

DNA Assembly is Not for You

Leave it to machines !

Valentin Zulkower
Edinburgh Genome Foundry
EASTBIO Dundee, May 18th, 2018

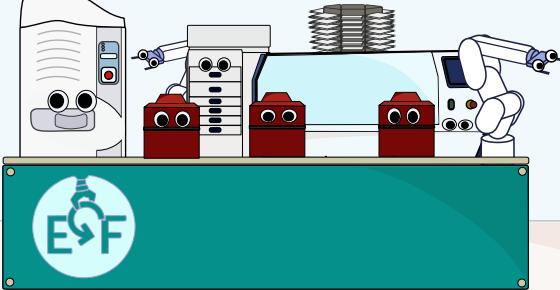
I NEED 18,000
ASSEMBLIES FOR MY
MASTERS PROJECT



GUYS ?



RIGHT AWAY !



Program

Main message: DNA Assembly is very human-unfriendly.

Program

Main message: DNA Assembly is very human-unfriendly.

Plan: We will see some common pitfalls of DNA design, assembly, and validation. And how software can help.

Program

Main message: DNA Assembly is very human-unfriendly.

Plan: We will see some common pitfalls of DNA design, assembly, and validation. And how software can help.

Design

Overhangs design

... without hangover !

Sequence design

Can you polish a gene?

Parts-based design

LEGO Madness !

Program

Main message: DNA Assembly is very human-unfriendly.

Plan: We will see some common pitfalls of DNA design, assembly, and validation. And how software can help.

Design

Overhangs design
... without hangover !

Sequence design
Can you polish a gene?

Parts-based design
LEGO Madness !

Assembly

Cloning simulation
Will it assemble ?

Automated pipetting
Hi robots !

Program

Main message: DNA Assembly is very human-unfriendly.

Plan: We will see some common pitfalls of DNA design, assembly, and validation. And how software can help.

Design

Overhangs design
... without hangover !

Sequence design
Can you polish a gene?

Parts-based design
LEGO Madness !

Assembly

Cloning simulation
Will it assemble ?

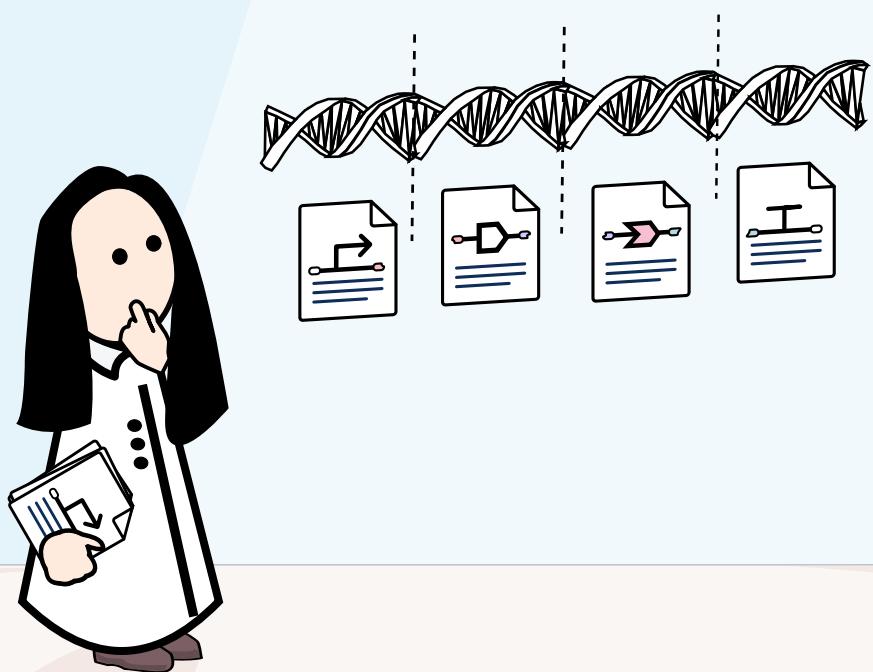
Automated pipetting
Hi robots !

Validation

Restriction digests
Say no to digestion problems!

Sanger sequencing
A primer on primers selection

1. Design



Overhangs design for Type 2S restriction assembly

Pitfalls of overhang design

1. Avoid similar overhangs

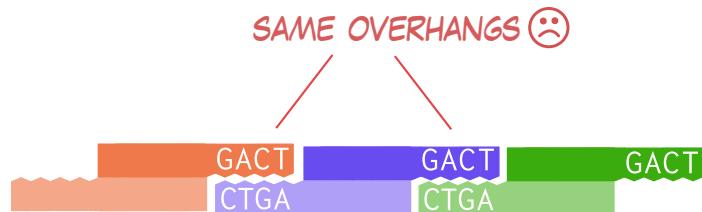
Pitfalls of overhang design

1. Avoid similar overhangs



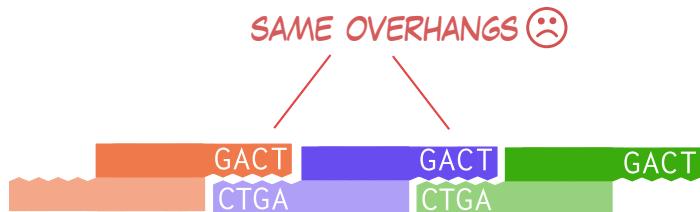
Pitfalls of overhang design

1. Avoid similar overhangs



Pitfalls of overhang design

1. Avoid similar overhangs



Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements

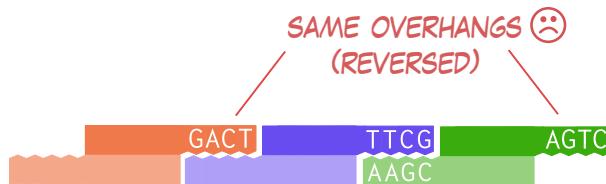
Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements



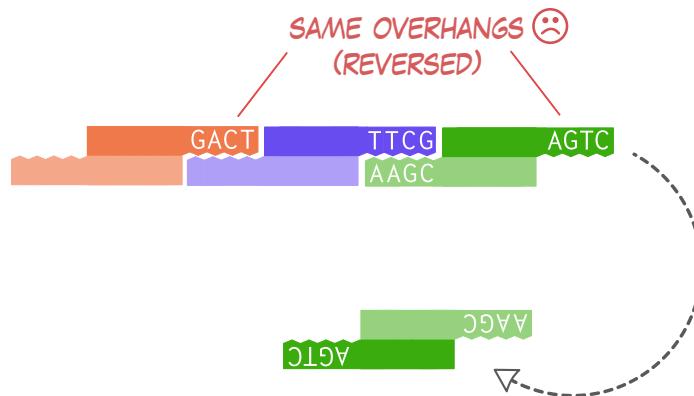
Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements



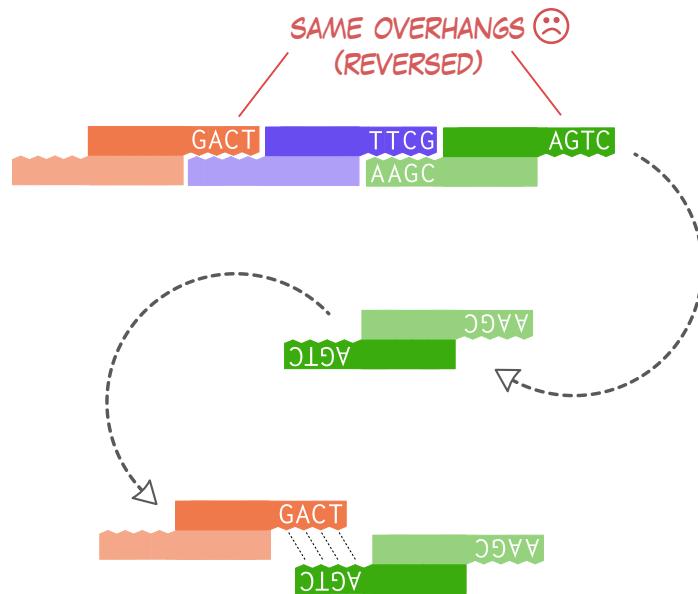
Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements



Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements

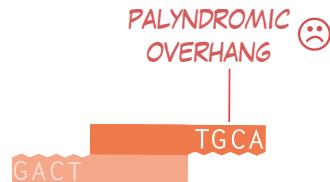


Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements
3. No palindromic sequences

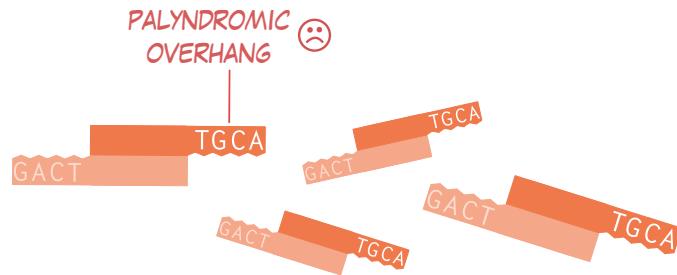
Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements
3. No palindromic sequences



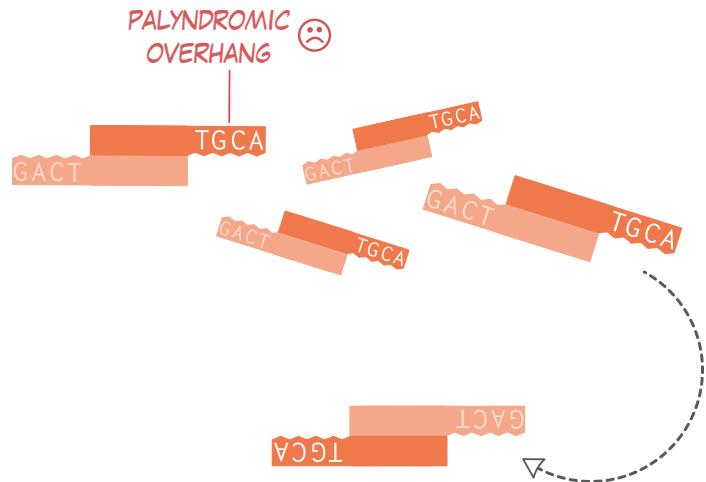
Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements
3. No palindromic sequences



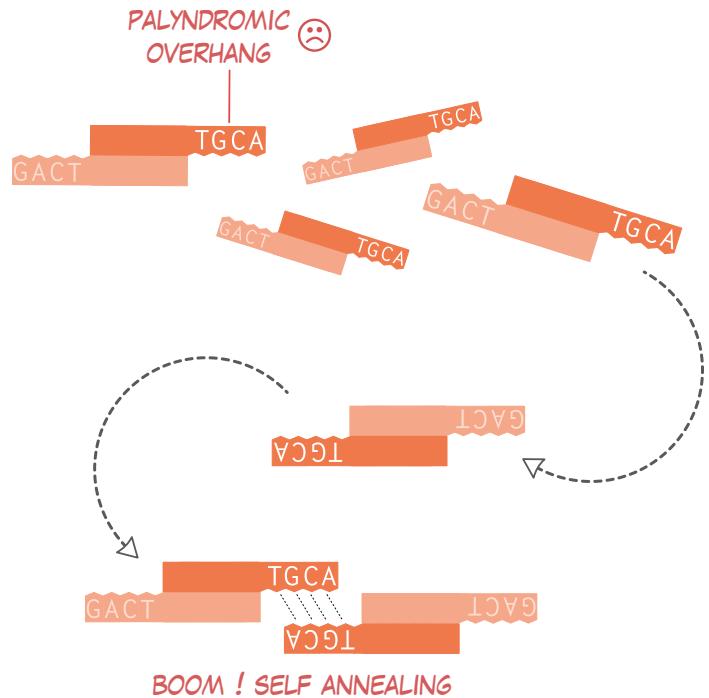
Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements
3. No palindromic sequences



Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements
3. No palindromic sequences

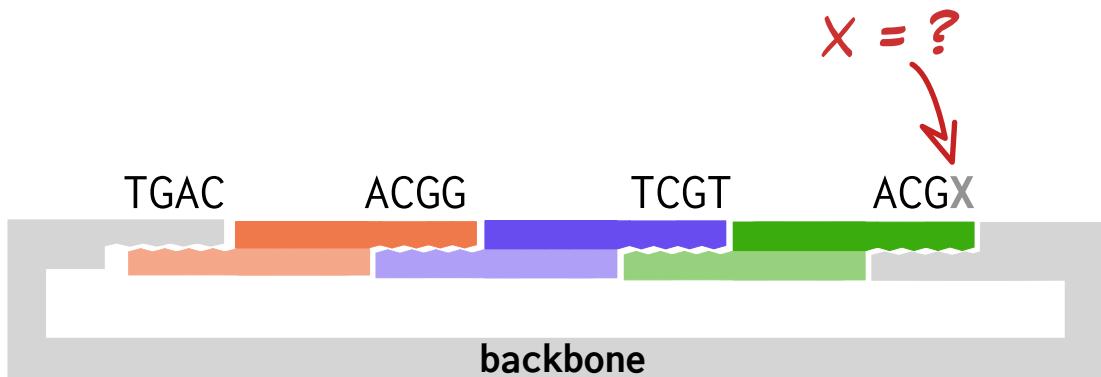


Pitfalls of overhang design

1. Avoid similar overhangs
2. Beware of reverse-complements
3. No palindromic sequences
4. Avoid 0% or 100% GC

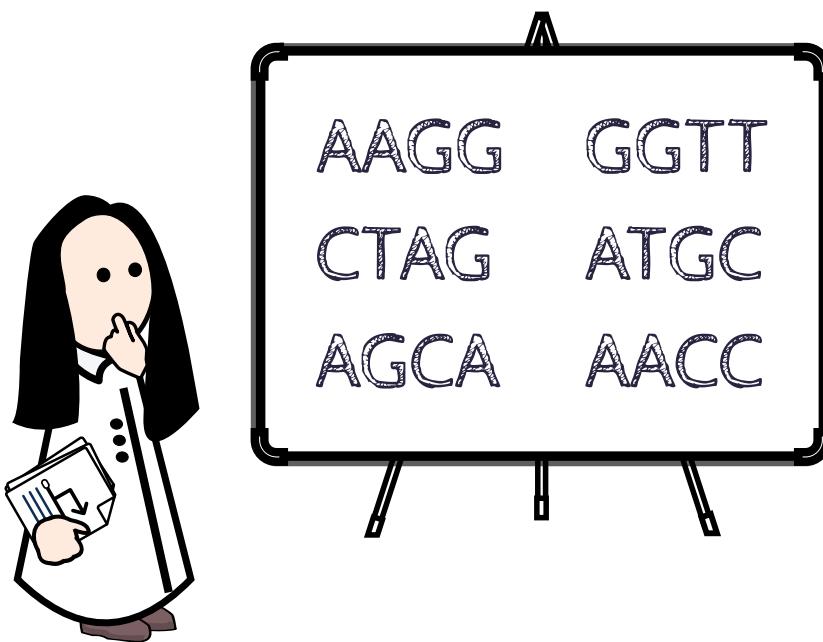
Overhangs design exercice 1

Which nucleotide will work best ?

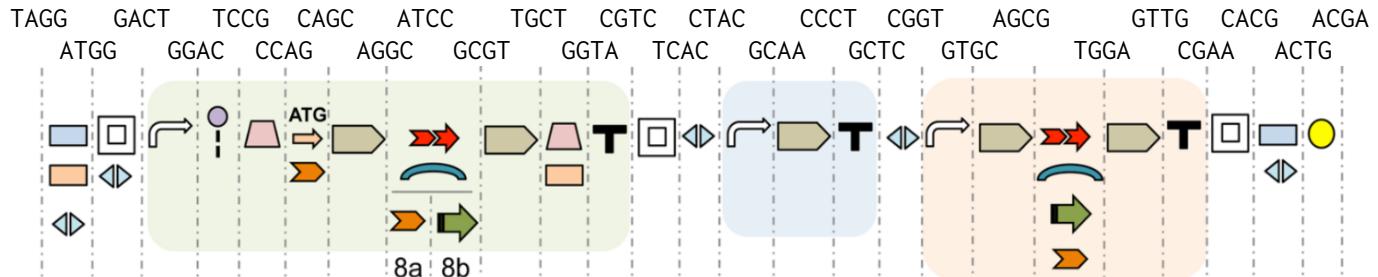


Overhangs design exercice 2

Is that a valid collection of overhangs ?



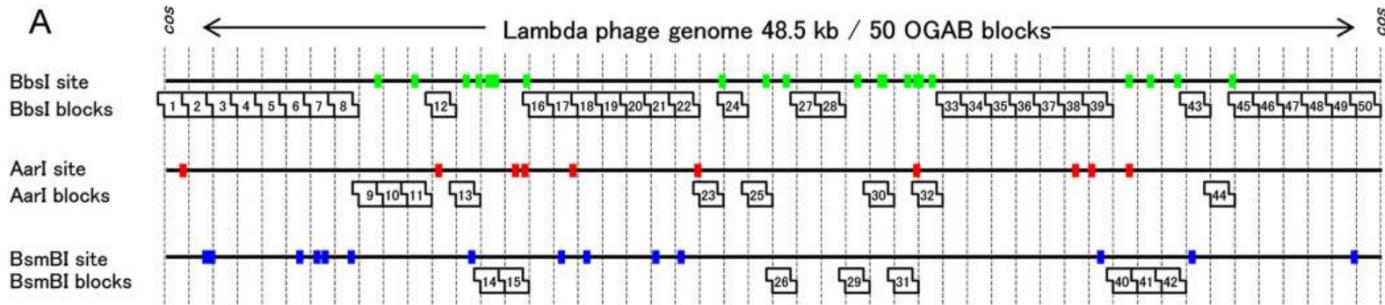
Real-world overhang design: EMMA



After Martella et al., ACS Syn. Bio. 2017

- ▶ 27 compatible overhangs
- ▶ Some mandatory overhangs (e.g. containing ATG)

Real-world overhang design: virus assembly

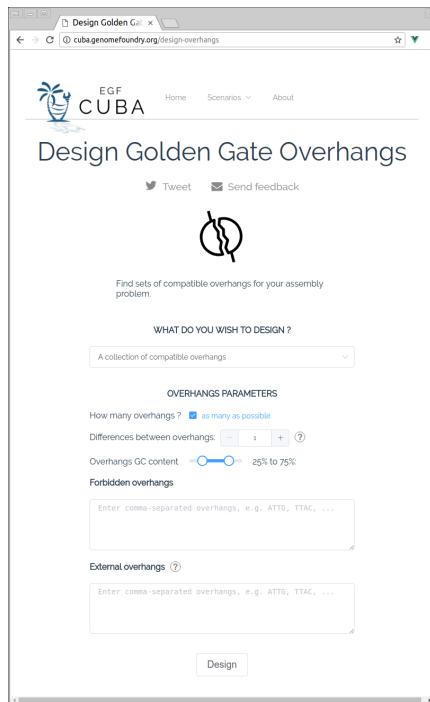


from Tsuge et al., Sci Rep. 2015

- ▶ 50 compatible overhangs
- ▶ Must be as equally spaced as possible in the sequence.

