, EXTENSION_NS EXTENSION_CLASS);
   };
}
```

https://github.com/SynBioDex/libSBOL

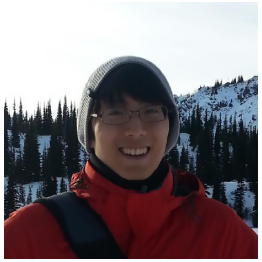
Example Serialization of Versioning Data

```
<?xml version="1.0"  encoding="utf-8"?>
<rdf:RDF xmlns:dcterms="http://purl.org/dc/terms/#"
   xmlns:host_context="sys-bio.org/HostContext#"
   xmlns:prov="http://www.w3.org/ns/prov#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:sbol="http://sbols.org/v2#">
  <host_context:Host  rdf:about="sys-bio.org/BB1">
    <sbol:persistentIdentity  rdf:resource="sys-bio.org/BB1"/>
    <host_context:generation>1</host_context:generation>
    <host_context:medium  rdf:resource="www.ebi.ac.uk/efo/EFO_0000579"/>
    <host_context:vendorId>sigmaaldrich.com/L2542</host_context:vendorId>
  </host_context:Host>
  <sbol:ModuleDefinition  rdf:about="sys-bio.org/CRISPRTemplate">
    <sbol:persistentIdentity  rdf:resource="sys-bio.org/CRISPRTemplate"/>
  </sbol:ModuleDefinition>
  <sbol:ModuleDefinition  rdf:about="sys-bio.org/CRPbCircuit">
    <sbol:persistentIdentity  rdf:resource="sys-bio.org/CRPbCircuit"/>
  </sbol:ModuleDefinition>
</rdf:RDF>
```

Acknowledgements

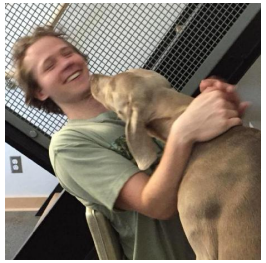
Michal Galdzicki

John Gennari



Kiri Choi

Herbert Sauro



Kyle Medley

NSF #1355909