

Defining Domain-Specific DNA Programming Languages Using GenoCAD 2.2 Grammar Editor

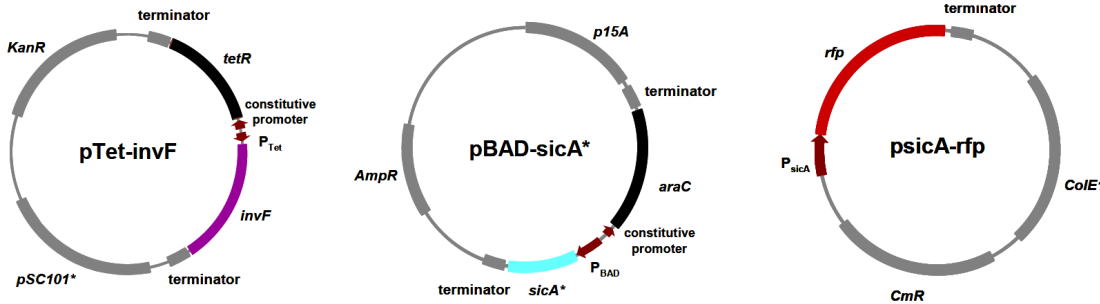
Jean Peccoud, Laura Adam, Mandy Wilson

Virginia Bioinformatics Institute

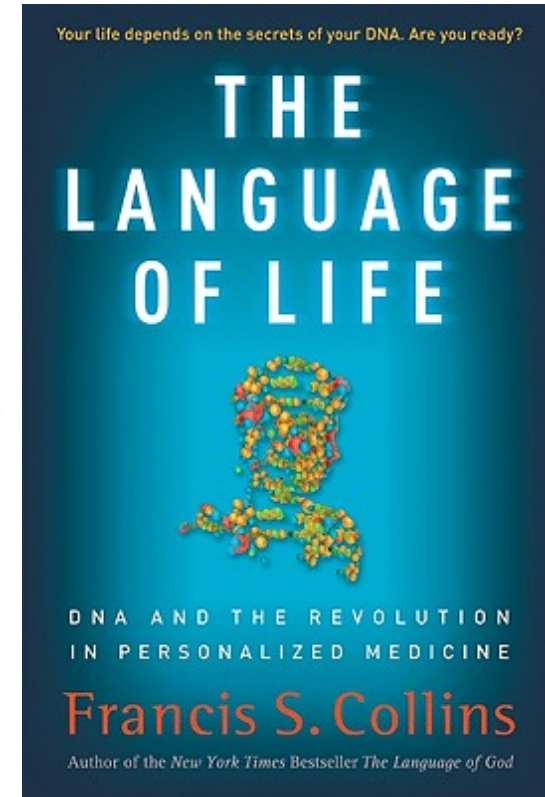
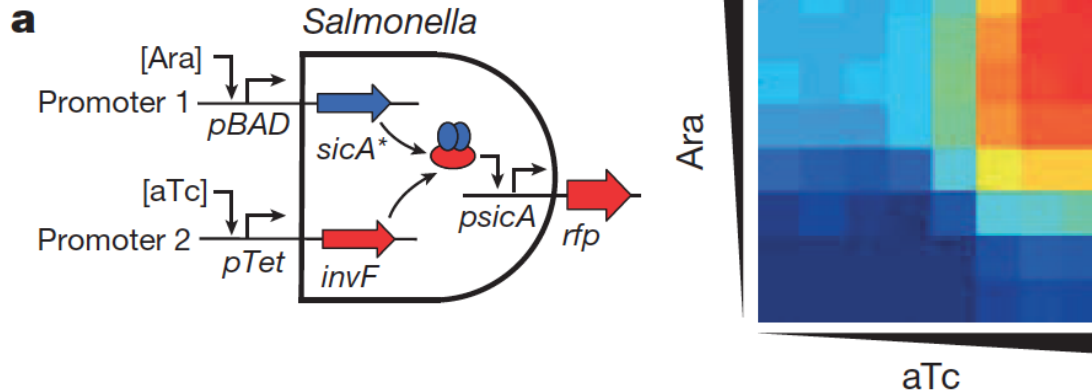
Virginia Tech

Synthetic DNA molecules as programs

Program



Behavior



Moon et al (2012) **Genetic programs constructed from layered logic gates in single cells** Nature (491) 249-53