Overhang design software

cuba.genomefoundry.org/design-overhangs



Overhang design software

cuba.genomefoundry.org/design-overhangs

collection size
overhang similarity tolerance
mandatory / forbidden overhangs
Allowed GC%



The screenshot shows the 'Design Golden Gate Overhangs' page. On the left, there's a sidebar with a blue icon of a person with a brain, followed by the text 'EGF CUBA' and links for 'Home', 'Scenarios', and 'About'. Below this is the main title 'Design Golden Gate Overhangs' with social sharing buttons for Twitter and 'Send feedback'. The main area has a heading 'WHAT DO YOU WISH TO DESIGN ?' with a dropdown menu set to 'A collection of compatible overhangs'. Under 'OVERHANGS PARAMETERS', there are three sections: 'How many overhangs?', 'Differences between overhangs' (set to 1), and 'Overhangs GC content' (set to 25% To 75%). There are also two input fields: 'Forbidden overhangs' and 'External overhangs', both with placeholder text 'Enter comma-separated overhangs, e.g. ATTC, TTAC, ...'. At the bottom is a large blue 'Design' button.

Overhang design software

cuba.genomefoundry.org/design-overhangs

collection size
overhang similarity tolerance
mandatory / forbidden overhangs
Allowed GC%



The screenshot shows the 'Design Golden Gate Overhangs' page. At the top, there's a logo for 'EGF CUBA' and navigation links for 'Home', 'Scenarios', and 'About'. Below the header, there are social sharing buttons for 'Tweet' and 'Send feedback'. The main title is 'Design Golden Gate Overhangs' with a magnifying glass icon. A sub-instruction says 'Find sets of compatible overhangs for your assembly problem.' Under the title, there's a dropdown menu set to 'A collection of compatible overhangs'. The 'OVERHANGS PARAMETERS' section includes: 'How many overhangs?' with a radio button for 'as many as possible', 'Differences between overhangs' (set to 1), 'Overhangs GC content' (set to 25% to 75%), and a 'Forbidden overhangs' input field containing 'AAAG, AAC...'. Below that is an 'External overhangs' input field containing 'TTAG, TTCA, TTTC'. At the bottom is a 'Design' button.



AAAG, AAC...
AGTA, ATAC, ATCT, ATTG, CACG, CTGG, GACC, GCTG,
GGAG, GTCG, TAGA, TATG, TCAC, TCGG, TGCG, TGGC,
TTAG, TTCA, TTTC,

Overhang design software

cuba.genomefoundry.org/design-overhangs

collection size
overhang similarity tolerance
mandatory / forbidden overhangs
Allowed GC%



The screenshot shows the 'Design Golden Gate Overhangs' page. At the top, there's a logo for 'EGF CUBA' and navigation links for 'Home', 'Scenarios', and 'About'. Below that is a 'Twitter' button and a 'Send feedback' link. The main title is 'Design Golden Gate Overhangs' with a magnifying glass icon. A sub-instruction says 'Find sets of compatible overhangs for your assembly problem.' Under the title, there's a dropdown menu set to 'A collection of compatible overhangs'. The 'OVERHANGS PARAMETERS' section includes: 'How many overhangs?' with a radio button for 'as many as possible', 'Differences between overhangs' (set to 1), 'Overhangs GC content' (set to 25% to 75%), and 'Forbidden overhangs' (a text input field containing 'AAAG, AAC...'). Below that is an 'External overhangs' section with another text input field and a 'Design' button at the bottom.



AAAG, AAC...
AGTA, ATAC, ATCT, ATTG, CACG, CTGG, GACC, GCTG,
GGAG, GTCG, TAGA, TATG, TCAC, TCGG, TGCG, TGGC,
TTAG, TTCA, TTTC,

sequence
cutting regions
forbidden overhangs
Allowed GC%



Overhang design software

cuba.genomefoundry.org/design-overhangs

collection size
overhang similarity tolerance
mandatory / forbidden overhangs
Allowed GC%

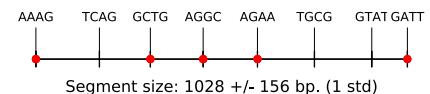


The screenshot shows the CUBA software interface for designing Golden Gate Overhangs. At the top, there's a logo for EGF CUBA, a navigation bar with Home, Scenarios, and About, and social sharing buttons for Twitter and Send feedback. Below that is a main title "Design Golden Gate Overhangs" with a magnifying glass icon. A sub-instruction says "Find sets of compatible overhangs for your assembly problem." Underneath, a dropdown menu says "A collection of compatible overhangs". The "OVERHANGS PARAMETERS" section includes a radio button for "as many as possible", a slider for "Differences between overhangs" from 1 to 75%, a range for "Overhangs GC content" from 25% to 75%, and input fields for "Forbidden overhangs" and "External overhangs", both with examples like ATTC, TTAC, etc. At the bottom is a "Design" button.

sequence
cutting regions
forbidden overhangs
Allowed GC%



AAAG, AACG, AAGT, AACG, ACCG, AGCC, AGGG,
AGTA, ATAC, ATCT, ATTG, CACG, CTGG, GACC, GCTG,
GGAG, GTCG, TAGA, TATG, TCAC, TCGG, TGCG, TGGC,
TTAG, TTCA, TTTC,



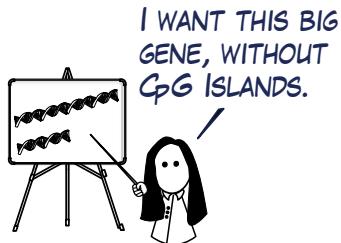
Sequence polishing: introductory exercise

Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

Sequence polishing: introductory exercise

Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

CUSTOMER



remove CG
keep protein

Sequence polishing: introductory exercise

Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

CUSTOMER

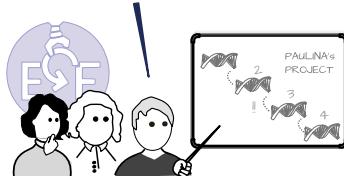
I WANT THIS BIG
GENE, WITHOUT
CpG ISLANDS.



remove CG
keep protein

EGF

LET'S REMOVE BSMBI SITES
AND INSERT A NCOI SITE



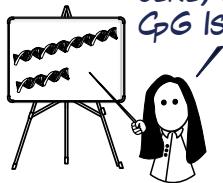
remove CGTCTC
insert CCATGG

Sequence polishing: introductory exercise

Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

CUSTOMER

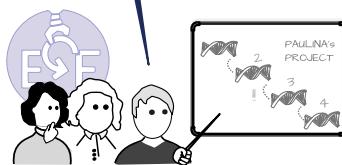
I WANT THIS BIG
GENE, WITHOUT
CpG ISLANDS.



remove CG
keep protein

EGF

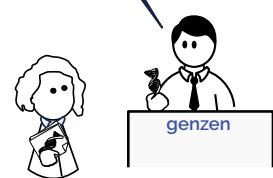
LET'S REMOVE BSMBI SITES
AND INSERT A NCOI SITE



remove CGTCTC
insert CCATGG

DNA PROVIDER

SORRY, WE CANNOT MAKE
PARTS WHICH CONTAIN
BSAI SITES.

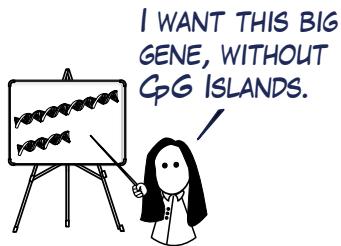


remove GGTCTC

Sequence polishing: introductory exercise

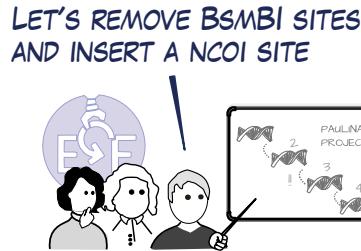
Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

CUSTOMER



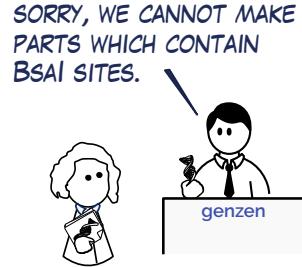
remove CG
keep protein

EGF



remove CGTCTC
insert CCATGG

DNA PROVIDER



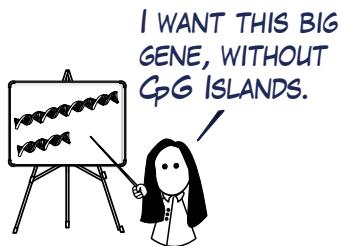
remove GGTCTC

M	S	S	G	G	S	R	L	H	G	V
A	T	G	A	G	T	C	G	C	G	A
A	G	C	A	G	T	G	G	A	G	T
T	C	A	T	C	A	G	G	G	G	C
T	C	C	T	C	C	G	G	T	T	T
T	C	G	T	C	G	T	C	C	T	A
T	C	T	T	C	T	C	C	G	T	T

Sequence polishing: introductory exercise

Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

CUSTOMER



remove CG
keep protein

EGF

LET'S REMOVE BSMBI SITES
AND INSERT A NCOI SITE



remove CGTCTC
insert CCATGG

DNA PROVIDER

SORRY, WE CANNOT MAKE PARTS WHICH CONTAIN BSAI SITES.



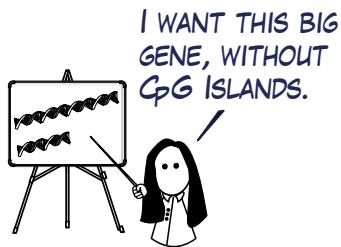
remove GGTCTC

M	S	S	G	G	S	R	L	H	G	V
A	T	G	A	G	C	G	T	C	A	C
A	G	C	A	G	T	G	G	A	G	T
T	C	A	T	C	A	G	G	C	G	C
T	C	C	T	C	C	G	G	T	G	T
T	C	G	T	C	G	T	C	C	G	C
T	C	T	T	C	T	C	C	G	T	T

Sequence polishing: introductory exercise

Sequence specifications can depend on its purpose, assembly method, synthesis limitations, and validation method.

CUSTOMER



remove CG
keep protein

EGF

LET'S REMOVE BSMBI SITES
AND INSERT A NCOI SITE



remove CGTCTC
insert CCATGG

DNA PROVIDER

SORRY, WE CANNOT MAKE PARTS WHICH CONTAIN BSAI SITES.



remove GGTCTC

M	S	S	G	G	S	R	L	H	G	V
A	T	G	A	G	C	G	T	C	A	T
A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A	A G C A G T G G A G G A A G C A G A C T A G G C C T A
T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G	T C A T T C A G G G C G G G C A G T A G G G C C T T G
T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T	T C C T T C C G G G G G G T T C A C G G A C C T T T
T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A	T C G T T C G T T C G T T C G T T C G T T C G T T A
T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G	T C T T T C T T C T T C T T C T T C T T C T T G

Real-world sequence polishing examples

Real-world sequence polishing examples

- ▶ *I want a 60,000bp sequence that is orthogonal to yeast, but same GC%*

Real-world sequence polishing examples

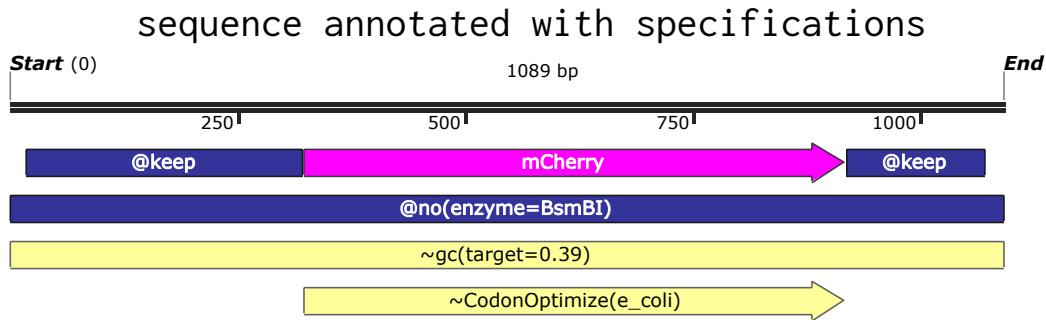
- ▶ *I want a 60,000bp sequence that is orthogonal to yeast, but same GC%*
- ▶ *I want this 50,000bp chromosome segment, but freed from as many common 4bp restriction sites as possible. Conserve ORFs if possible.*

Real-world sequence polishing examples

- ▶ I want a 60,000bp sequence that is orthogonal to yeast, but same GC%
- ▶ I want this 50,000bp chromosome segment, but freed from as many common 4bp restriction sites as possible. Conserve ORFs if possible.
- ▶ I want to domesticate these genes for E. coli. They should also pass the constraints of Synthesis Company (no homopolymers, no hairpins, no repeats)

Sequence polishing software

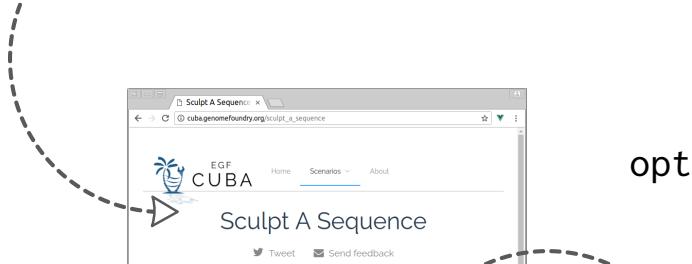
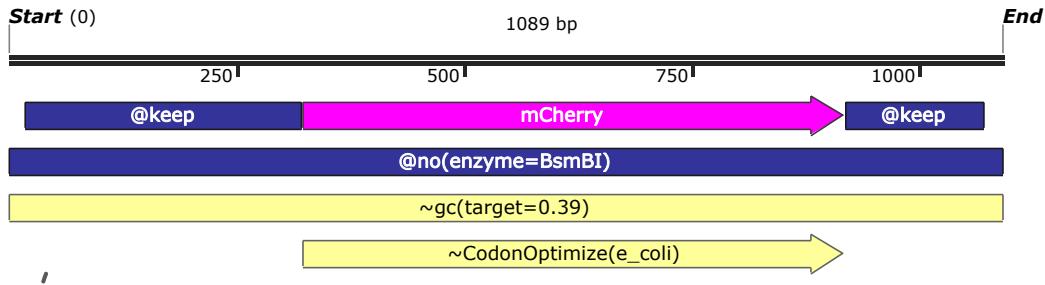
cuba.genomefoundry.org/sculpt_a_sequence



Sequence polishing software

cuba.genomefoundry.org/sculpt_a_sequence

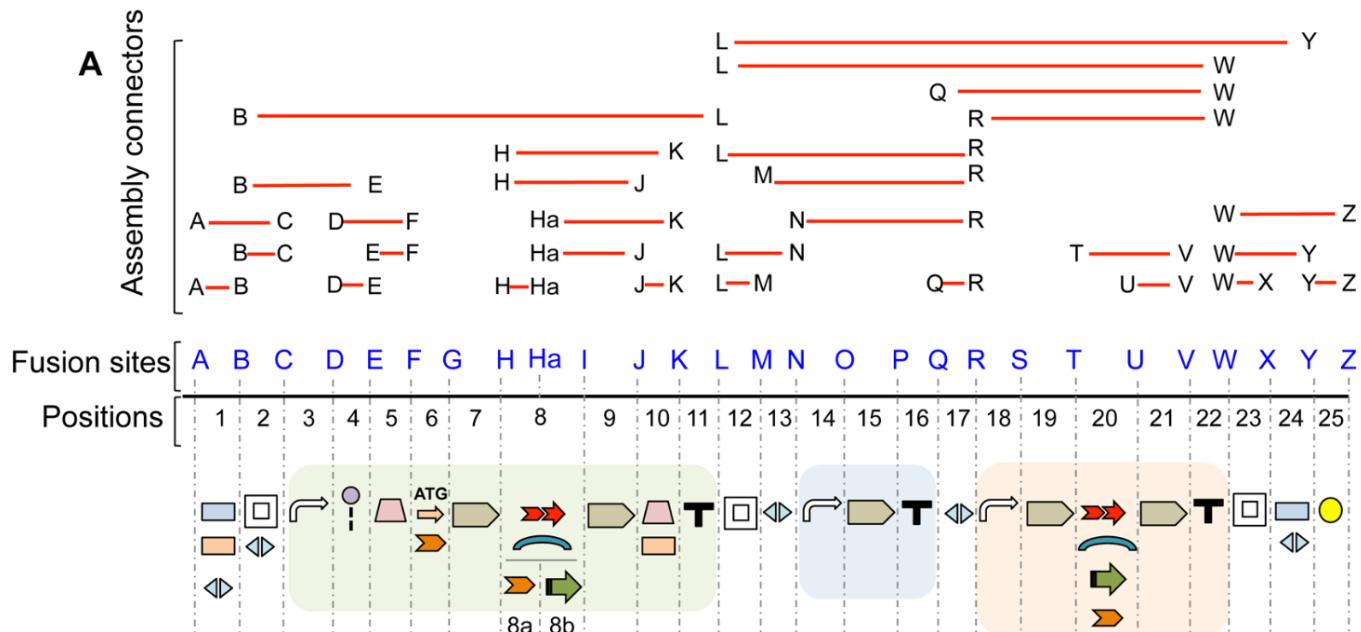
sequence annotated with specifications



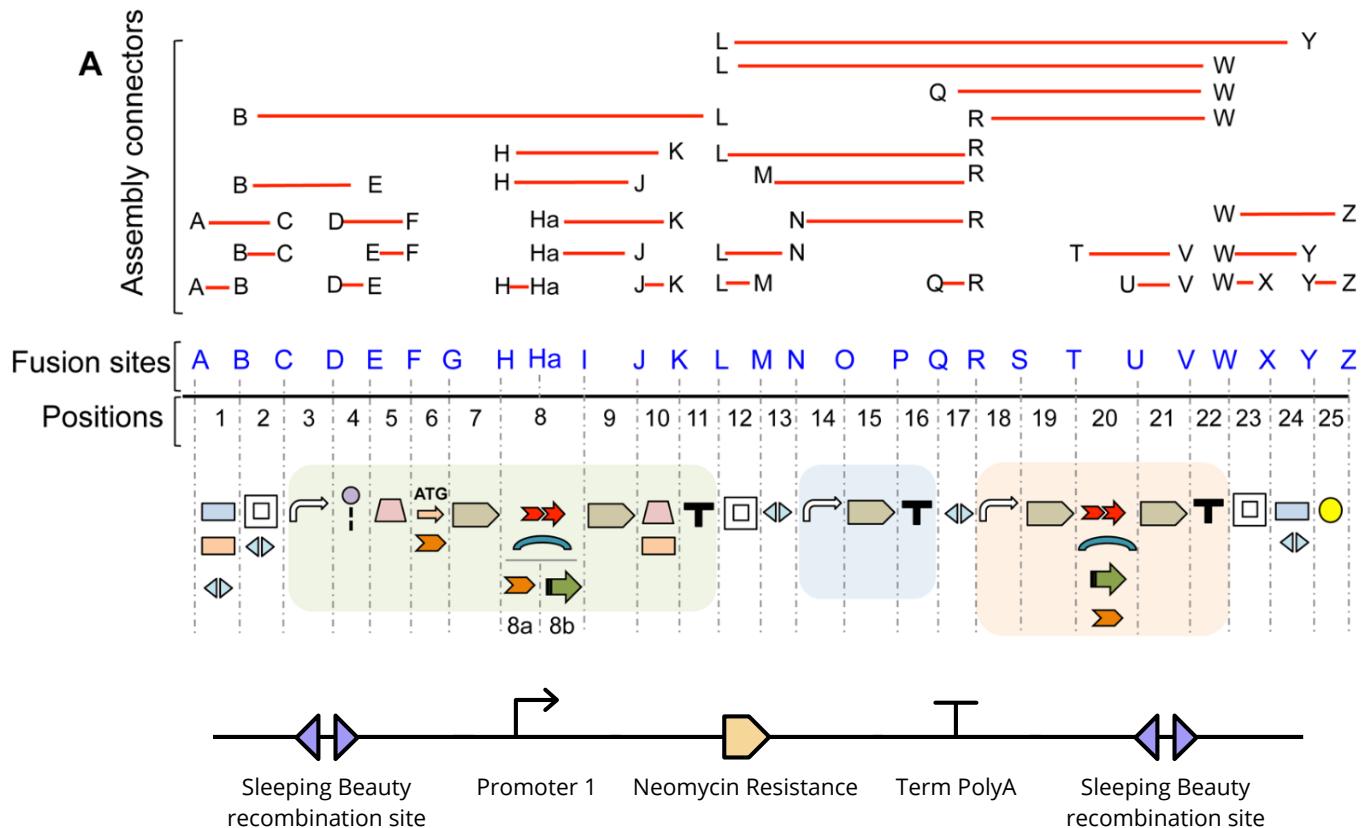
optimized sequence



Parts based assembly: easy peasy ?

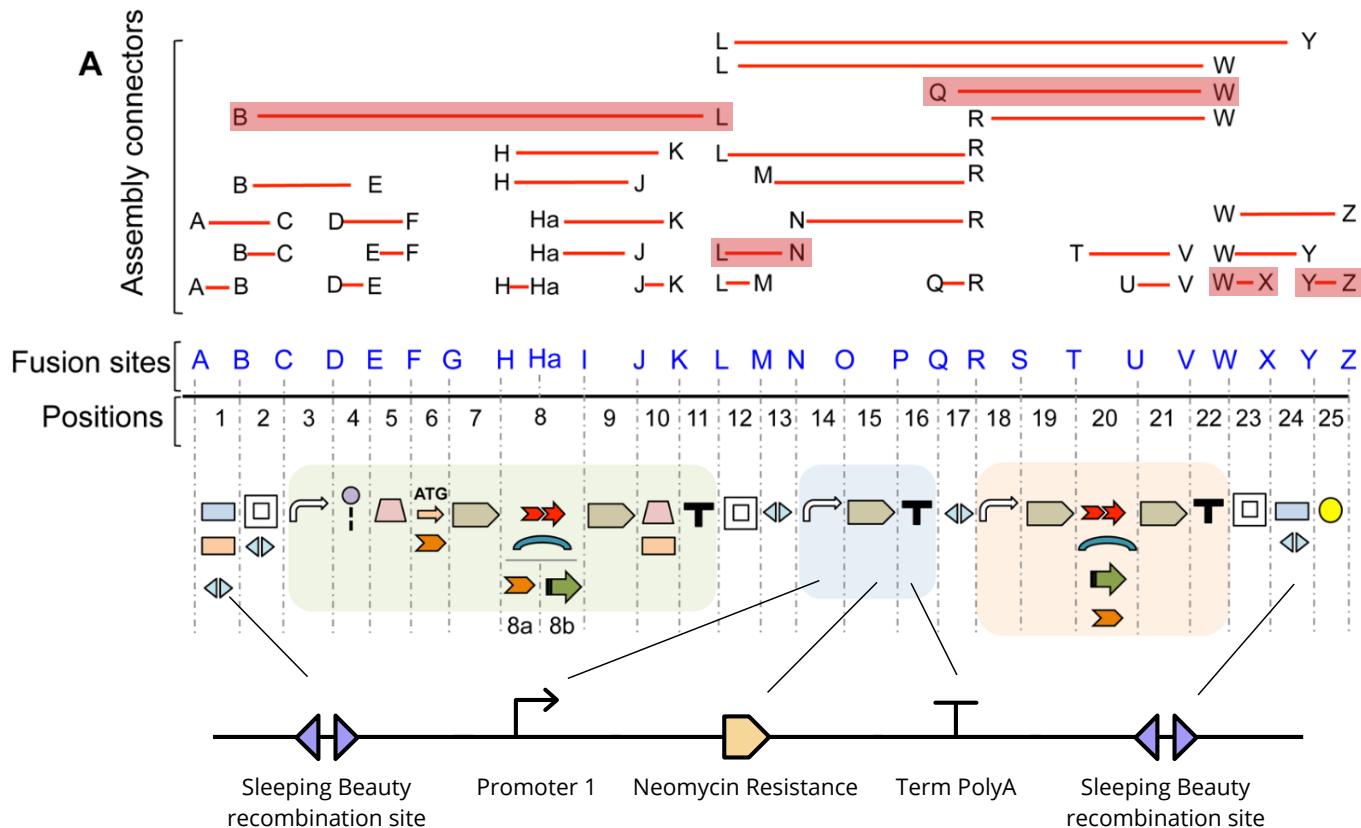


Parts based assembly: easy peasy ?



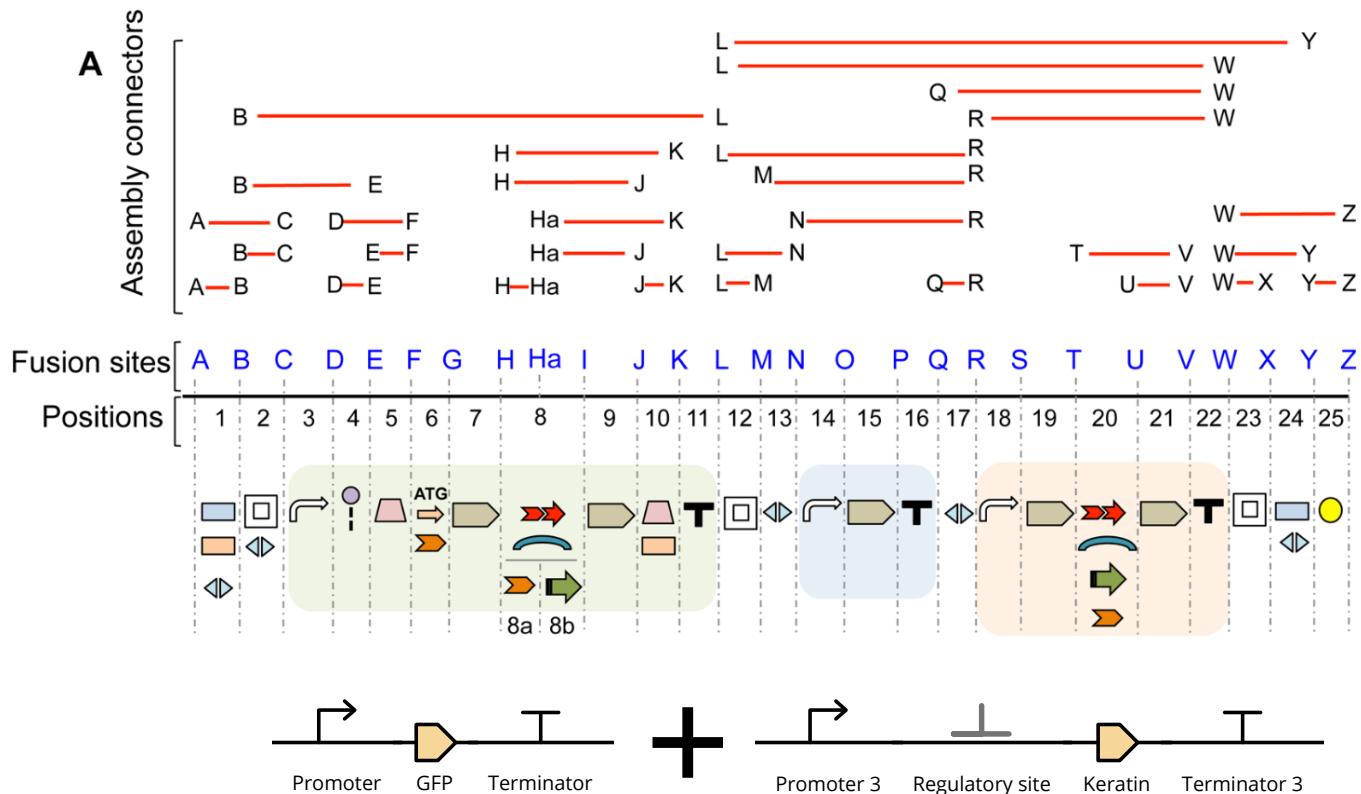
Question: Where would you place the parts ?

Parts based assembly: easy peasy ?



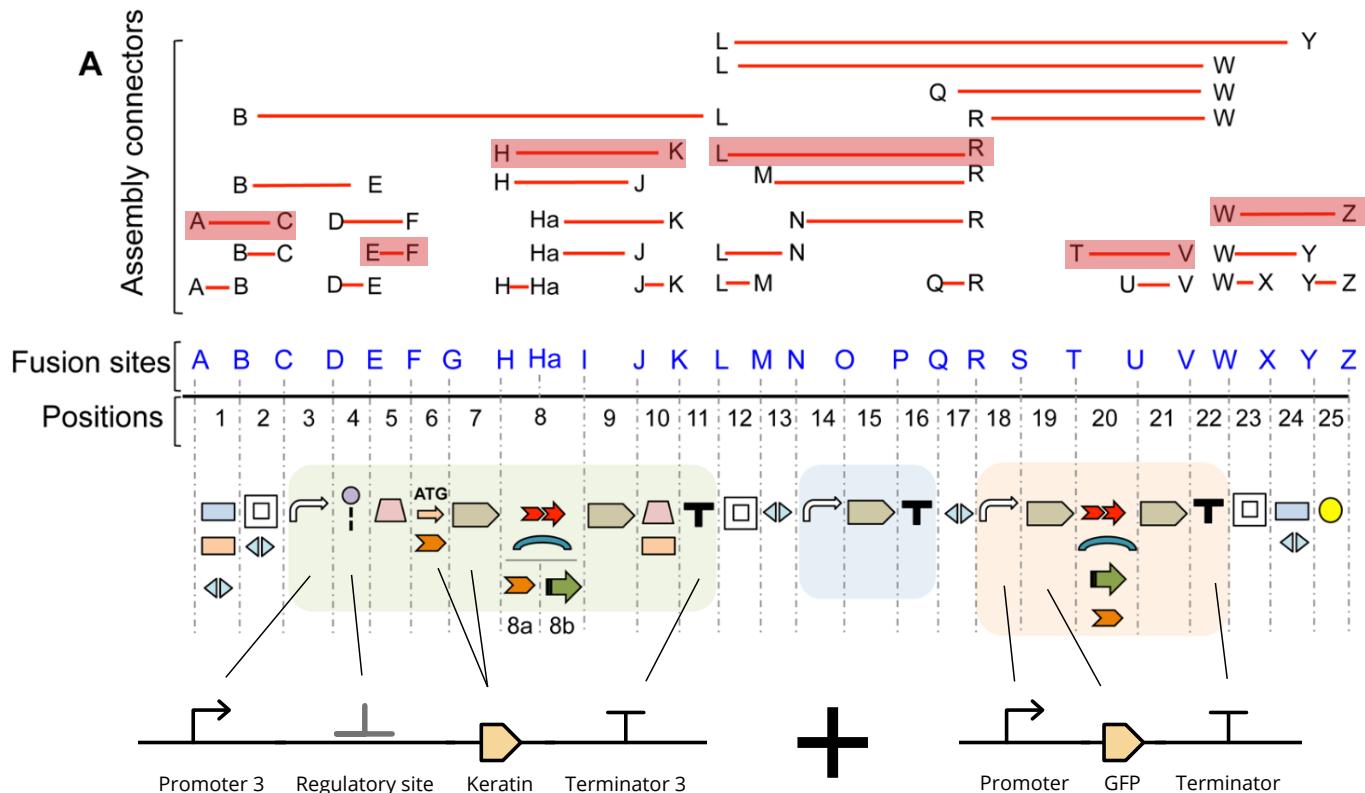
Question: Where would you place the parts ?

Parts based assembly: easy peasy ?



Question: Where would you place the parts ?

Parts based assembly: easy peasy ?



Question: Where would you place the parts ?

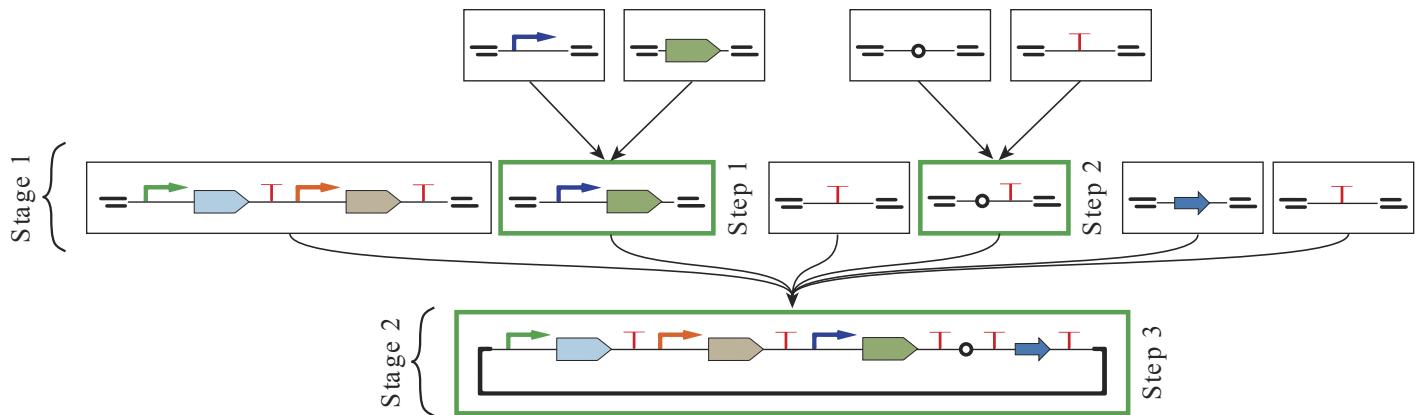
Part-based design with Autodesk Genetic Constructor

The screenshot displays the Autodesk Genetic Constructor software interface, showing four different genetic constructs:

- 5 - Episomal, single bicistronic unit**: A grid with positions 3 through 11. Positions 3, 6, 7, 8a, 9, and 11 contain CAGp. Positions 6 and 7 contain ATG-BoxC (LTae) and Ni-IgGL sequence. Position 7 also contains mTgFP2. Positions 8a and 9 contain L7Ae - Weiss, mNeonGreen, and mRuby2. Position 8a also contains mTgFP2. Position 9 also contains DmRc. Position 11 contains SV40 polyA.
- 6A - Episomal, Two bicistronic units**: A grid with positions 3 through 18. Positions 3, 6, 7, 8, 9, 11, and 18 contain CAGp. Positions 6 and 7 contain ATG-BoxC (LTae) and Ni-IgGL sequence. Position 7 also contains mTgFP2. Position 8 contains L7Ae and mNeonGreen. Position 9 contains Linker 2 and FcFy Luciferase -3XFLAG. Position 11 contains SV40 polyA. Position 18 contains CMVp.
- 2A - Episomal, single tagged fusion protein**: A grid with positions 3 through 9. Positions 3, 6, 7, 8, and 9 contain CAGp. Positions 6 and 7 contain Ni-IgGL sequence, EF1αp, Ni-SV40-NLS, Ni-myristylation signal, 3X-FLAG, ATG-BoxC (LTae), Ni-Palmitoylation sequence, Kozak, ATG, and Ni-MLS. Position 7 also contains mTgFP2. Position 8 contains L7Ae, L7Ae - Weiss, FcFy Luciferase, mRuby2, Bx81, mKate2, and mNeonGreen. Position 9 contains p2a Porcine teschovirus-1, Linker 2, Linker 1, mNeonGreen, DmRc, BSDR, NeoR, PuoR, α-Tubulin, mRuby2, and mTagBFP2.
- 1B - Single tagged protein**: A grid with positions 3 through 11. Positions 3, 6, 7, and 11 contain CAGp. Positions 6 and 7 contain ATG-BoxC (LTae) and Kozak_ATG.