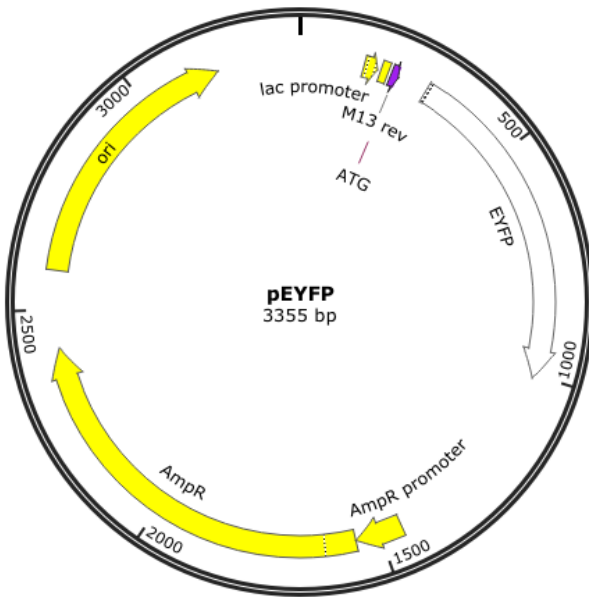
Different Dialects

Plasmids for different purposes

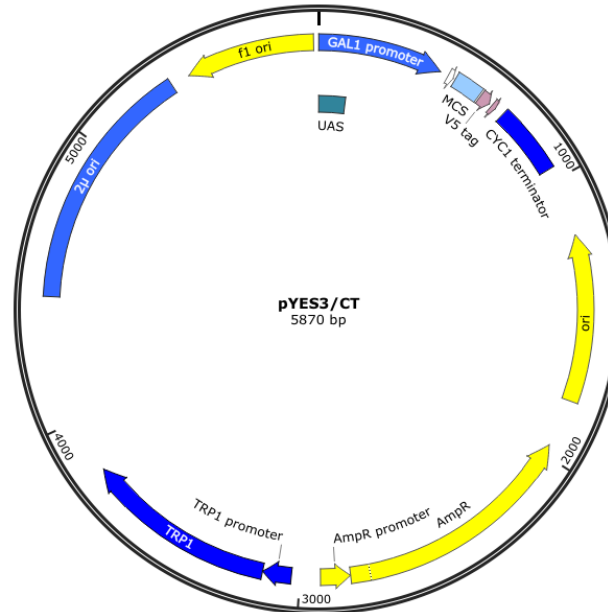
Created with SnapGene®

Created with SnapGene®

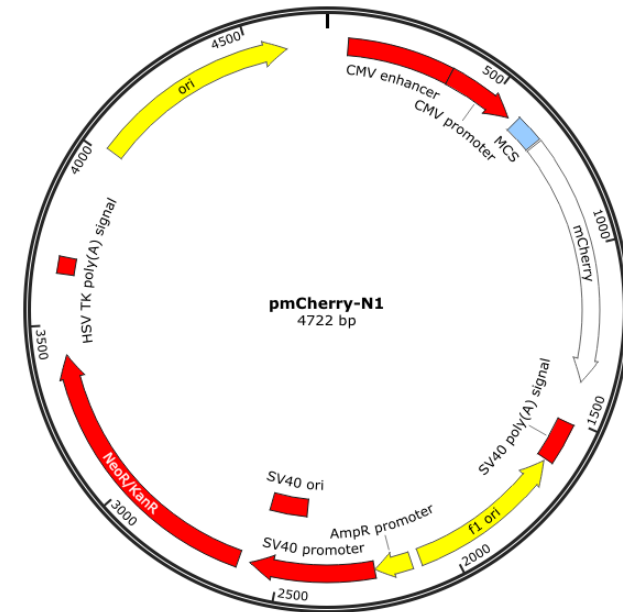
Created with SnapGene®



Bacterial



Yeast



Mammalian

Domain Specific Languages

No universal language

- ▶ C, SQL, HTML, Flash

Languages express design strategies

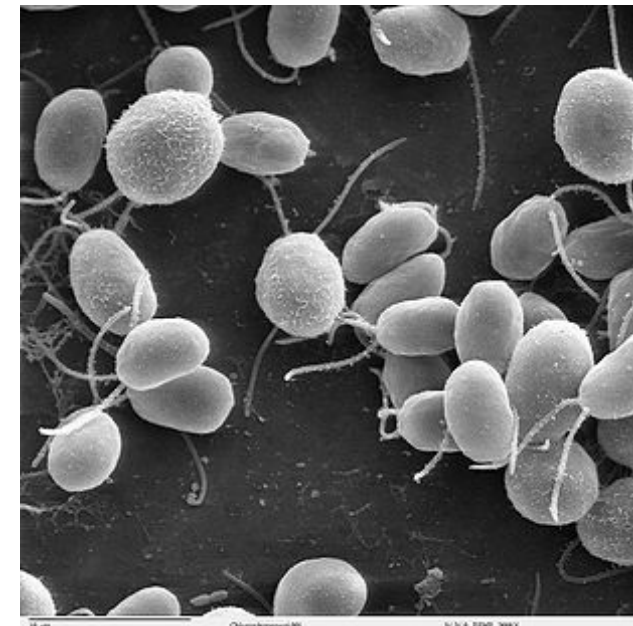
- ▶ Domain-specific
 - bacterial, yeast, mammalian
- ▶ Project-specific
 - gene therapy, synthetic biology
- ▶ Organization-specific
 - know-how, intellectual property

Empower end-users to develop their own DSL

- ▶ Gene expression in the chloroplast of microalgae



Pieter Bruegel the Elder - The Tower of Babel



Chlamydomonas reinhardtii

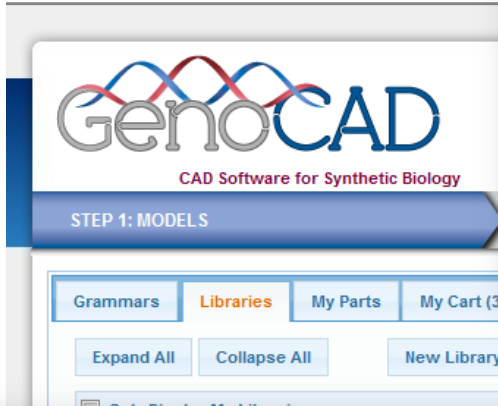


CHAPTER EIGHT

A STEP-BY-STEP INTRODUCTION TO RULE-BASED DESIGN OF SYNTHETIC GENETIC CONSTRUCTS USING GENOCAD

Mandy L. Wilson, Russell Hertzberg, Laura Adam, and
Jean Peccoud

Meth. Enz. (2011) 498



Library Name: Public Parts Library (E. coli expression grammar)
Description: This is the base library for the E. coli expression grammar

Published online 18 February 2010

Nucleic Acids Research, 2010, Vol. 38, No. 8 2637–2644
doi:10.1093/nar/gkq086

GenoCAD for iGEM: a grammatical approach to the design of standard-compliant constructs

Yizhi Cai, Mandy L. Wilson and Jean Peccoud*

Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Washington St MC 0477,
Blacksburg VA 24061, USA

Received December 13, 2009; Revised January 28, 2010; Accepted February 1, 2010



Published online 8 May 2009

W40–W47 *Nucleic Acids Research*, 2009, Vol. 37, Web Server issue
doi:10.1093/nar/gkp361

Writing DNA with GenoCAD™

Michael J. Czar, Yizhi Cai and Jean Peccoud*

Virginia Bioinformatics Institute at Virginia Polytechnic and State University, Blacksburg, VA 24061, USA

Received March 13, 2009; Revised April 21, 2009; Accepted April 22, 2009



2013/07/24

Example of formal grammar

A **grammar** is a:

- Set of rules describing how to form sentences from a language's vocabulary

R1: Sentence → Subject + Verb + Object

R2: Subject → NounPhrase

R3: Object → NounPhrase

R4: NounPhrase → NounPhrase + Modifier

R5: Modifier → PrepositionalPhrase

How to build a sentence?

Sentence

Subject

Verb

Object

NounPhrase

NounPhrase

NounPhrase

Modifier

Prepositional
Phrase

R1

R3

R4

R5

R1: **Sentence** → Subject + Verb + Object

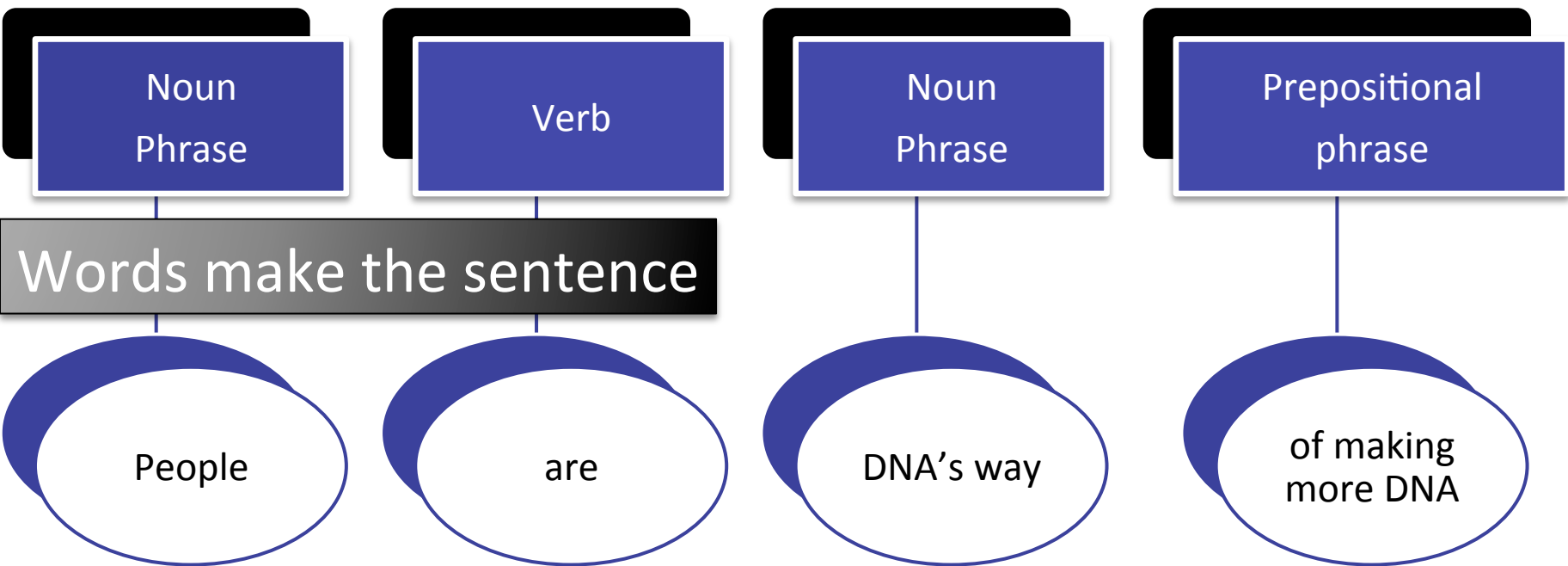
R2: **Subject** → NounPhrase

R3: **Object** → NounPhrase

R4: **NounPhrase** → NounPhrase + Modifier

R5: **Modifier** → PrepositionalPhrase

Rules make the structure



~ Edward O. Wilson, 1975

Identify and Define Categories

Category	Definition
5FLR / 3FLR	5' / 3' Flanking region for homologous recombination
SIS	Short Interval Sequences used to make polycistronic cassettes
STP	Stop codon
ATG	Start codon
GEN	Gene or protein domain. By convention does not include start and stop codons.
CDS	Open reading frame composed of several protein domains. Does not include start and stop codons.
TAG	Epitope tags. By convention does not include Start or Stop codons.
PBS	Sequence associated with the initiation of transcription and translation.
TCS	Targeted expression cassette. Expression cassette flanked with two adjacent genomic sequences for homologous recombination.
CAS	Expression cassette delimited by a promoter in 5' and a transcription terminator in 3'.

Grammar Editor – Add/Edit Categories

Manage Grammar

C. reinhardtii Chloroplast Grammar_v5

Add / Import Grammar

Name: C. reinhardtii Chloroplast Grammar_v5
Description: This grammar captures rules to design expression vectors for the Chlamydomonas reinhardtii chloroplast. Authors: Sakiko Okumoto, Mandy L. Wilson, Jean Peccoud.
Starting Category: Start / Transcription unit (S)
Icon Set: main_icon_set

Edit

Export

Copy

Test

Categories

New Category

Category Detail


Delete Category

Edit Category

Expand All

Collapse All

- 📁 Rewritable Categories
 - Start / Transcription unit (S)
 - Cassette (CAS)
 - Open reading frame (CDS)
 - Gene (GEN)
 - Terminator (TER)
 - Targeted expression cassette (TCS)
- 📁 Terminal Categories
 - Opening reverse complement delimiter (|)
 - Closing reverse complement delimiter (|)
 - Opening plasmid delimiter (|)
 - Closing plasmid delimiter (|)
 - Promoter-RBS (PBS)**
 - Start codon (ATG)
 - Vector (VEC)
 - Short Interval Sequence (SIS)
 - Stop codon (STP)
 - Epitope tags (TAG)
 - 5' Flanking Region (5FLR)
 - 3' Flanking Region (3FLR)
- 📁 Orphaned Categories
 - Opening chromosome delimiter (|)

Letter: PBS
Description: Promoter-RBS
Genbank Qualifier: N/A
Icon:  icon_c.png







Category Rules

Add Rule

Child Rules

Code	Rule	Edit	Delete
There are no child rules for the selected category.			

Parent Rules

Code	Rule	Edit	Delete
PRCT			
MRUL			

Grammar Editor – Add/Edit Categories

Manage Grammar

Name:
Description:
Starting Category:
Icon Set:

Categories

Expand All

- Rewritable
 - Start / T...
 - Cassett...
 - Open re...
 - Gene (C...
 - Termina...
 - Targete...
- Terminal C...
- Opening...
- Closing...
- Opening...
- Closing...
- Promote...
- Start co...
- Vector (...
- Short In...
- Stop codon (STP)
- Epitope tags (TAG)
- 5' Flanking Region (5FLR)
- 3' Flanking Region (3FLR)
- Orphaned Categories
 - Opening chromosome delimiter (/)

Edit Category

Letter: *

Description: *

Genbank Qualifier:

Icon:

Save **Cancel**

- Export Grammar
- Export
- Test
- Category
- Add Rule
- Edit Delete

Code	Rule	Edit	Delete
PRCT			
...			