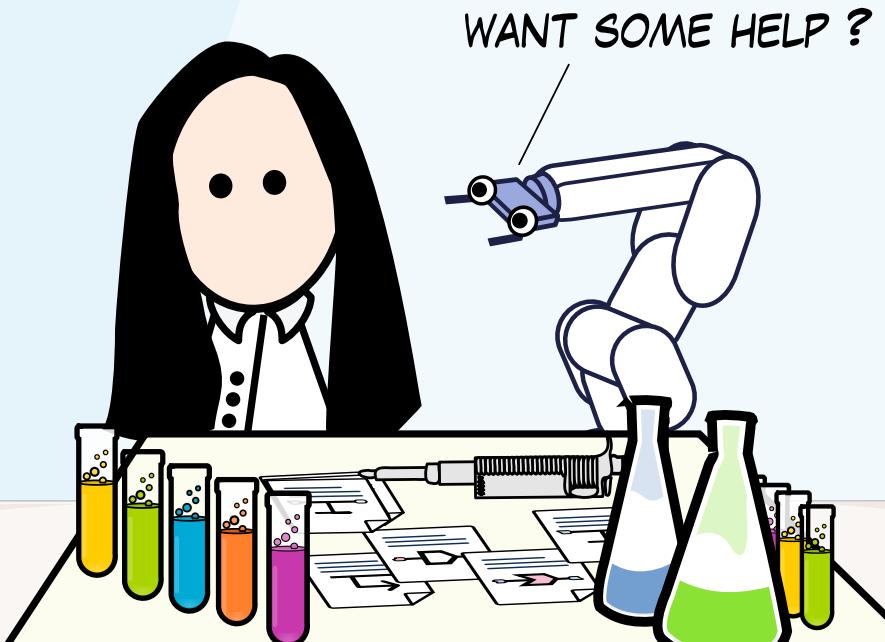
The interface includes a sidebar for "Projects" and "By Project" search, and tabs for "SEQUENCE" and "GSL EDITOR".

Automated part selection with RavenCAD

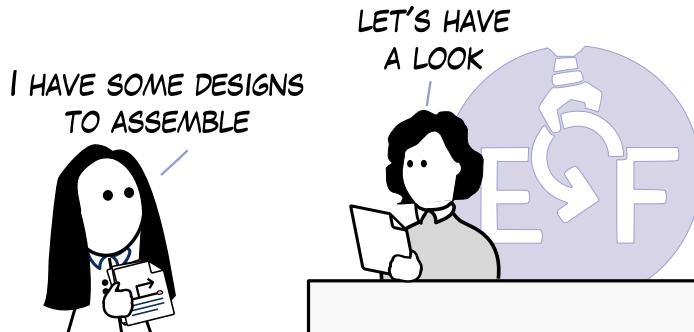


From Appleton et al., Nat Methods 2014

2. Assembly

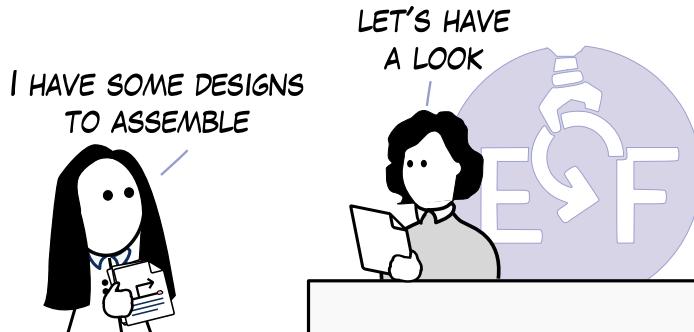


Assembly pre-validation



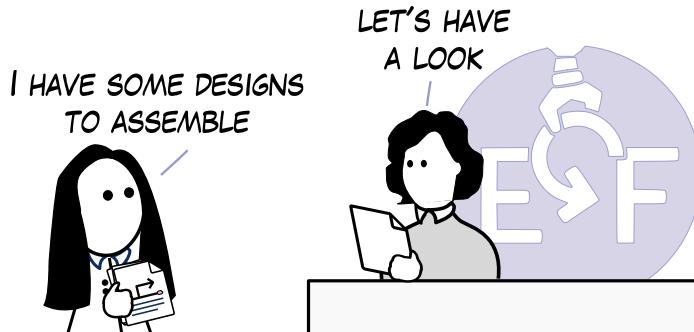
Assembly 1	promoter_PKA	neogreen	term_pkA	insulator	HA_0001	HC_amp
Assembly 2	promoter_PkA	mCherry	term_pkA	insulator	HA_1002	HC_amp
Assembly 3	promoter_PKA	YFP	term_pkA	insulator	HA_0202	HC_amp
Assembly 4	promoter_AR1	neogreen	UTR53	insulator	HA_0003	HC_amp
Assembly 5	promoter_AR1	neogreen	neogreen	term_pkA	HA_0203	HC_amp
Assembly 6	promoter_ZDF	YFP	UTR53	term_pkA	insulator	HA_0001
Assembly 7	promoter_ZDF	mCherry	term_pkA	insulator	HA_0003	HC_amp
Assembly 8	promoter_ZDF	neogreen	term_pkA	insulator	HA_0401	HC_amp

Assembly pre-validation



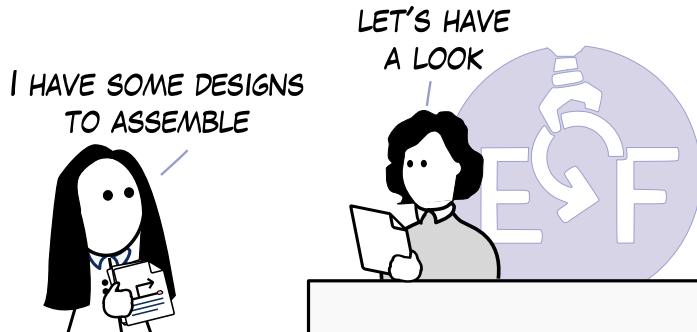
Assembly 1	promoter_PKA	neogreen	term_pkA	insulator	HA_0001	HC_amp
Assembly 2	promoter_PkA	mCherry TYPO?	term_pkA	insulator	HA_1002	HC_amp
Assembly 3	promoter_PKA	YFP	term_pkA	insulator	HA_0202	HC_amp
Assembly 4	promoter_AR1	neogreen	UTR53	insulator	HA_0003	HC_amp
Assembly 5	promoter_AR1	neogreen	neogreen	term_pkA	HA_0203	HC_amp
Assembly 6	promoter_ZDF	YFP	UTR53	term_pkA	insulator	HA_0001
Assembly 7	promoter_ZDF	mCherry	term_pkA	insulator	HA_0003	HC_amp
Assembly 8	promoter_ZDF	neogreen	term_pkA	insulator	HA_0401	HC_amp

Assembly pre-validation



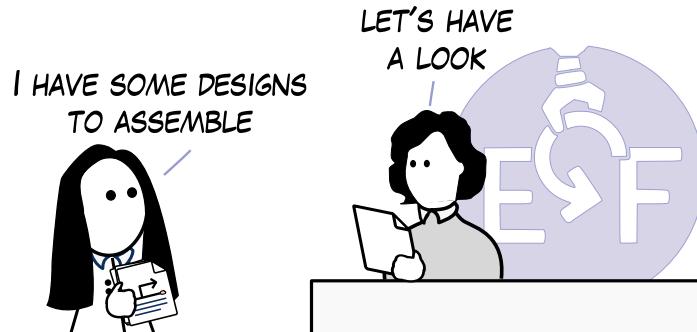
Assembly 1	promoter_PKA	neogreen	term_pkA	insulator	HA_0001	HC_amp
Assembly 2	promoter_PkA	mCherry	term_pkA	insulator	HA_1002	HC_amp
Assembly 3	promoter_PKA	YFP	term_pkA	insulator	HA_0202	HC_amp
Assembly 4	promoter_AR1	neogreen	UTR53	insulator NO TERMINATOR?	HA_0003	HC_amp
Assembly 5	promoter_AR1	neogreen	neogreen	term_pkA	HA_0203	HC_amp
Assembly 6	promoter_ZDF	YFP	UTR53	term_pkA	insulator	HA_0001
Assembly 7	promoter_ZDF	mCherry	term_pkA	insulator	HA_0003	HC_amp
Assembly 8	promoter_ZDF	neogreen	term_pkA	insulator	HA_0401	HC_amp

Assembly pre-validation



Assembly 1	promoter_PKA	neogreen	term_pkA	insulator	HA_0001	HC_amp
Assembly 2	promoter_PkA	mCherry ↖TYPO?	term_pkA	insulator	HA_1002	HC_amp
Assembly 3	promoter_PKA	YFP	term_pkA	insulator	HA_0202	HC_amp
Assembly 4	promoter_AR1	neogreen	UTR53 NO TERMINATOR?	insulator	HA_0003	HC_amp
Assembly 5	promoter_AR1	neogreen DUPLICATE?	↔ neogreen	term_pkA	HA_0203	HC_amp
Assembly 6	promoter_ZDF	YFP	UTR53	term_pkA	insulator	HA_0001
Assembly 7	promoter_ZDF	mCherry	term_pkA	insulator	HA_0003	HC_amp
Assembly 8	promoter_ZDF	neogreen	term_pkA	insulator	HA_0401	HC_amp

Assembly pre-validation



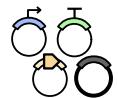
Assembly 1	promoter_PkA	neogreen	term_pkA	insulator	HA_0001	HC_amp
Assembly 2	promoter_PkA	mCherry ↖TYPO?	term_pkA	insulator	HA_1002	HC_amp
Assembly 3	promoter_PkA	YFP	term_pkA	insulator	HA_0202	HC_amp
Assembly 4	promoter_AR1	neogreen	UTR53	insulator NO TERMINATOR?	HA_0003	HC_amp
Assembly 5	promoter_AR1	neogreen	neogreen DUPLICATE?	term_pkA	HA_0203	HC_amp
Assembly 6	promoter_ZDF	YFP	UTR53	term_pkA	insulator	HA_0001 X NO BACKBONE?
Assembly 7	promoter_ZDF	mCherry	term_pkA	insulator	HA_0003	HC_amp
Assembly 8	promoter_ZDF	neogreen	term_pkA	insulator	HA_0401	HC_amp

Assembly pre-validation with simulated cloning

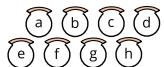
Assembly pre-validation with simulated cloning

SIMULATED CLONING

GENETIC PARTS



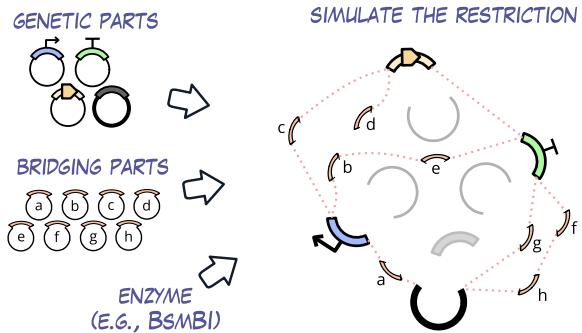
BRIDGING PARTS



ENZYME
(E.G., BSMBI)

Assembly pre-validation with simulated cloning

SIMULATED CLONING



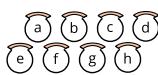
Assembly pre-validation with simulated cloning

SIMULATED CLONING

GENETIC PARTS

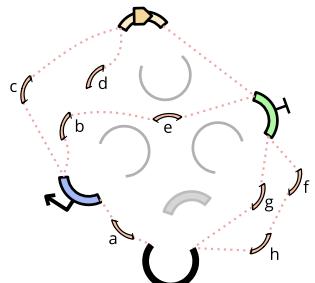


BRIDGING PARTS

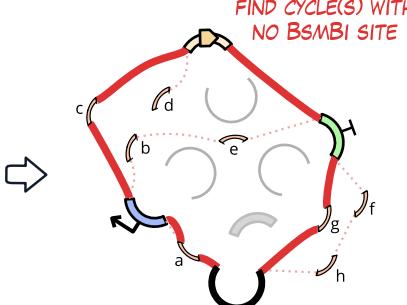


ENZYME
(E.G., BSMBI)

SIMULATE THE RESTRICTION



FIND CYCLE(S) WITH
NO BSMBI SITE



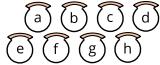
Assembly pre-validation with simulated cloning

SIMULATED CLONING

GENETIC PARTS

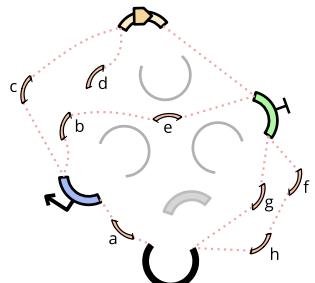


BRIDGING PARTS

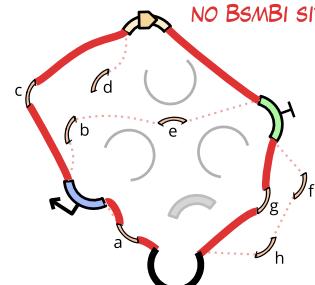


ENZYME
(E.G., BSMBI)

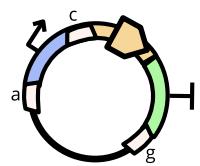
SIMULATE THE RESTRICTION



FIND CYCLE(S) WITH
NO BSMBI SITE



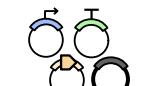
COMPUTE THE FINAL
ASSEMBLY SEQUENCES



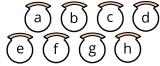
Assembly pre-validation with simulated cloning

SIMULATED CLONING

GENETIC PARTS

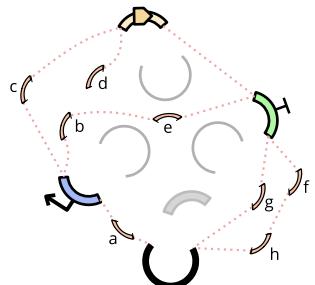


BRIDGING PARTS

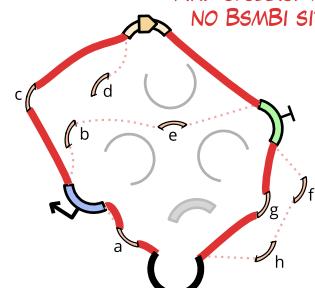


ENZYME
(E.G., BSMBI)

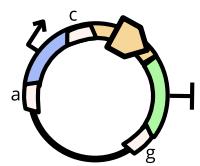
SIMULATE THE RESTRICTION



FIND CYCLE(S) WITH
NO BSMBI SITE



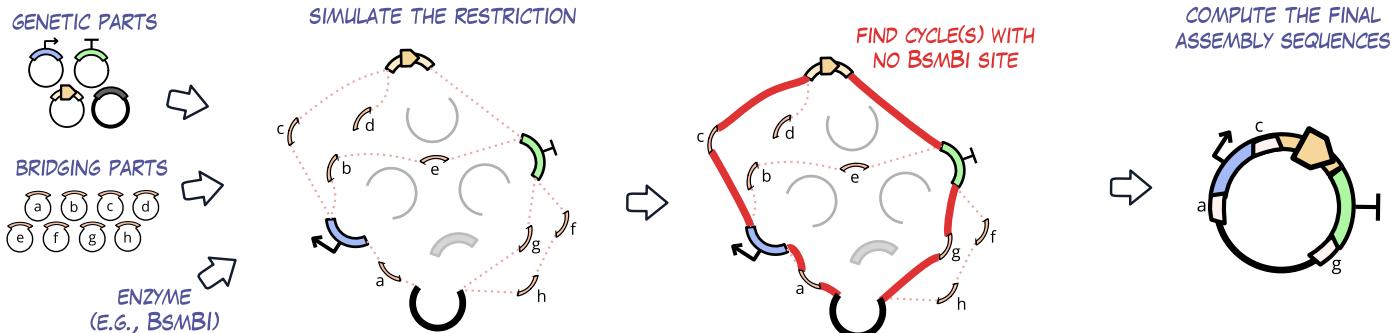
COMPUTE THE FINAL
ASSEMBLY SEQUENCES



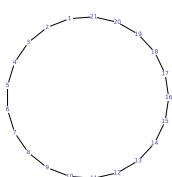
PARTS CONNECTIONS GRAPH

Assembly pre-validation with simulated cloning

SIMULATED CLONING



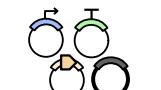
PARTS CONNECTIONS GRAPH



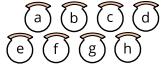
Assembly pre-validation with simulated cloning