2013/07/24

BOSC 2013

Reserved Categories

Category	Definition
[and]	Negative orientation delimiters
(and)	Plasmid delimiters
{ and }	Chromosome delimiters



Define rewriting rules

Code	Rule	Comment
CAS	S -> TCS	This rule is used to design only one expression cassette
1PLAS	S -> (VEC TCS)	This rule is used to specify the expression cassette along with the vector where it is inserted. The output is the entire plasmid sequence.
2PLAS	S -> (VEC TCS) (VEC TCS)	This rule is for designs that involve two plasmids.
TGS	TCS -> 5FLR CAS 3FLR	Specifies the flanking regions for homologous recombination.
PRCT	CAS-> PBS CDS TER	A gene expression cassette is composed of a promoter, open reading frame, and a transcription terminator.
2CAS	CAS -> CAS CAS	This rule makes it possible to have more than one expression cassette on a construct.
rCAS	CAS -> [CAS]	This rule is used to specify that the cassette is coded on the negative strand.
2CDS	CDS -> CDS SIS CDS	This rule makes it possible to design polycistronic constructs.
SGEN	CDS -> ATG GEN STP	The open reading frame is composed of a single gene flanked by a start and stop codon.
TGEN	GEN → GEN TAG	This rule is used to add a tag to a coding sequence. It can be used iteratively to add more than one tag.
2GEN	GEN-> GEN GEN	This rule can be used to fuse two coding sequences that are not tags.

Grammar Editor – Add/Edit Rules

PRCT CAS-> PBS CDS TER A gene expression cassette is composed of a promoter, open reading frame, and a transcription terminator.

Manage Grammar

C. reinhardtii Chloroplast Grammar_v4

Add / Import Grammar

Name: C. reinhardtii Chloroplast Grammar_v4

Description: This grammar captures rules to design expression vectors for the Chlamydomonas reinhardtii chloroplast. Authors: Sakiko Okumoto, Mandy L. Wilson, Jean Peccoud.

Starting Category: Start / Transcription unit (S)

Icon Set: main_icon_set

Buttons: Edit, Export, Copy, Test, Delete

Categories

New Category

Expand All Collapse All

- Rewritable Categories
 - Start / Transcription unit (S)
 - Cassette (CAS)**
 - Open reading frame (CDS)
 - Gene (GEN)
 - Terminator (TER)
 - Targeted expression cassette (TCS)
- Terminal Categories
 - Opening reverse complement delimiter (|)
 - Closing reverse complement delimiter (|)
 - Opening plasmid delimiter (|)
 - Closing plasmid delimiter (|)
 - Promoter-RBS (PBS)
 - Start codon (ATG)
 - Vector (VEC)
 - Short Internal Sequence (SIS)
 - Stop codon (STP)
 - Epitope tags (TAG)

Category Detail


Delete Category

Edit Category

Letter: CAS

Description: Cassette







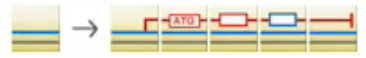


Genbank Qualifier: N/A

Icon:  icon_m.png

Category Rules

Add Rule

Child Rules

Code	Rule	Edit	Delete
2CAS			
rCAS			
NSC			

2013/07/24

BOSC 2013

14

Categories

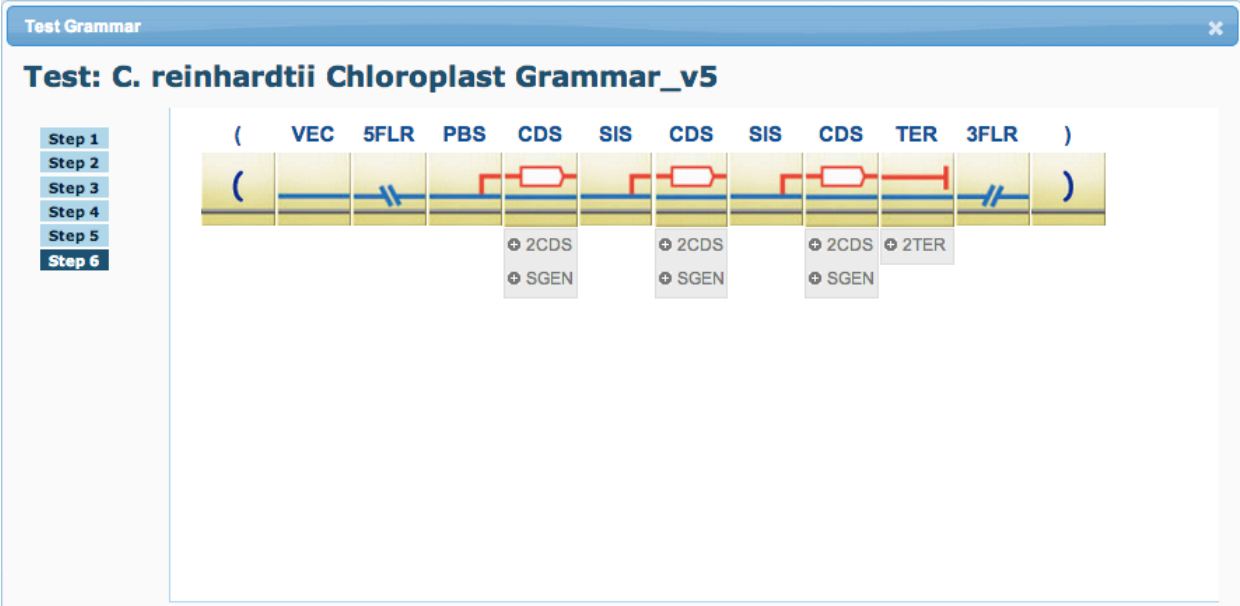
New Category

Expand All Collapse All

- 📁 Rewritable Categories
 - Start / Transcription unit (S)
 - Cassette (CAS)
 - Open reading frame (CDS)
 - Gene (GEN)
 - Terminator (TER)
 - Targeted expression cassette (TCS)
- 📁 Terminal Categories
 - Opening reverse complement delimiter (|)
 - Closing reverse complement delimiter (|)
 - Opening plasmid delimiter (|)
 - Closing plasmid delimiter (|)
 - Promoter-RBS (PBS)**
 - Start codon (ATG)
 - Vector (VEC)
 - Short Interval Sequence (SIS)
 - Stop codon (STP)
 - Epitope tags (TAG)
 - 5' Flanking Region (5FLR)
 - 3' Flanking Region (3FLR)
- 📁 Orphaned Categories
 - Opening chromosome delimiter (|)
 - Closing chromosome delimiter (|)

Debugging Tools

- Rewritable**: used on the left side
- Terminal**: used only on the right side
- Orphan**: not used



Start Over Done

Add parts in libraries

Grammars Libraries My Parts My Cart (0)