SIMULATED CLONING

GENETIC PARTS

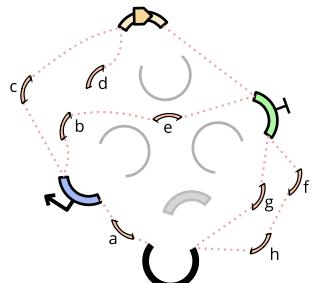


BRIDGING PARTS

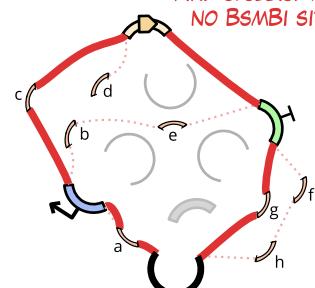


ENZYME
(E.G., BSMBI)

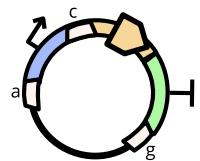
SIMULATE THE RESTRICTION



FIND CYCLE(S) WITH
NO BSMBI SITE

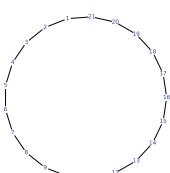


COMPUTE THE FINAL
ASSEMBLY SEQUENCES

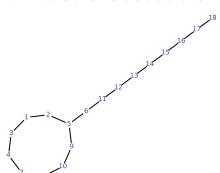


PARTS CONNECTIONS GRAPH

VALID ASSEMBLY

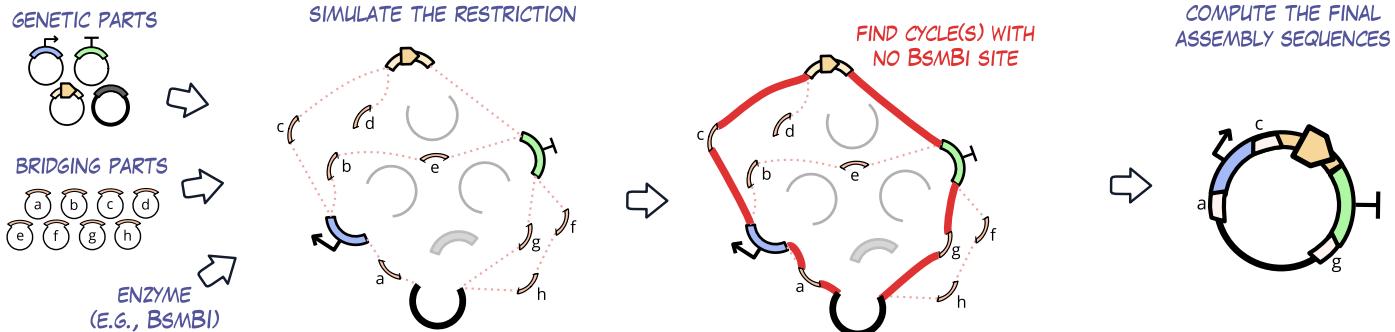


ONE PART HAS
A WRONG OVERHANG

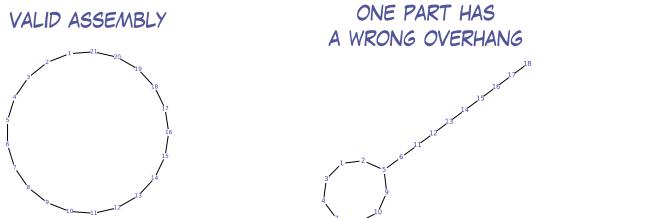


Assembly pre-validation with simulated cloning

SIMULATED CLONING



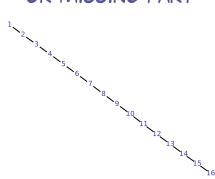
PARTS CONNECTIONS GRAPH



ONE PART HAS A WRONG OVERHANG

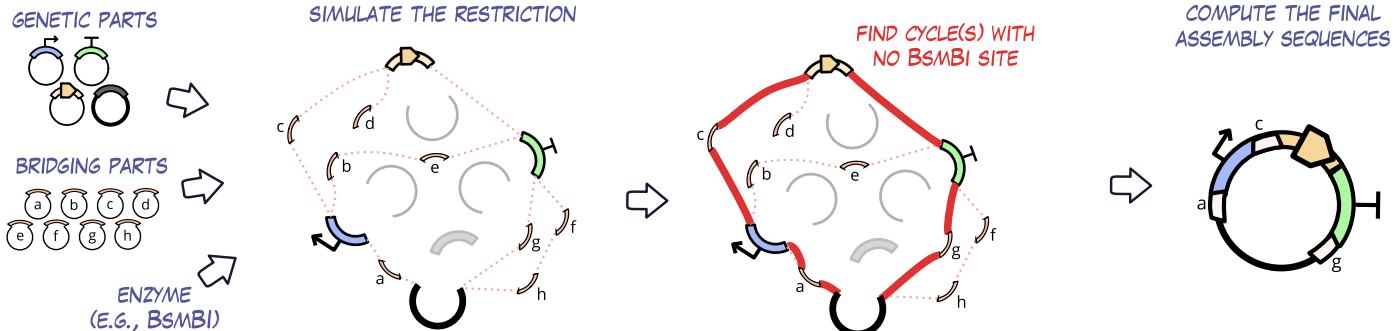


WRONG OVERHANG OR MISSING PART

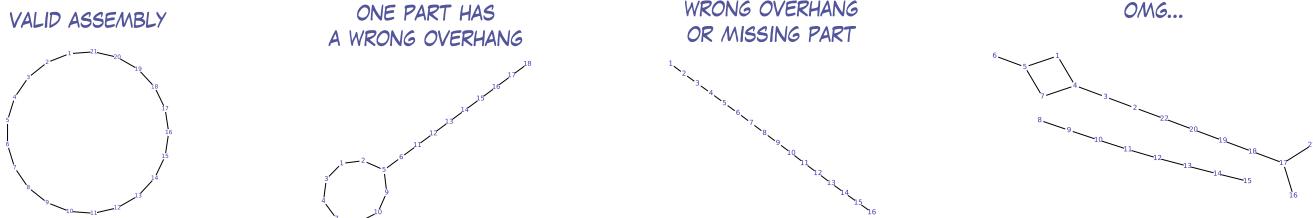


Assembly pre-validation with simulated cloning

SIMULATED CLONING

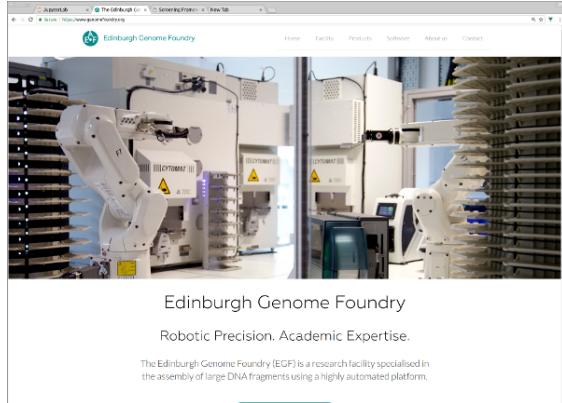


PARTS CONNECTIONS GRAPH

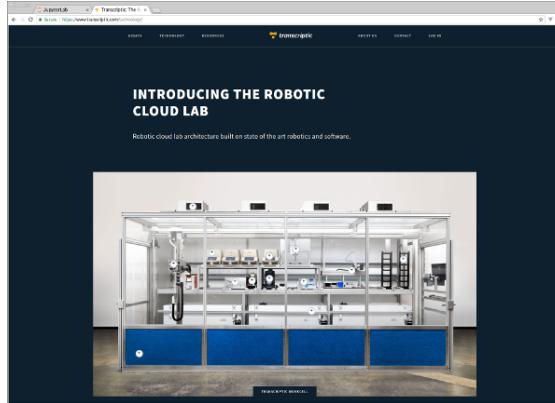


No hands ! DNA assembly in the cloud.

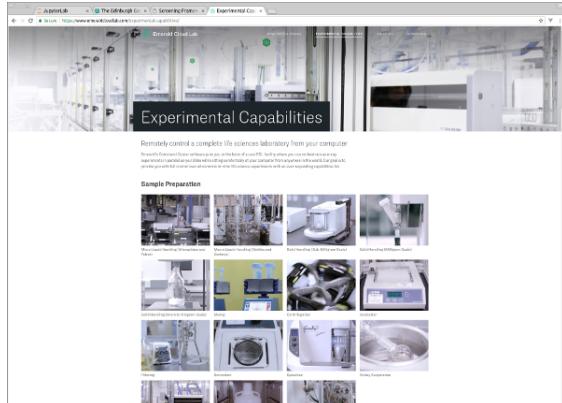
Edinburgh Genome Foundry



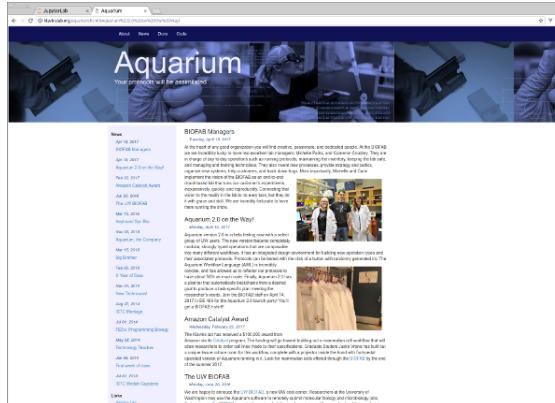
Transcriptic



Emerald Therapeutics

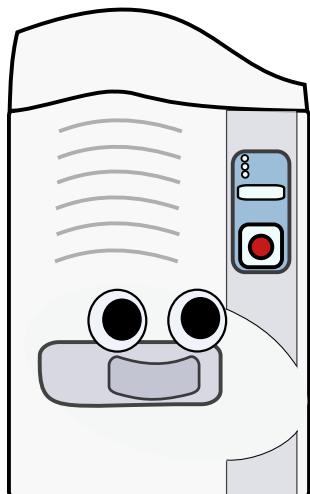


Aquarium



Assembly time !

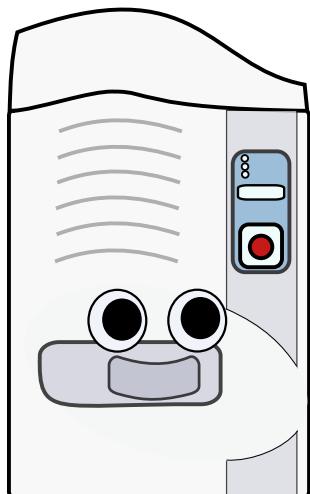
Question: What does this picklist do ?



Source Plate	Source Well	Destination Plate	Destination well	Volume (nL)
P1	A1	P2	A1	50
P1	B3	P2	A1	21
P1	B5	P2	A1	48
P1	C12	P2	A1	20
P1	A1	P2	A2	40
P1	F11	P2	A2	48
P1	G2	P2	A2	25
P1	G5	P2	A2	21
P1	A1	P2	A3	50
P1	B5	P2	A3	24
P1	B6	P2	A3	81
P1	B7	P2	A3	20

Assembly time !

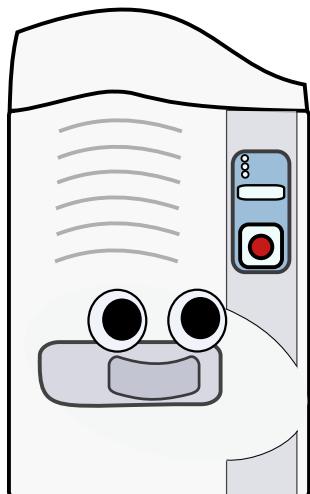
Question: What does this picklist do ?



Source Plate	Source Well	Destination Plate	Destination well	Volume (microL)
P1	A1	P2	A1	100
P1	A5	P2	B1	100
P1	B5	P2	C1	100
P1	B12	P2	D1	100
P1	C10	P2	E1	100
P1	C11	P2	F1	100
P1	C12	P2	G1	100
P1	F3	P2	H1	100
P1	G1	P2	A2	100
P1	G2	P2	B2	100
P1	H11	P2	C2	100
P1	H12	P2	D2	100

Assembly time !

Question: What does this picklist do ?



Source Plate	Source Well	Destination Plate	Destination well	Volume (microL)
P1	A1	P1	B1	25
Trough	--	P1	B1	75
P1	B1	P1	C1	25
Trough	--	P1	C1	75
P1	C1	P1	D1	25
Trough	--	P1	D1	75
P1	D1	P1	E1	25
Trough	--	P1	E1	75
P1	E1	P1	F1	25
Trough	--	P1	F1	75
P1	F1	P1	G1	25
Trough	--	P1	G1	75

Remarks on automated pipetting

Remarks on automated pipetting

- ▶ Automated pipetting improves precision, reproducibility, speed, costs.

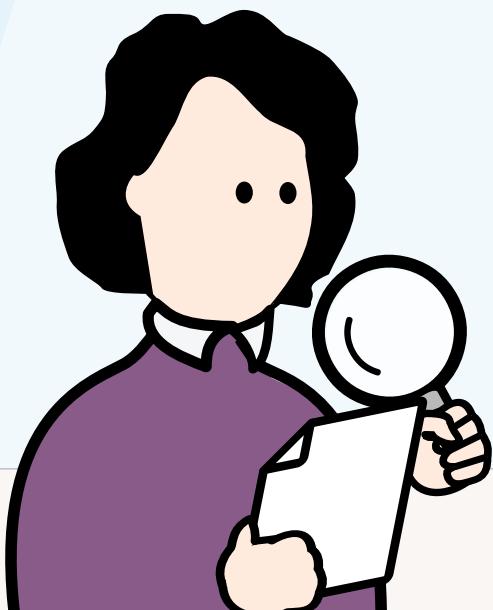
Remarks on automated pipetting

- ▶ Automated pipetting improves precision, reproducibility, speed, costs.
- ▶ It does **not** resolve the problem's complexity (you still need to write picklists), or the risk of error.

Remarks on automated pipetting

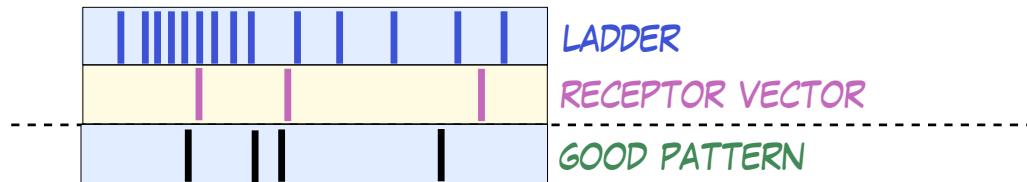
- ▶ Automated pipetting improves precision, reproducibility, speed, costs.
- ▶ It does **not** resolve the problem's complexity (you still need to write picklists), or the risk of error.
- ▶ See PR-PR, Antha, Plateo, for examples of software helping with picklist generation.

3. Validation



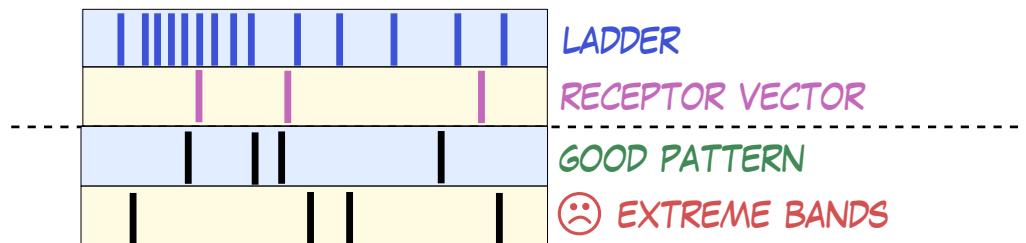
Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



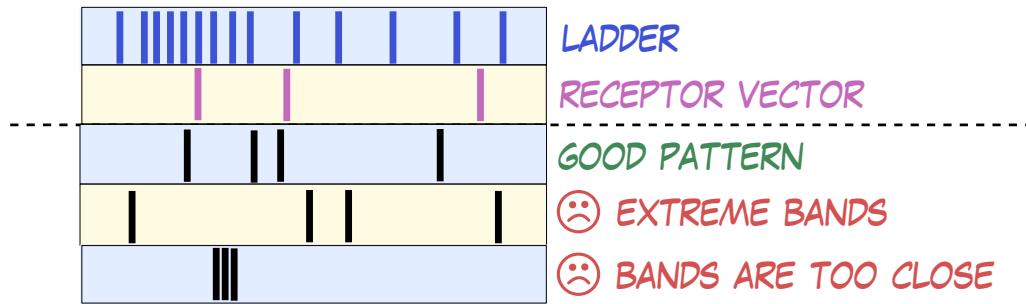
Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



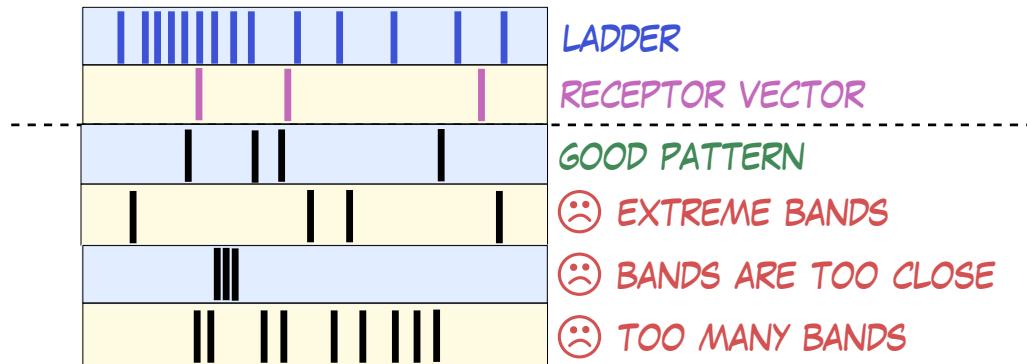
Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