Expand All Collapse All New Library

Only Display My Libraries

- Synthetic Biology Open Language Grammar (SBOL)
- SBOL Attribute Grammar
- Basic Grammar -- No Simulation
- Basic Grammar with Simulation
- Imported Concept grammar (Cai, 2007)
- Imported E. coli Expression Grammar
- Imported BioBrick_standards
- PNAS_SynBio_Grammar
- Gene Regulatory Networks
- VSV
- Training Set E. Coli Grammar
- C. reinhardtii Chloroplast Grammar_v5
- 043013**
 - Opening reverse complement delimiter (|)
 - Closing reverse complement delimiter (|)
 - Opening plasmid delimiter (|)
 - Closing plasmid delimiter (|)
 - Opening chromosome delimiter (|)
 - Closing chromosome delimiter (|)
 - Promoter-RBS (PBS)
 - Gene (GEN)
 - Short Interval Sequence (SIS)
 - Terminator (TER)
 - Epitope tags (TAG)
 - 5' Flanking Region (5FLR)
 - 3' Flanking Region (3FLR)
- Orphaned Parts

Libraries

Add Selected to My Cart

Copy Library

Library Name: 043013

Description: homologous recombination regions added

Edit Library

Delete Library

Remove Selected from Library

Add New Part

Show 25 entries

Filter records:

<input type="checkbox"/>	Part ID	Part Name	Modified	
3' Flanking Region (3FLR)				
<input type="checkbox"/>	a0c2p	cp47245-50693	04/30/2013	
5' Flanking Region (5FLR)				
<input type="checkbox"/>	a0c2o	cp45186-47244	04/30/2013	
Closing chromosome delimiter ()				
<input type="checkbox"/>	a0c1e	Closing chromosome delimiter	04/30/2013	
Closing plasmid delimiter ()				
<input type="checkbox"/>	a0c1c	Closing plasmid delimiter	04/30/2013	
Closing reverse complement delimiter ()				
<input type="checkbox"/>	a0c1a	Closing reverse complement delimiter	04/30/2013	
Epitope tags (TAG)				
<input type="checkbox"/>	a0c2d	7x His	04/30/2013	
<input type="checkbox"/>	a0c2e	cmyc	04/30/2013	
<input type="checkbox"/>	a0c2f	flag	04/30/2013	
<input type="checkbox"/>	a0c2g	HA	04/30/2013	

Grammar Summary

Grammars Libraries My Parts My Cart (0)

Add / Import Grammar

- Public Grammars
 - Basic Grammar -- No Simulation
 - Basic Grammar with Simulation
 - Gene Regulatory Networks
- User Grammars
 - Synthetic Biology Open Language Grammar (SBOL)
 - SBOL Attribute Grammar
 - Imported Concept grammar (Cai, 2007)
 - Imported E. coli Expression Grammar
 - Synthetic Biology Open Language Grammar (SBOL)
 - Imported BioBrick_standards
 - My copy of Basic Grammar FF -- No Simulation
 - PNAS_SynBio_Grammar
 - PNAS_SynBio_Grammar_norbs
 - PNAS_SynBio_Grammar_norbs vLaura
 - Copy of Gene Regulatory Networks
 - Draft Chloroplast Grammar_v2
 - Chloroplast Grammar_v3
 - VSV
 - Training Set E. Coli Grammar
 - New Grammar
 - C. reinhardtii Chloroplast Grammar_v4
 - C. reinhardtii Chloroplast Grammar_v5**

Grammar Summary

Manage Grammar

Name:	C. reinhardtii Chloroplast Grammar_v5
Description:	This grammar captures rules to design expression vectors for the Chlamydomonas reinhardtii chloroplast. Authors: Sakiko Okumoto, Mandy L. Wilson, Jean Peccoud.
Icon Set:	main_icon_set
Supports Attributes?:	No
# Categories:	20
# Libraries:	1
# Parts:	53
# Rules:	14

Rule-based design of DNA sequences

Part:BBa_J13004

Designed by Jeff Tabor Group: iGEM_UTAustin (2005-06-28)



Composite

DNA Available
Experience Works
[Get This Part](#)

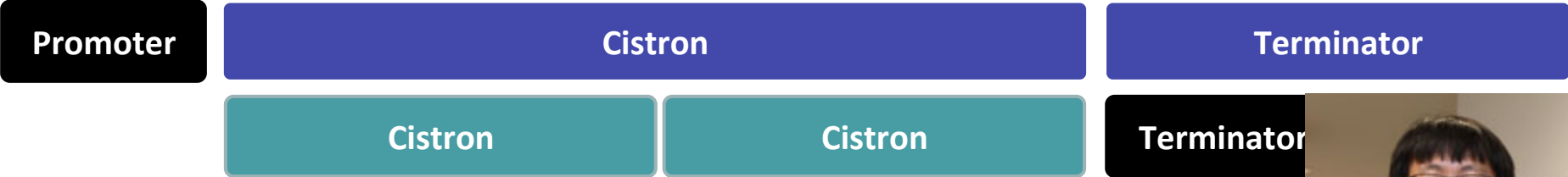
polycistronic CFP/YFP expression cassette

A single mRNA containing CFP and YFP under the strong RBS B0034 is driven by the tet repressible promoter, R0040.

Sequence and Features

Format: [Subparts](#) | [Ruler](#) | [SS](#) | [DS](#) Search: Length: 1689 bp Context: Part only [Get selected sequence](#)

Construct



BIOINFORMATICS ORIGINAL PAPER Vol. 23 no. 20 2007, pages 2760–2767
doi:10.1093/bioinformatics/btm446

Systems biology

A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts

Yizhi Cai¹, Brian Hartnett¹, Claes Gustafsson² and Jean Peccoud^{1,*}

2013/07/24



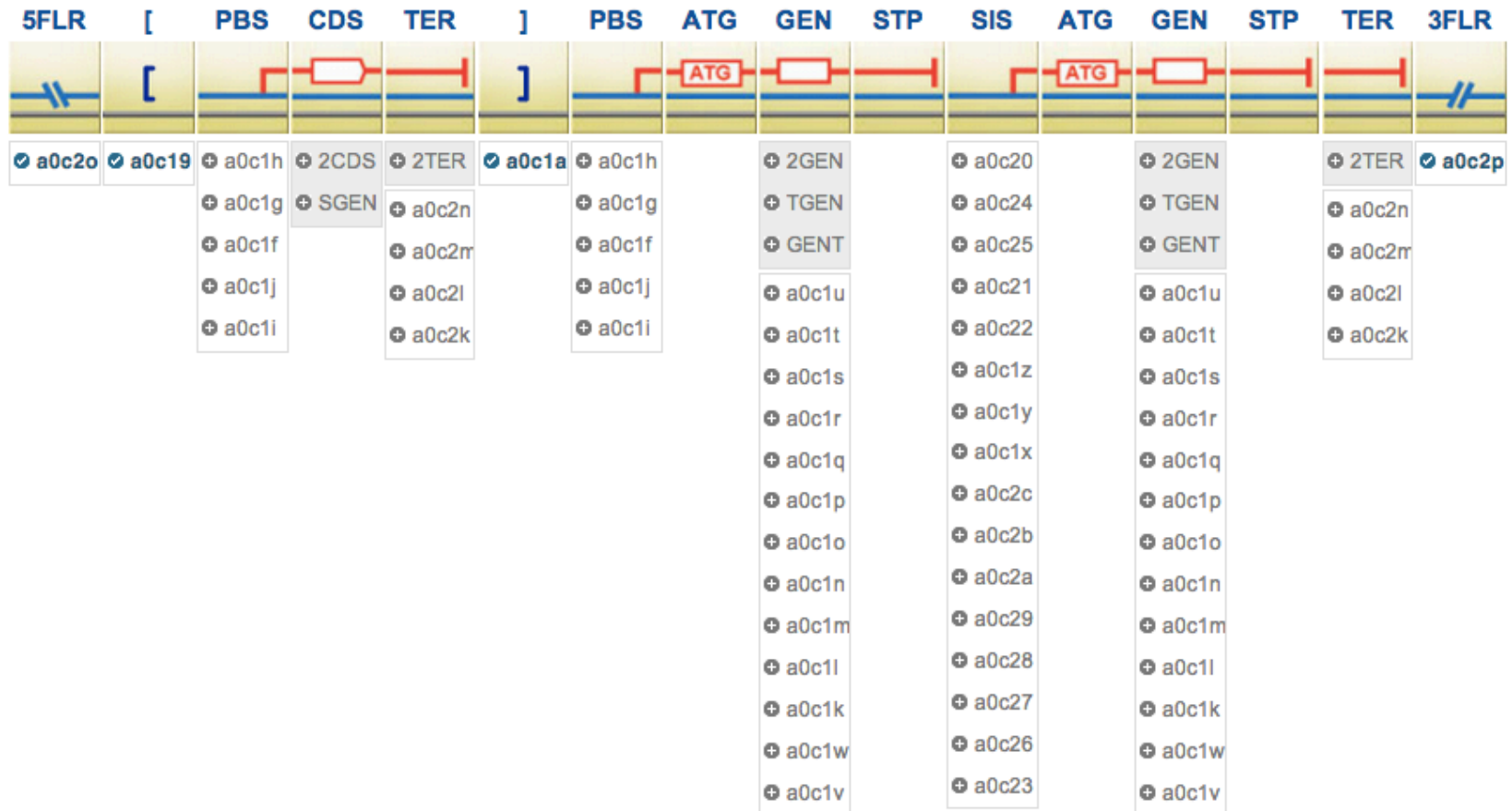
Point-and-click design tool

New Design

History

Save Design...

- Step 1
- Step 2
- Step 3
- Step 4
- Step 5
- Step 6
- Step 7
- Step 8
- Step 9
- Step 10**



A dynamic language

You can change the model ...

- ▶ Add or Change a Rule in a Grammar
- ▶ Delete a Rule from a Grammar

- ▶ R Parent Rules
- ▶ C Parent Rules
- ▶ C Library Name: Test Library

CAA
GAG
CCGT
GGG
GCTA

... but

signs?

Available Designs

Welcome, Mandy Wilson | My Profile | Log Out

GenocAD
CAD Software for Synthetic Biology

STEP 1: MODELS STEP 2: DESIGN

Available Designs

My Designs Delete Selected

Show 10 entries

<input type="checkbox"/>	Name	Description	Status	Modified	Clone	View
<input checked="" type="checkbox"/>	Repressilator	This is an oscillatory gene circuit.	✓	1/15/2011 09:44:07 AM		
<input checked="" type="checkbox"/>	Repressilator - LPG	This is an oscillatory gene circuit.	✗	1/15/2011 09:45:20 AM		
<input checked="" type="checkbox"/>	Copy of Copy of Repressilator	This is an oscillatory gene circuit.	✗	1/15/2011 09:44:31 AM		
<input type="checkbox"/>	SBOL Design1 for Export Test		✗	1/24/2012 11:27:32 AM		
<input type="checkbox"/>	SBOLDesign2		✗	2/02/2011 10:22:33 AM		
<input type="checkbox"/>	SBOL3ReallyLong		✓	2/02/2011 10:25:17 AM		
<input type="checkbox"/>	SBOL4Reallyshort		✓	1/24/2012 11:29:11 AM		
<input type="checkbox"/>	Test SBOL icons		⚠	2/06/2011 11:04:16 AM		
<input type="checkbox"/>	Test Design		⚠	2/22/2011 09:34:33 AM		
<input type="checkbox"/>	Test Design		⚠	2/22/2011 09:34:50 AM		

Showing 1 to 10 of 17 entries

Filter records:

Port Selected to: FASTA Tab-Delimited

Start New Design

First Prev 1 2 Next Last

Different design statuses



Valid – the sequence could be decomposed into its parts, and the parts' categories make up a grammar-sanctioned framework.



Needs validation – either grammar, part, or library has changed, and the sequence has not been validated since



Under construction – design is unfinished, so cannot be compiled.