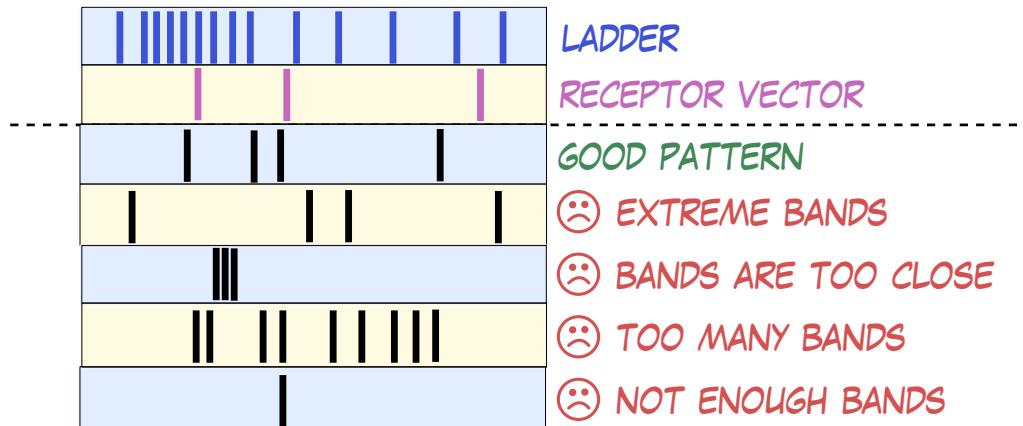
Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



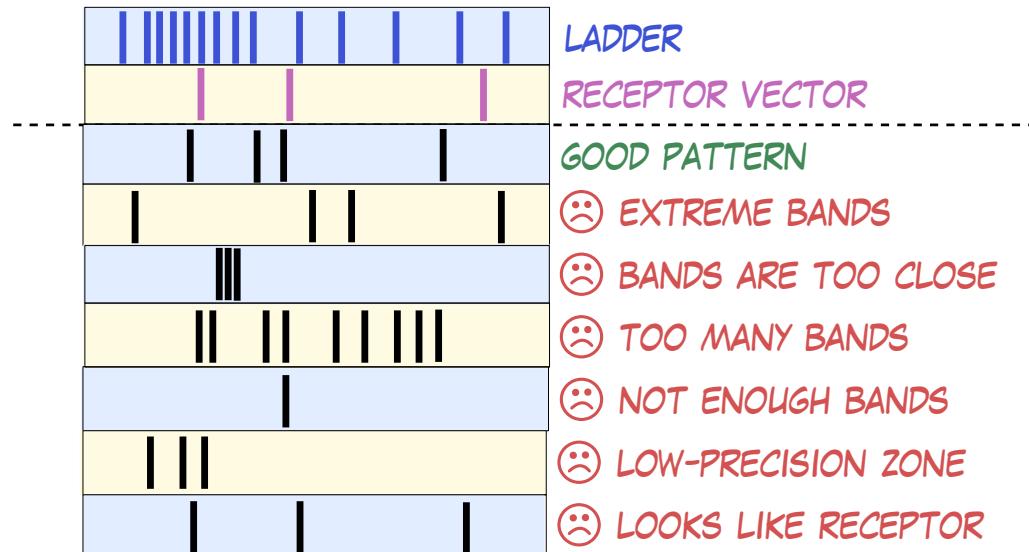
Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



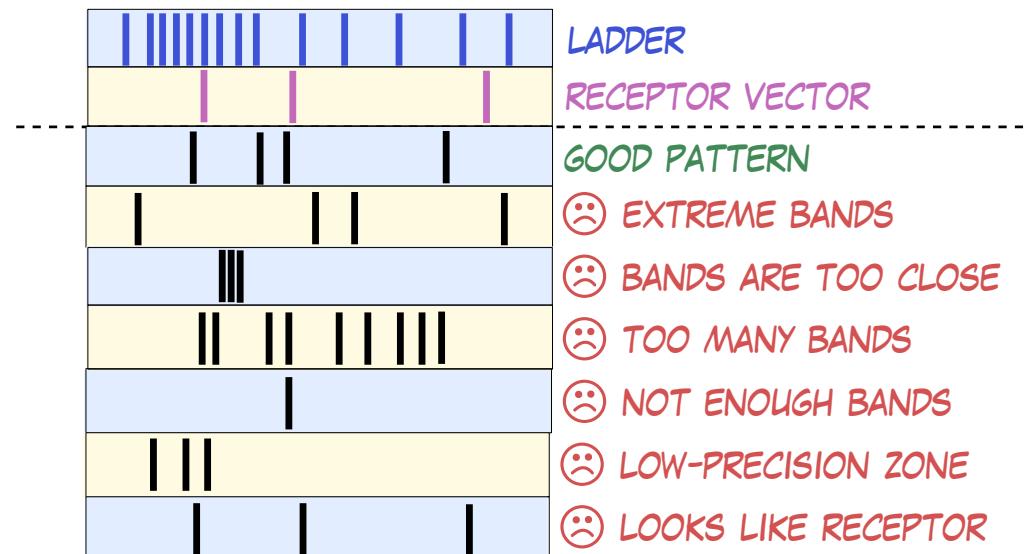
Computer-aided restriction digests

Shoot yourself in the foot with a bad choice of enzymes !



Computer-aided restriction digests

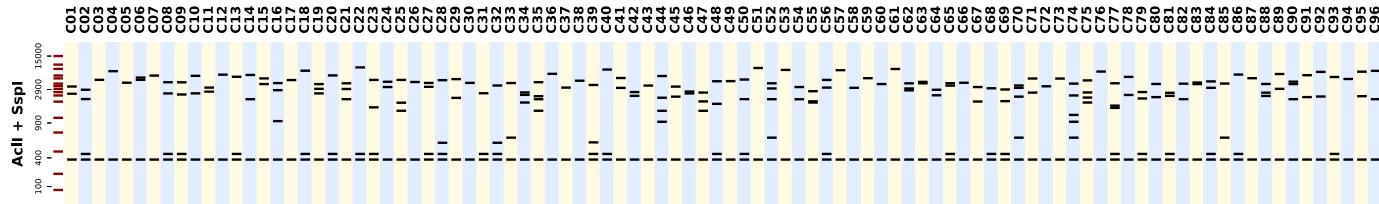
Shoot yourself in the foot with a bad choice of enzymes !



Other problems: methylation-sensitive enzymes, enzymes with incompatible activation temperatures.

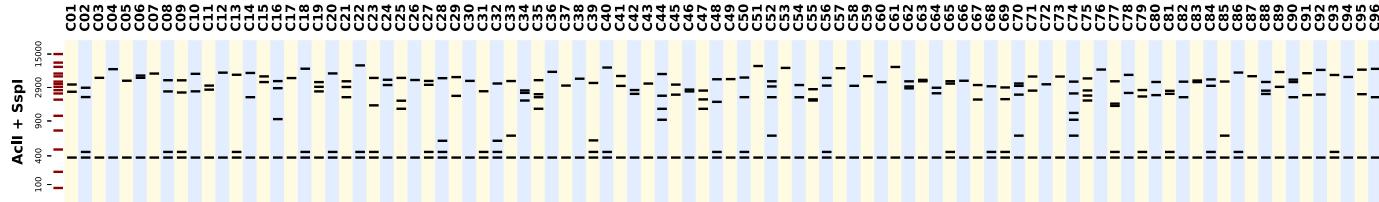
EGF restriction digest software

AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION

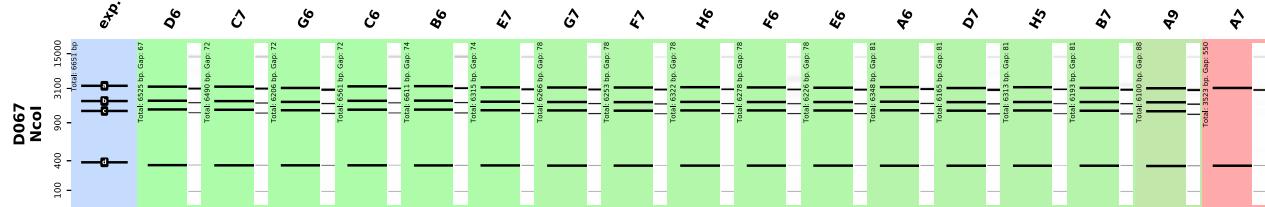


EGF restriction digest software

AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION

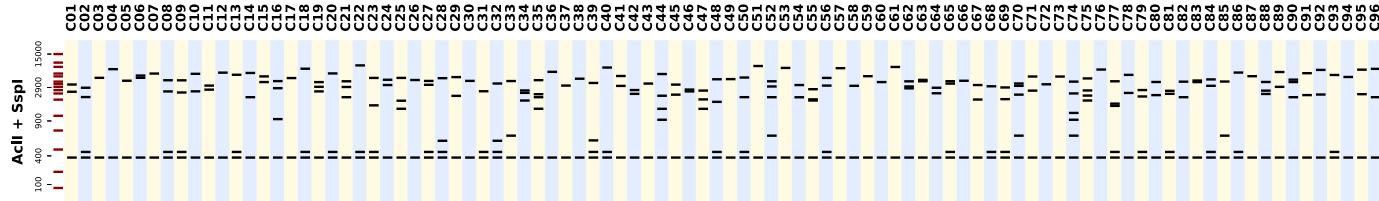


EXTENSIVE REPORT GENERATION FOR CONSTRUCT VALIDATION

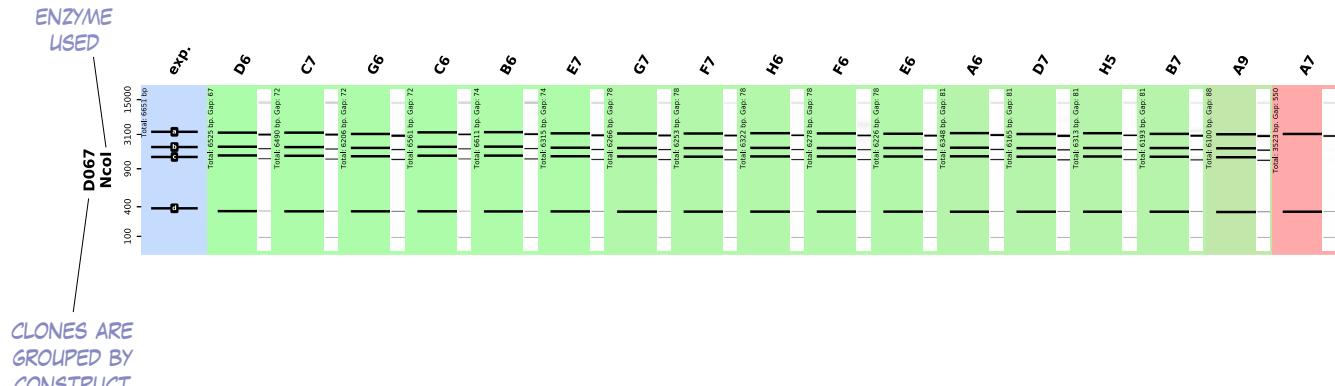


EGF restriction digest software

AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION

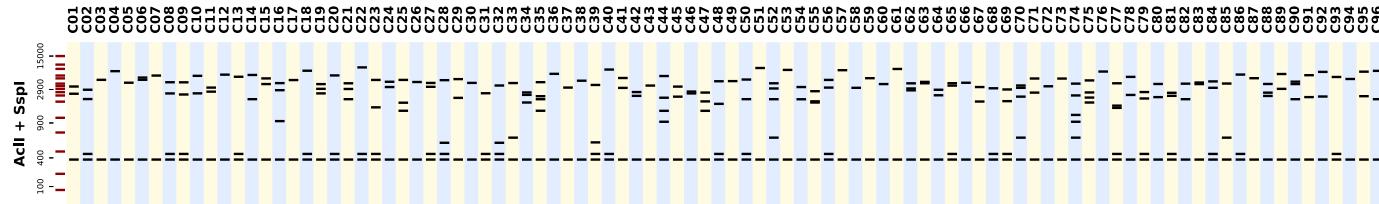


EXTENSIVE REPORT GENERATION FOR CONSTRUCT VALIDATION

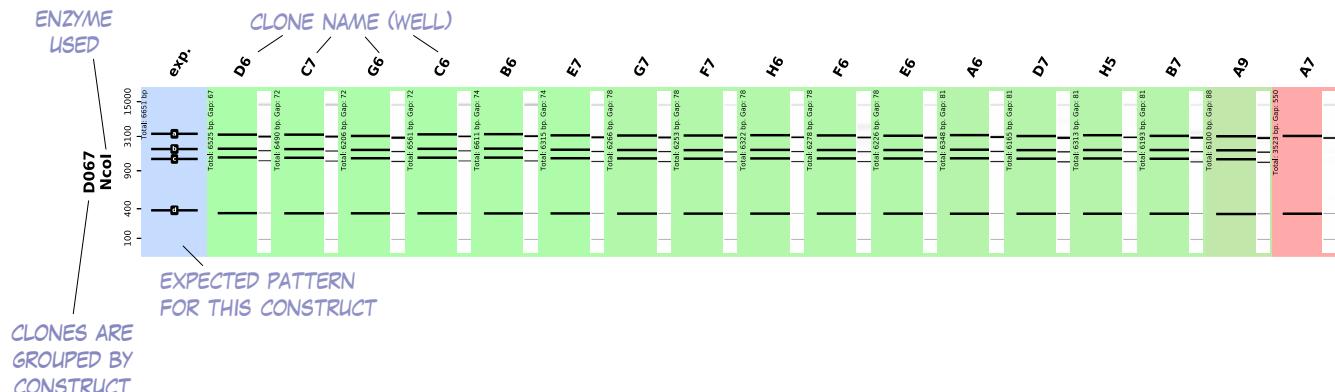


EGF restriction digest software

AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION

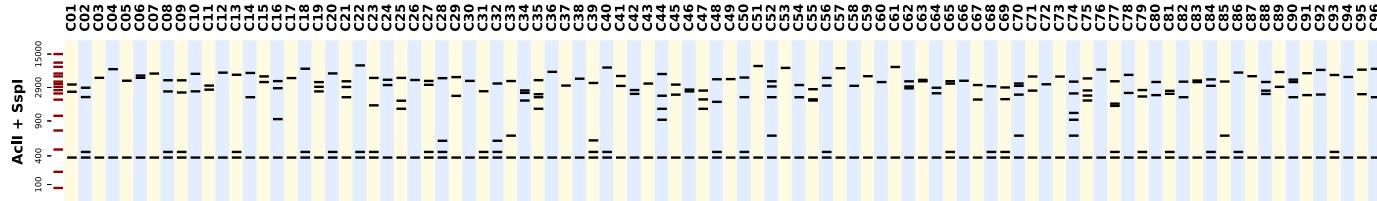


EXTENSIVE REPORT GENERATION FOR CONSTRUCT VALIDATION

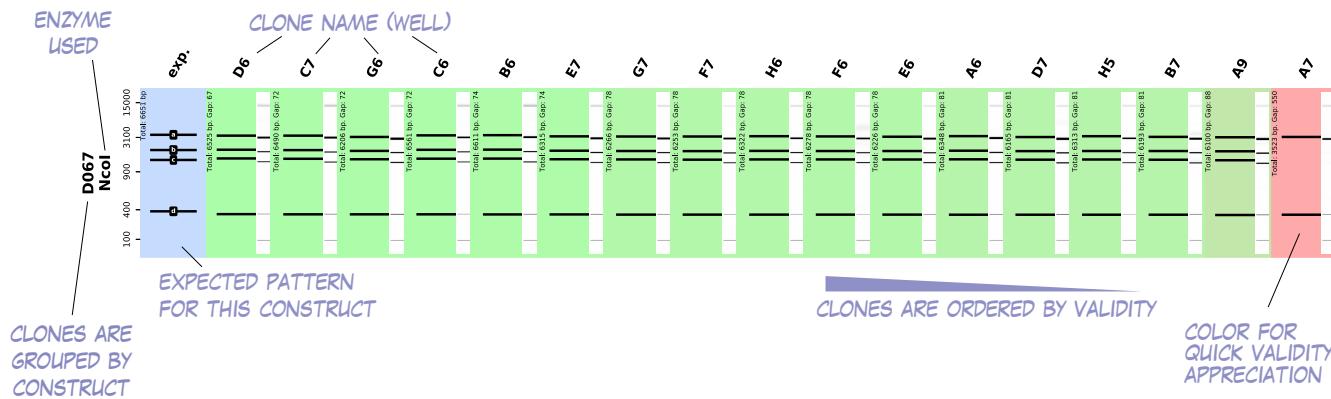


EGF restriction digest software

AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION

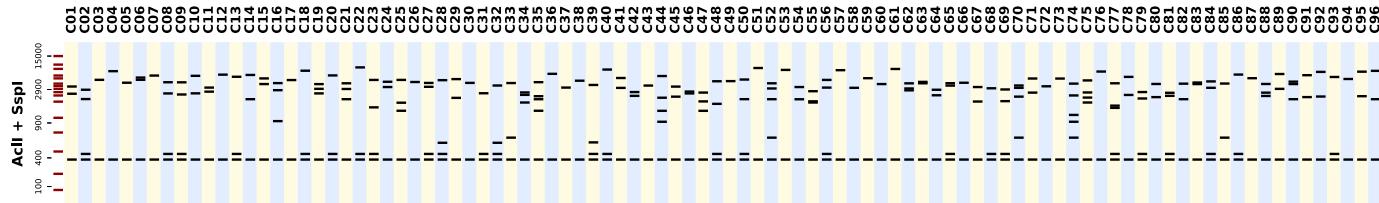


EXTENSIVE REPORT GENERATION FOR CONSTRUCT VALIDATION

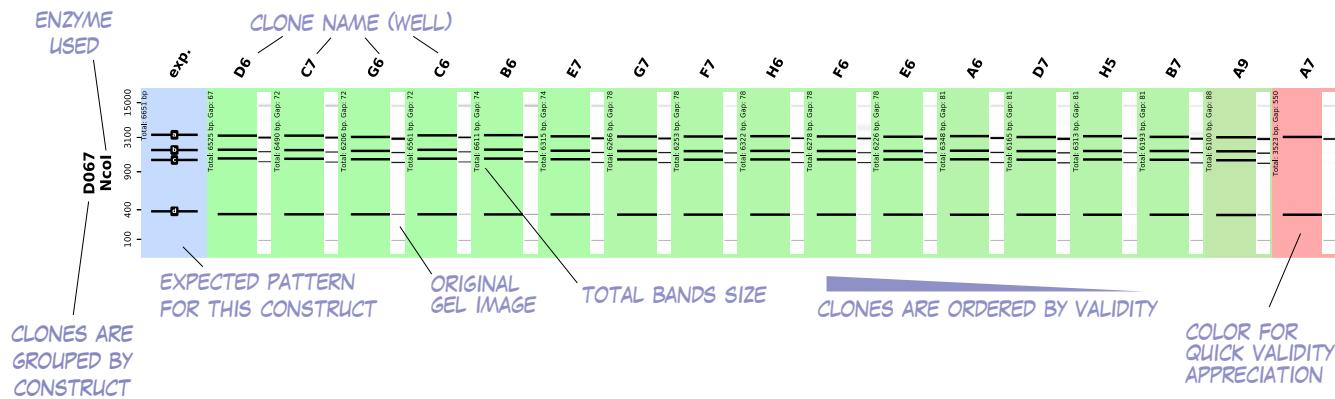


EGF restriction digest software

AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION



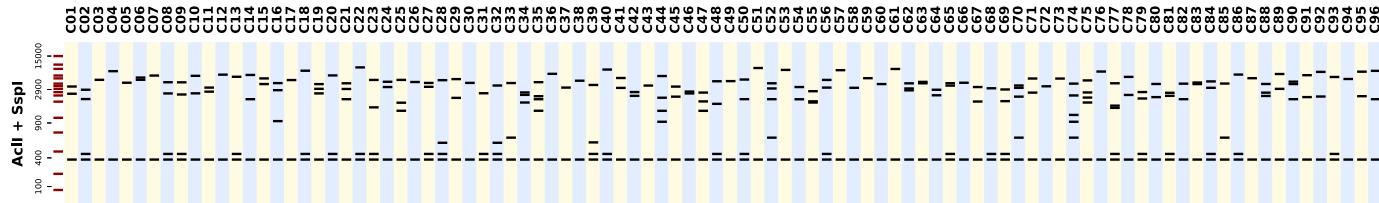
EXTENSIVE REPORT GENERATION FOR CONSTRUCT VALIDATION



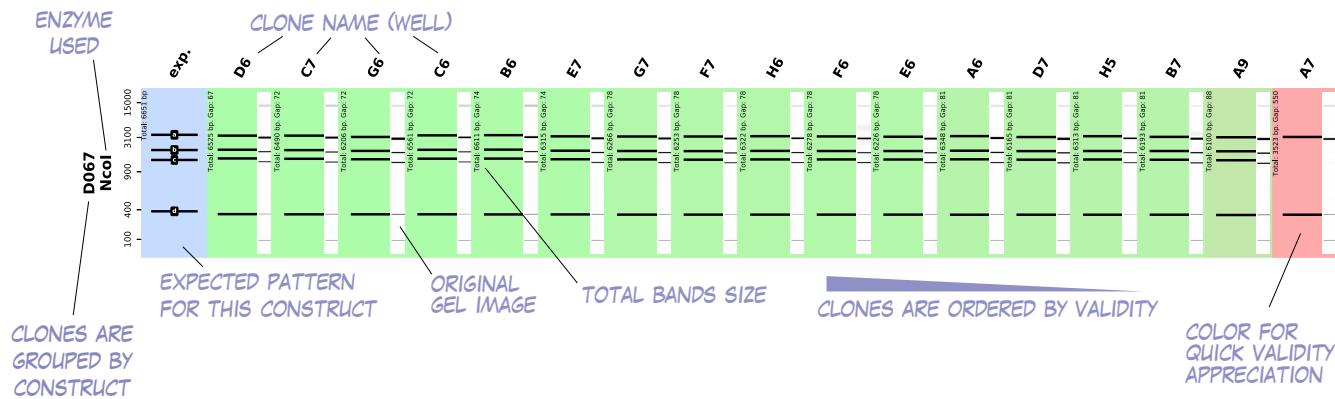
CLONES ARE GROUPED BY CONSTRUCT

EGF restriction digest software

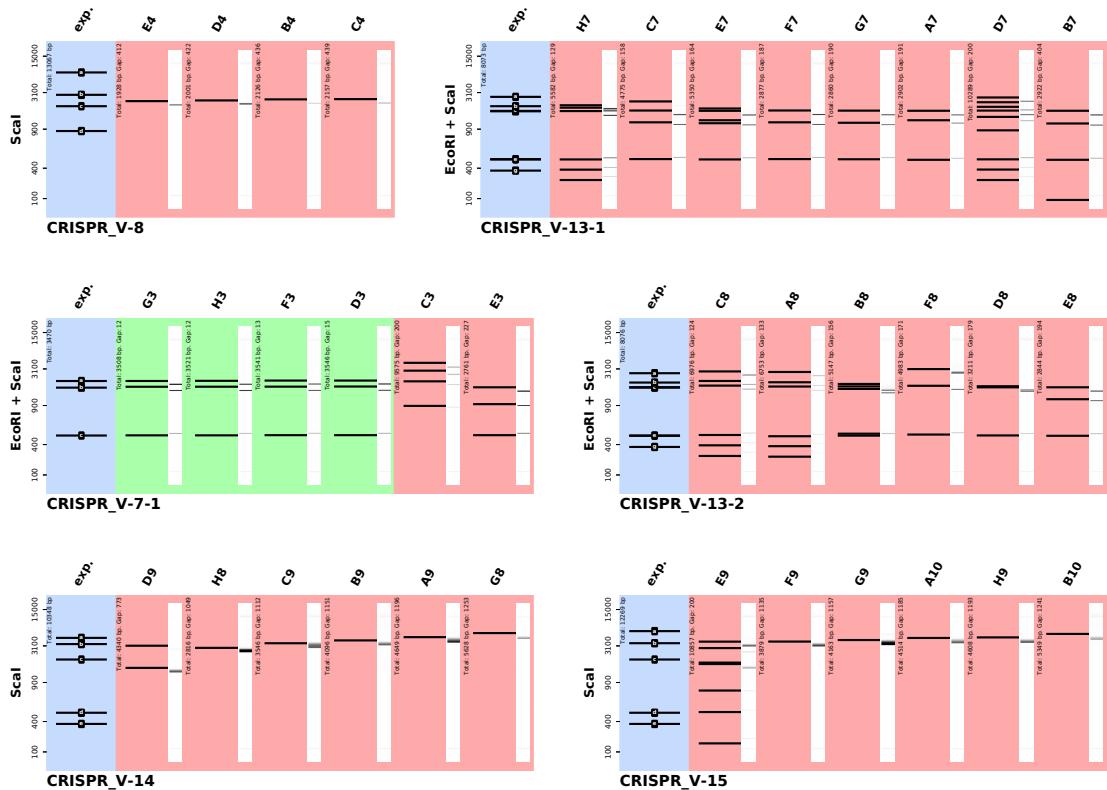
AUTOMATIC DIGESTION SELECTION FOR WHOLE-PLATE VALIDATION



EXTENSIVE REPORT GENERATION FOR CONSTRUCT VALIDATION



Aaaaaaargh ! Aaaaaaaaaaaaaaaargh !



What do you think happened ?

Find the saboteurs !

FAILED
ASSEMBLIES

SUCCESSFUL
ASSEMBLIES

PARTS

NAME	Value	max_Di	max_M	min_P	max_P	min_V	max_V								
ASSEMBLY-A	600	/	/												
ASSEMBLY-B	600	/	/												
ASSEMBLY-C	600	/													
ASSEMBLY-D	600	/													
ASSEMBLY-E	600	/													
ASSEMBLY-F	600	/													
ASSEMBLY-G	600	/													
ASSEMBLY-H	600	/													
ASSEMBLY-I	600	/													
ASSEMBLY-J	600	/													
ASSEMBLY-K	600	/													
ASSEMBLY-L	600	/													
ASSEMBLY-M	600	/													
ASSEMBLY-N	600	/													
ASSEMBLY-O	600	/													
ASSEMBLY-P	600	/													
ASSEMBLY-Q	600	/													
ASSEMBLY-R	600	/													
ASSEMBLY-S	600	/													
ASSEMBLY-T	600	/													
ASSEMBLY-U	600	/													
ASSEMBLY-V	600	/													
ASSEMBLY-W	600	/													
ASSEMBLY-X	600	/													
ASSEMBLY-Y	600	/													
ASSEMBLY-Z	600	/													

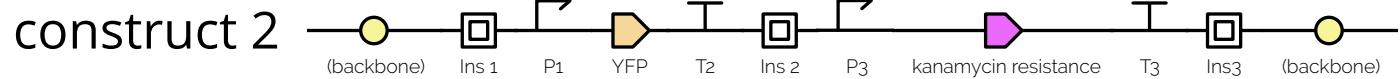
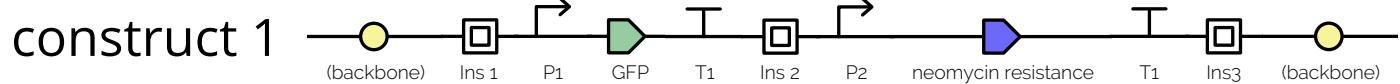
Find the saboteurs !