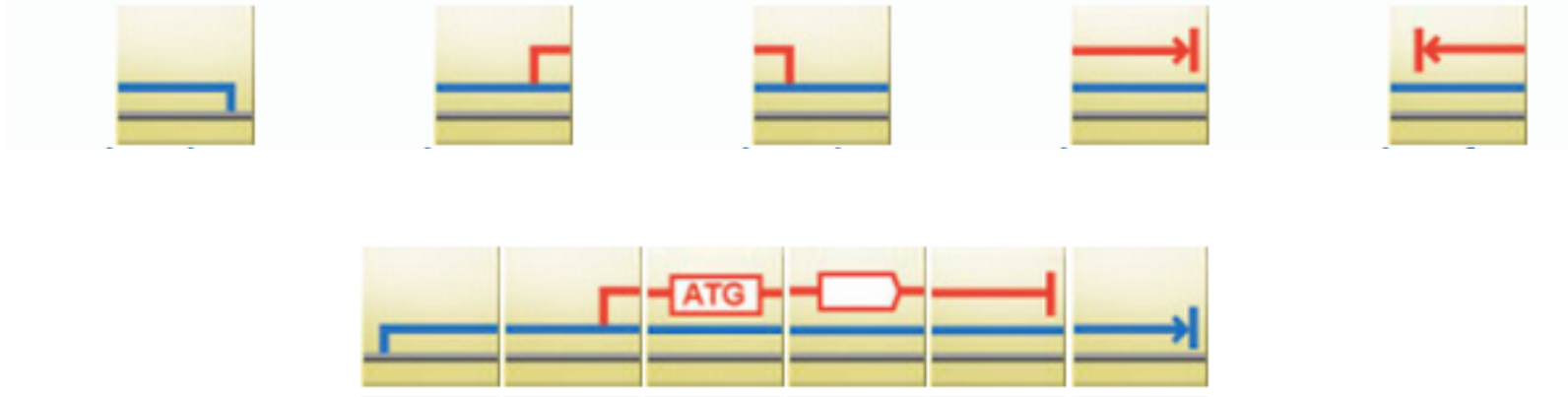


Out of Date – although design is still valid with respect to grammars and libraries, the parts have changed.

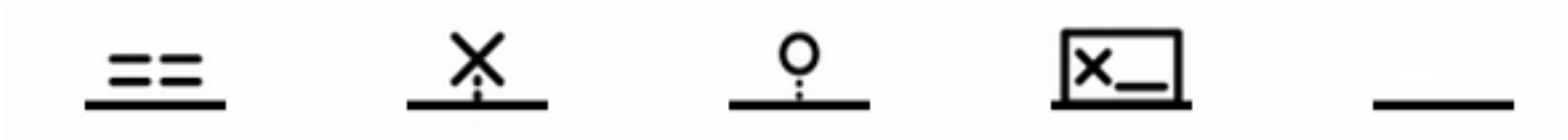


Invalid – the sequence cannot be resolved.

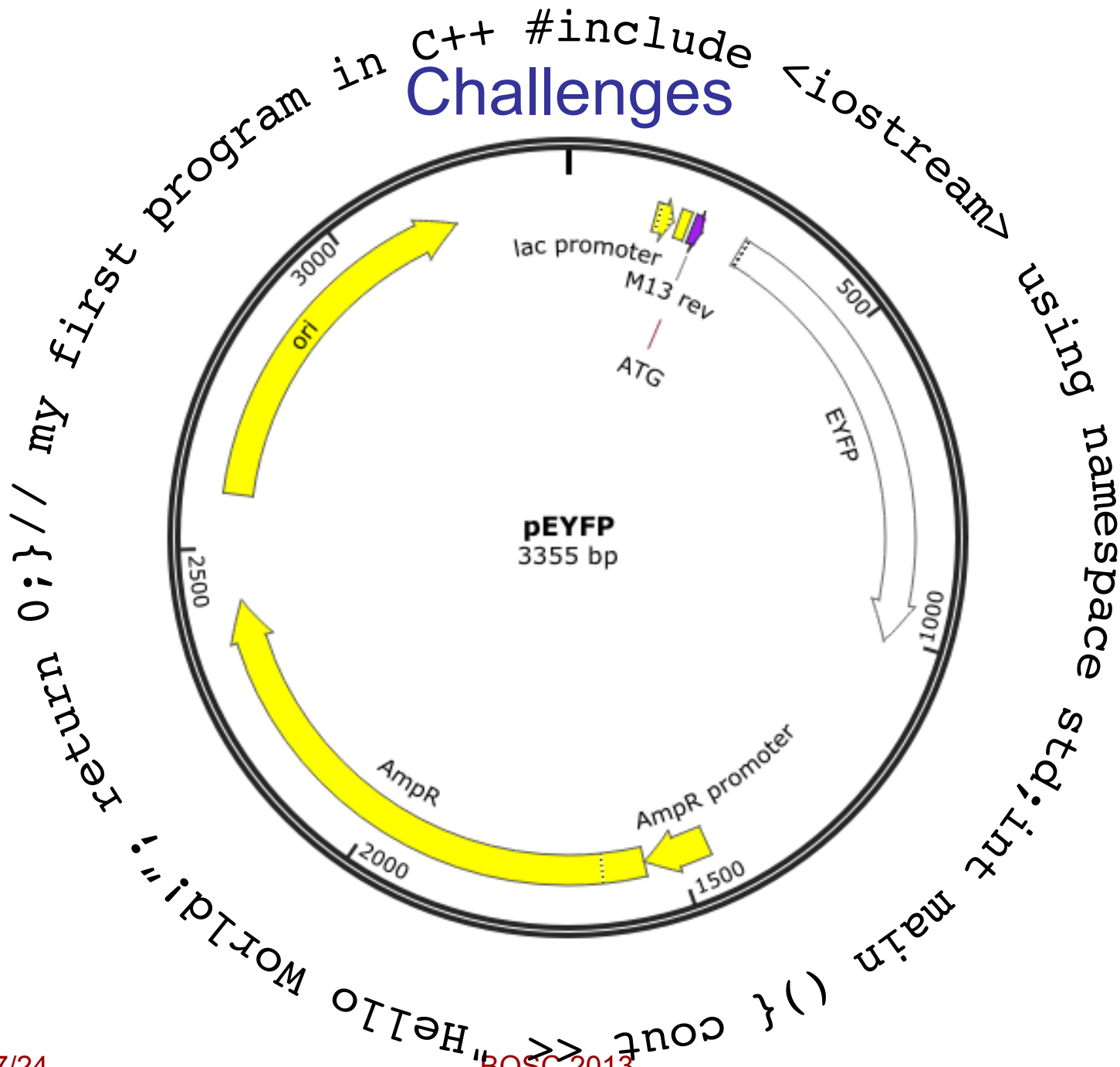
Graphical Language Icons



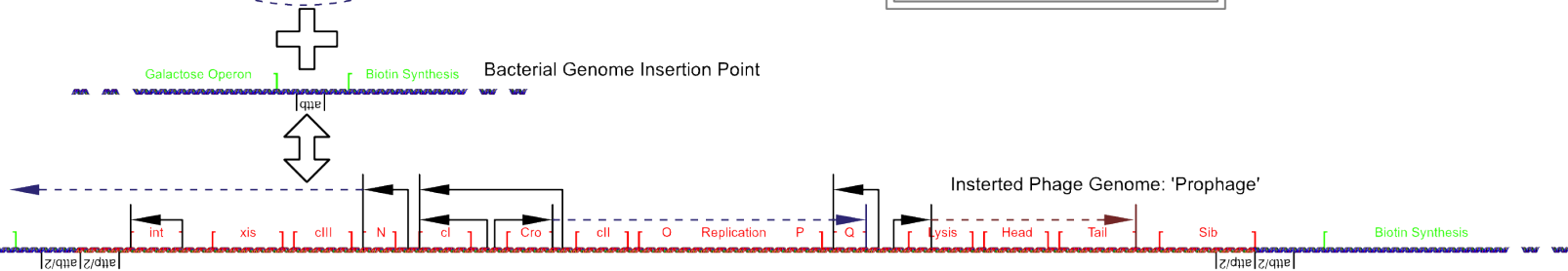
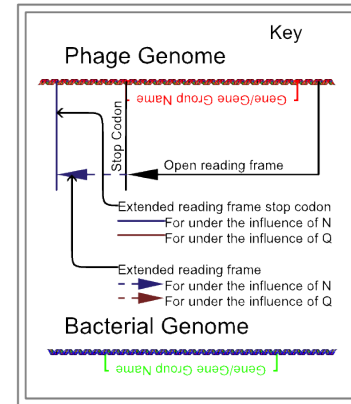
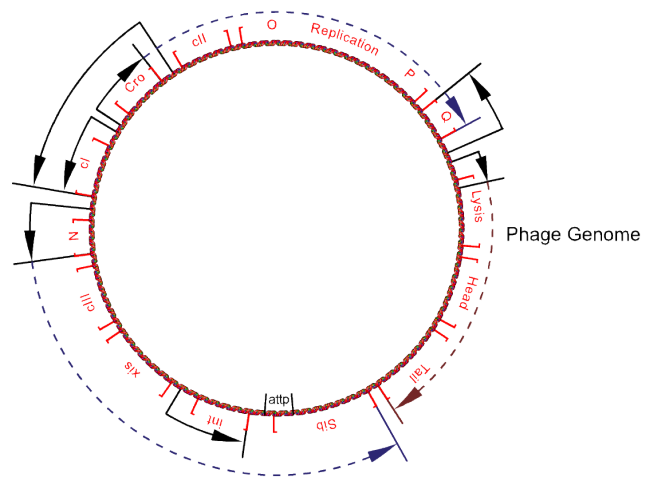
Icon sets: visual language – color – structure



Challenges



Never odd or even. No lemon, no melon.



Richard Wheeler (Zephyris) 2005; Schematic representation of the insertion of the bacteriophage lambda

Custom languages vs. Standardization

DSL in a standardized world

- ▶ GenBank
- ▶ Sequence Ontology
- ▶ Synthetic Biology Open Language

Need to map custom concepts to standards

- ▶ Data exchange
- ▶ Graphical representation
- ▶ Sequence annotation

Can we standardize customization?

SBLE 2013 – October 26 - Indianapolis

Workshop Synthetic Biology and Language Engineering

- ▶ Bring together language designers and synthetic biologists with the goal of analyzing the different programming paradigms that have been or could be explored to write these biological programs more effectively
- ▶ Part of:
 - International Conference on Software Language Engineering (SLE)
 - International Conference on Generative Programming: Concepts & Experiences (GPCE)
 - Systems, Programming, Languages and Applications: Software for Humanity (SPLASH)

<http://planet-sl.org/sble-at-sle2013/>

Acknowledgements

VBI SynBio Group

- ▶ M. Wilson
- ▶ D. Ball
- ▶ M. Lux
- ▶ L. Adam
- ▶ C. Overend

GenoCAD Alumni

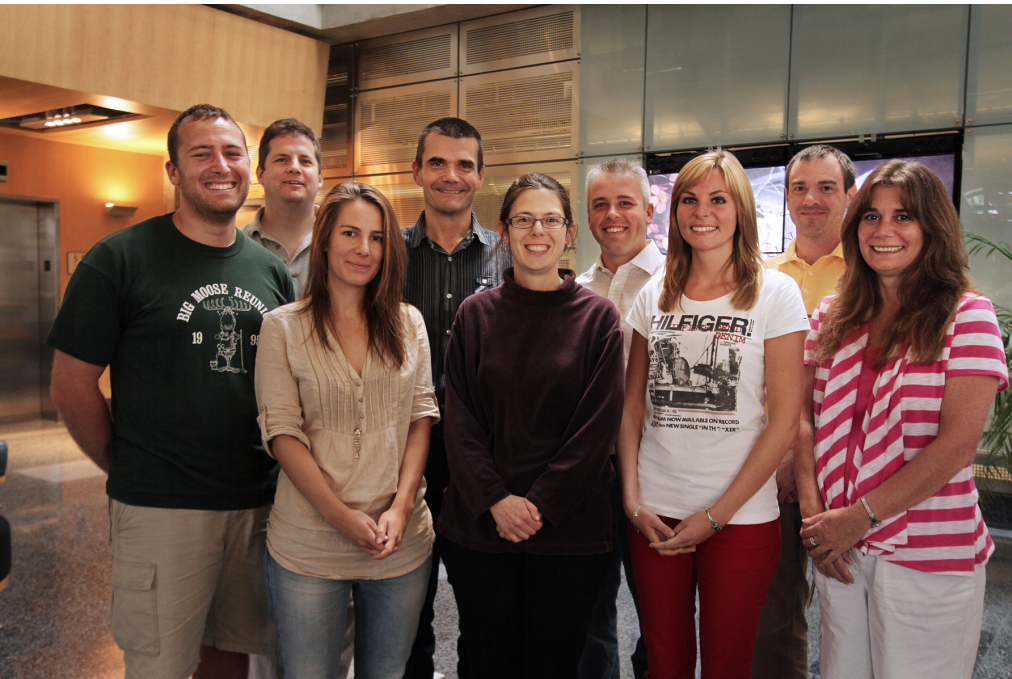
- ▶ Yizhi Patrick Cai (JHU)
- ▶ Mike Czar (Carillon)

GenoCAD Collaborators

- ▶ VBI: S. Hoops, J. Lewis
- ▶ SBOL: H. Sauro, C. Myers, D. Densmore, C. Rodriguez, M. Galdzicki and many more
- ▶ Language: Eric Van Wyck

Find GenoCAD

- ▶ www.genocad.org
- ▶ www.facebook.com/genocad/
- ▶ @genocad
- ▶ Google +
- ▶ LinkedIn



Questions?



CAD Model of VBI



Photo of VBI