FAILED ASSEMBLIES

SUCCESSFUL ASSEMBLIES

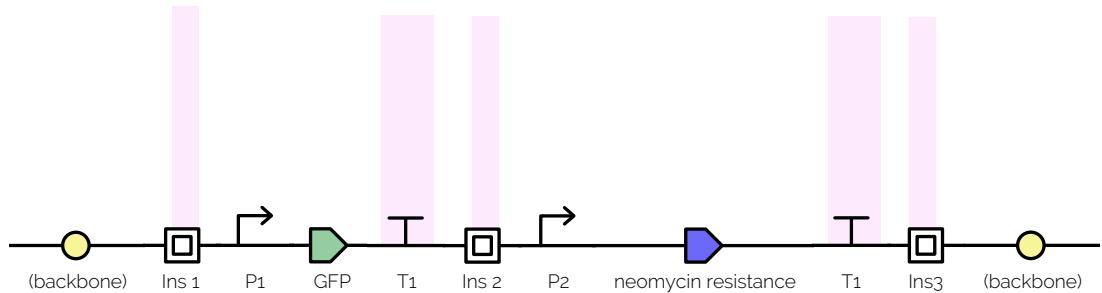
PARTS

Optimal Sanger Sequencing plan

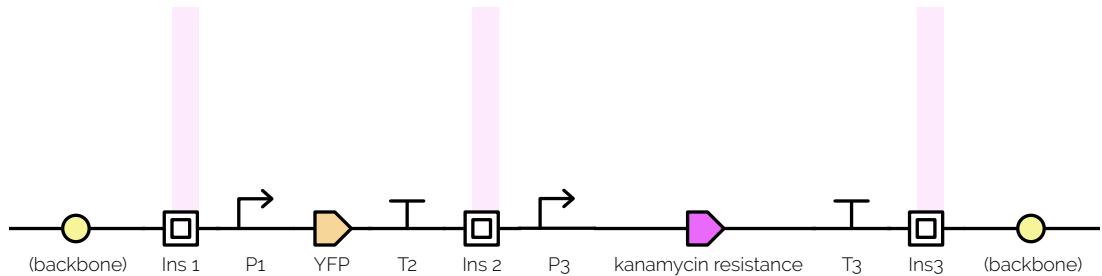


Optimal Sanger Sequencing plan

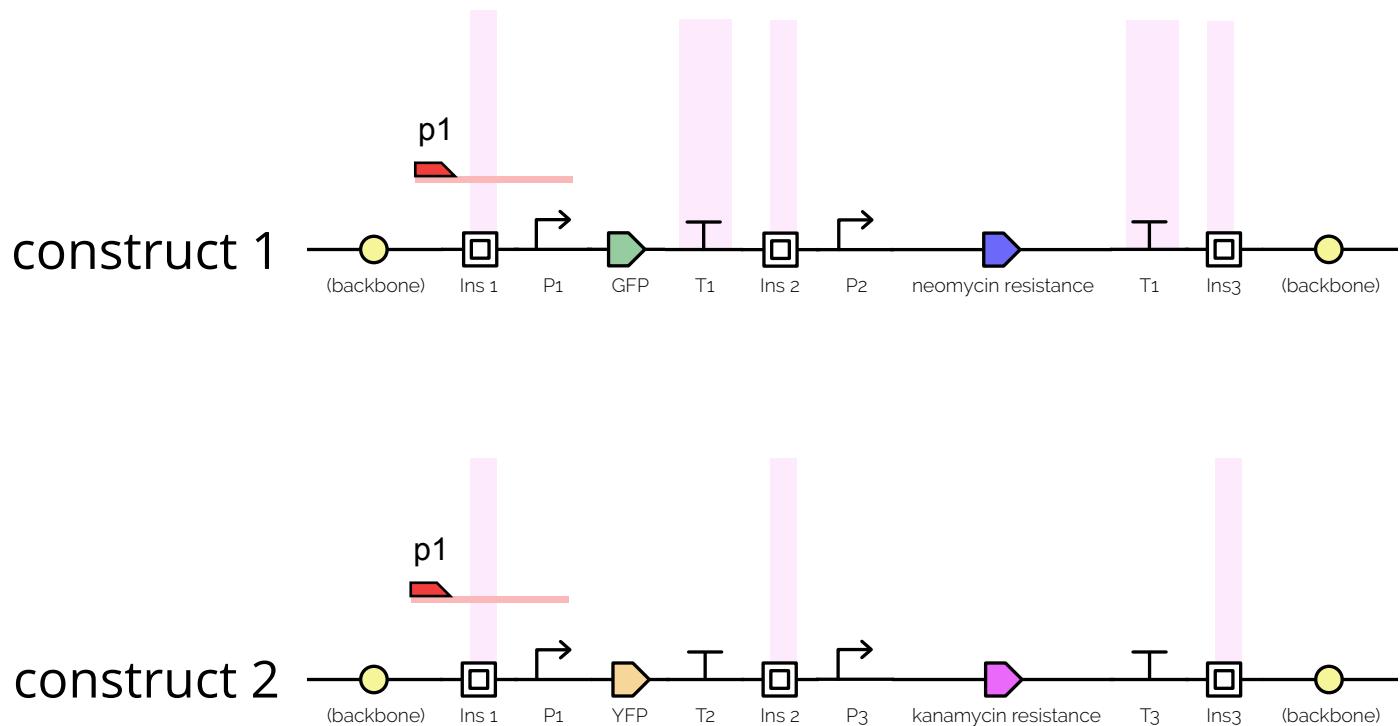
construct 1



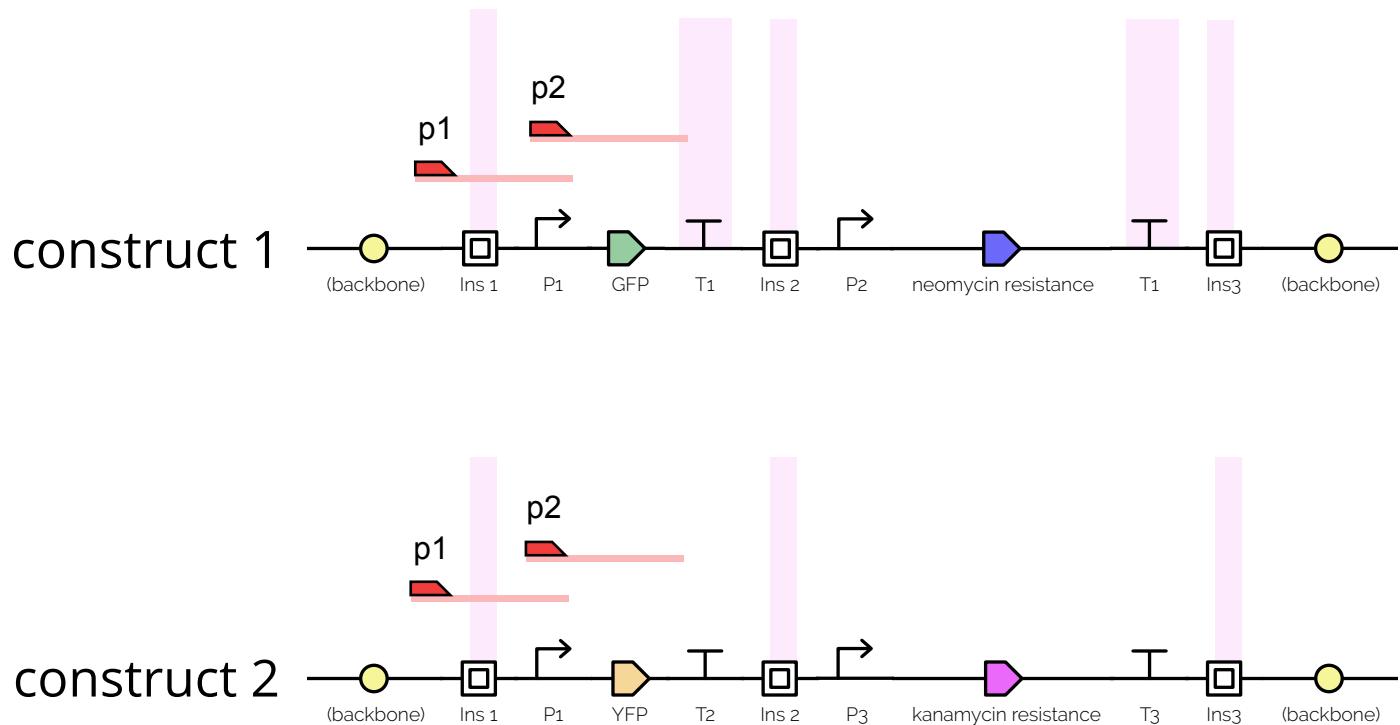
construct 2



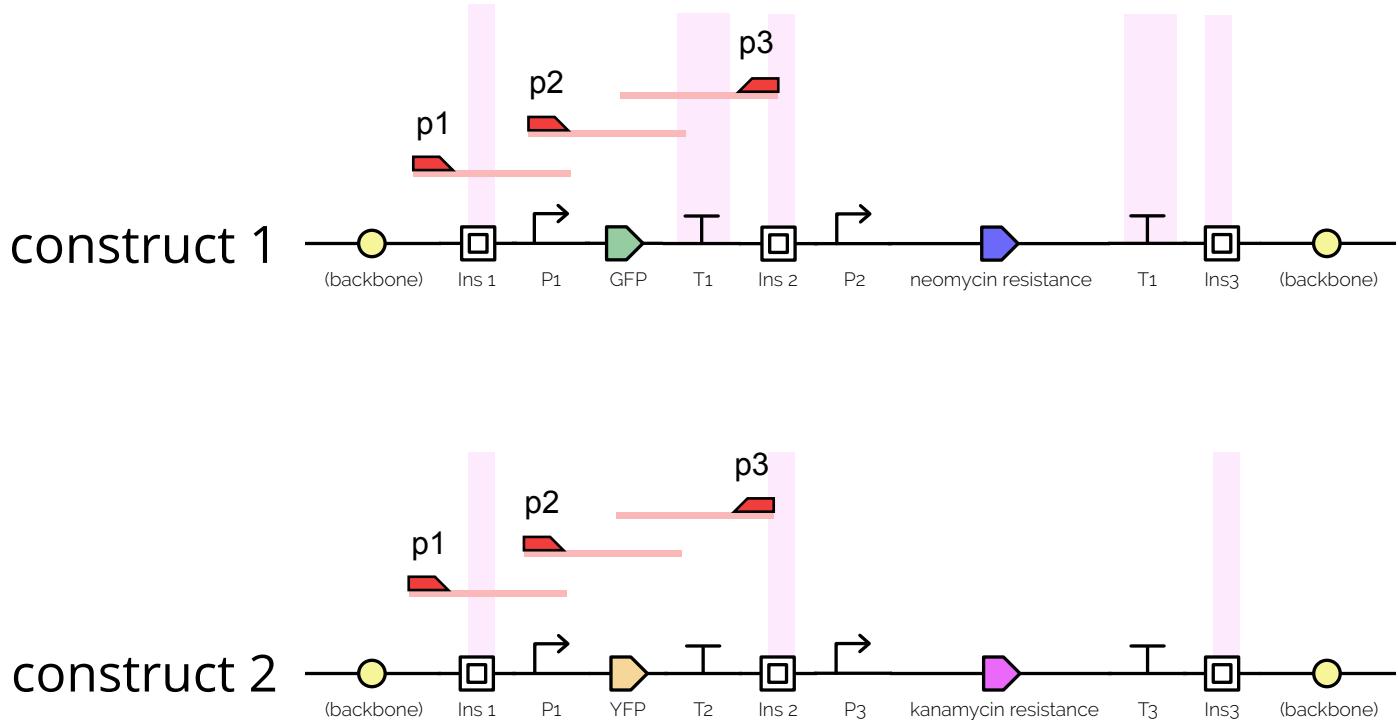
Optimal Sanger Sequencing plan



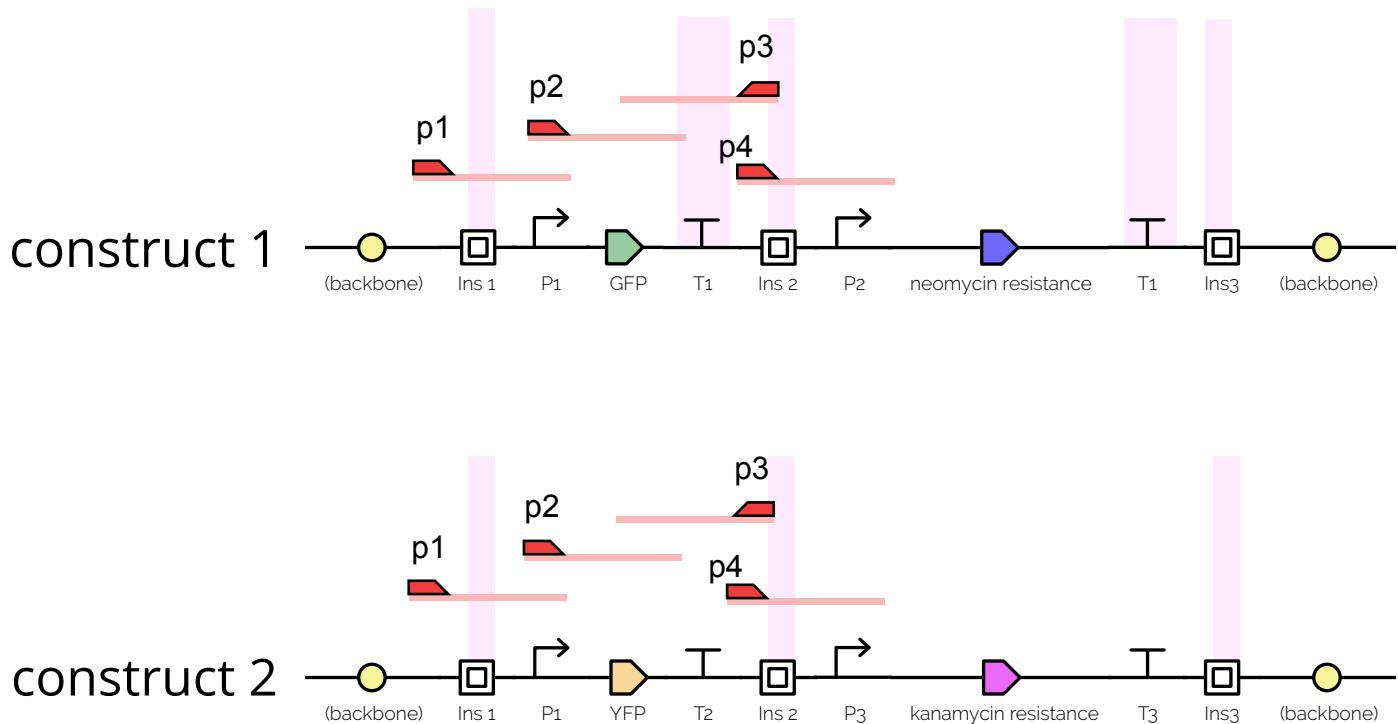
Optimal Sanger Sequencing plan



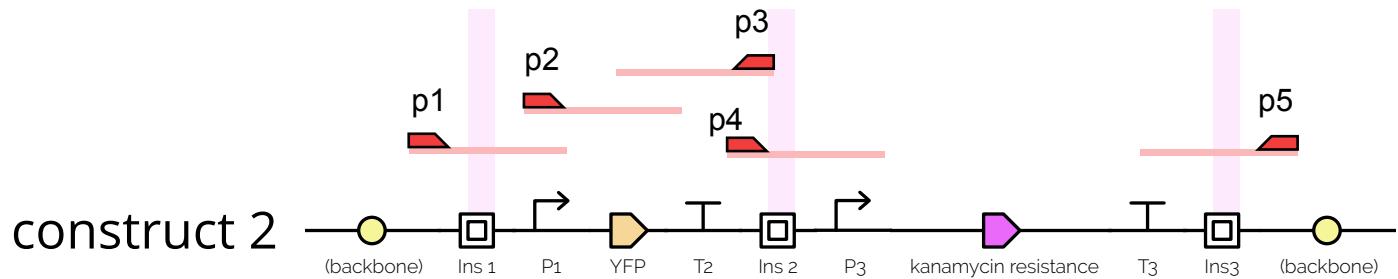
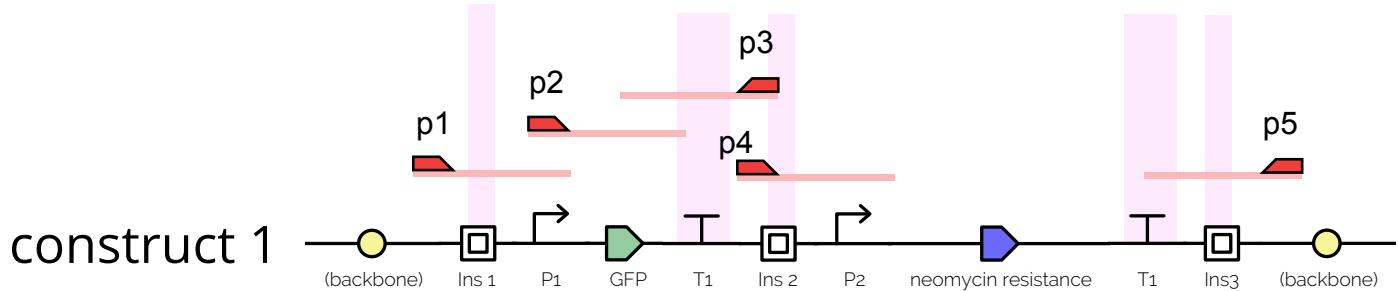
Optimal Sanger Sequencing plan



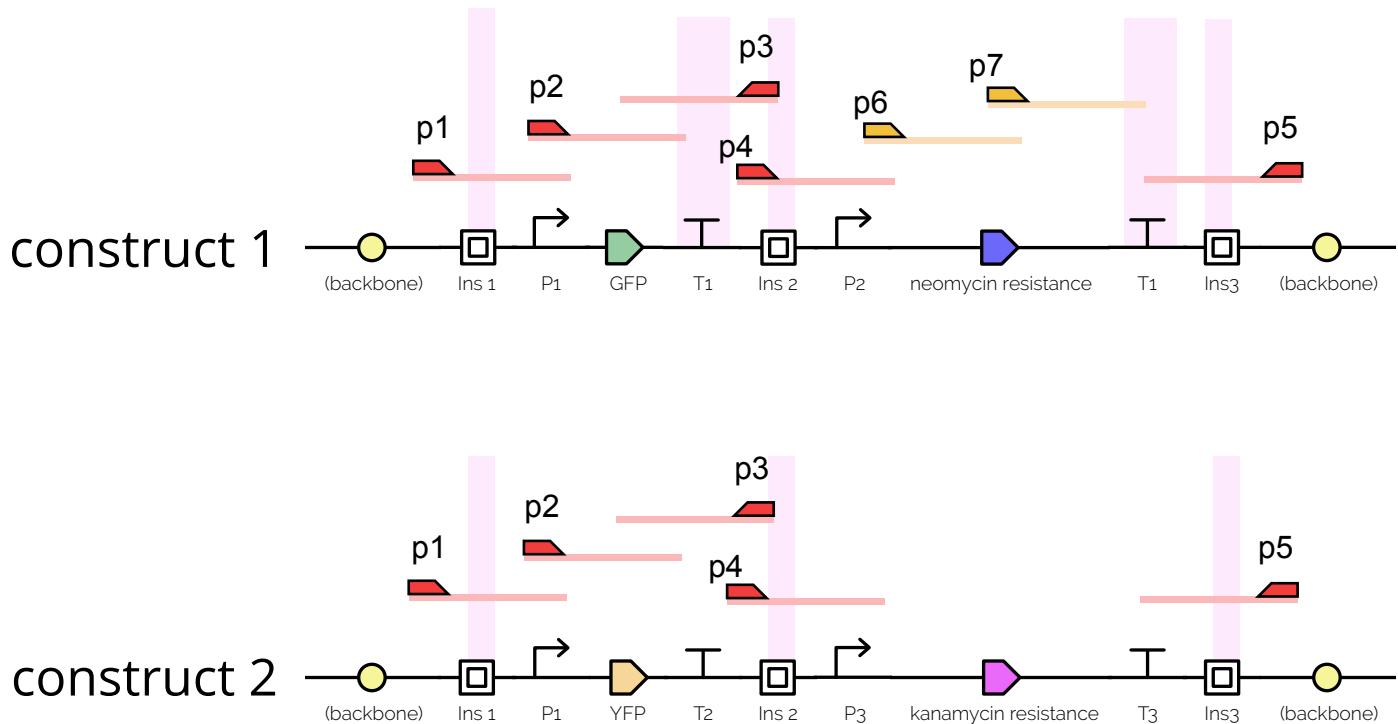
Optimal Sanger Sequencing plan



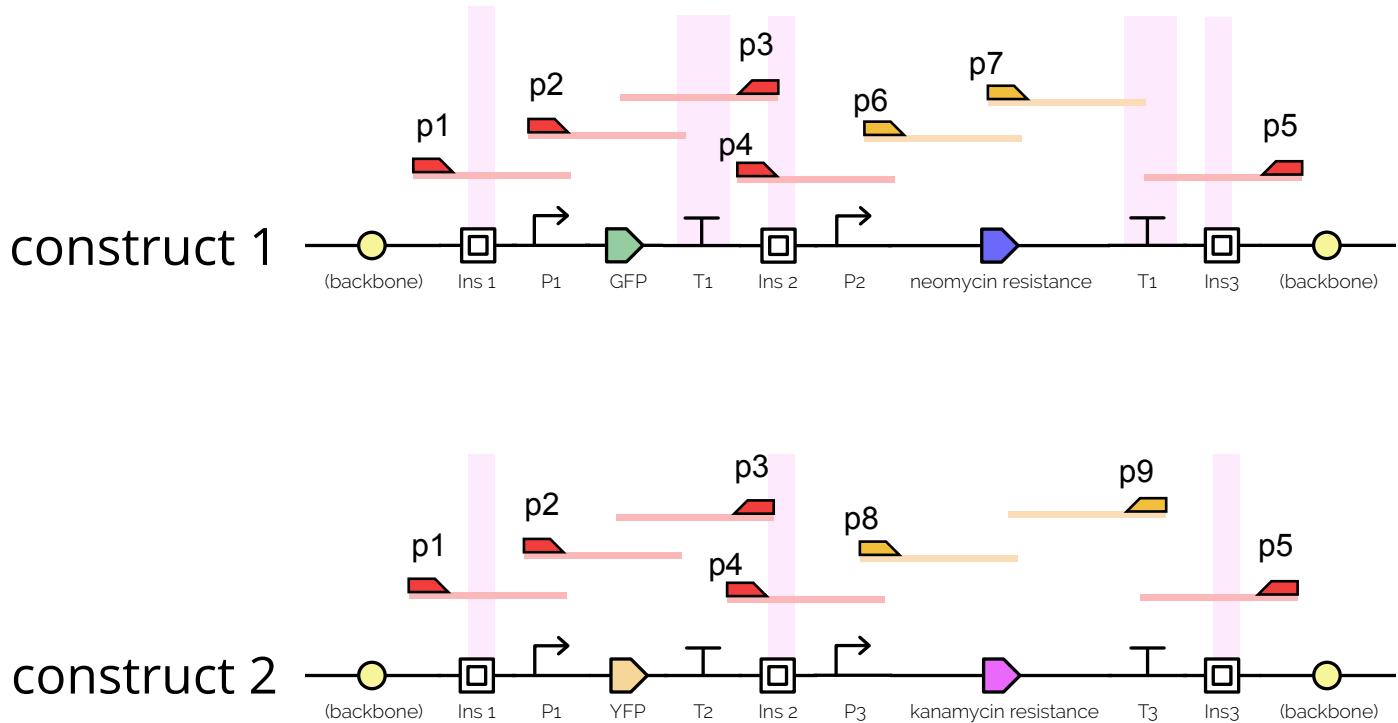
Optimal Sanger Sequencing plan



Optimal Sanger Sequencing plan



Optimal Sanger Sequencing plan



Primer selection software

constructs
sequences

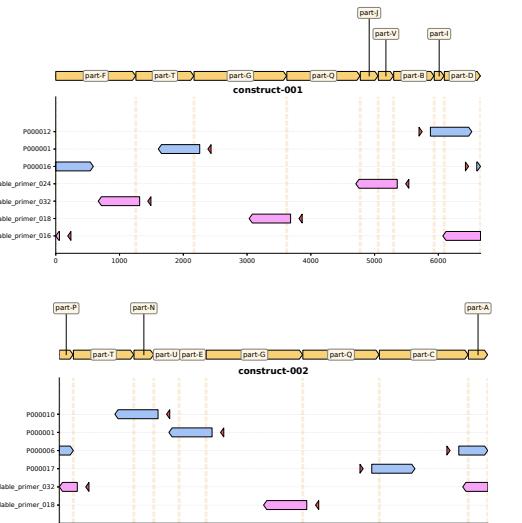


Screenshot of the "Select Primers" interface from the EGF CUBA genome browser. The page title is "Select Primers". It features a logo of a palm tree and the text "Find the best primers to your constructs, for validation or identification purposes." A dropdown menu shows "Validation Type: Sanger Sequencing". Below this are sections for "CONSTRUCTS SEQUENCES" (with a "Drop files here or click to select" box) and "AVAILABLE PRIMERS" (with a similar "Drop files here or click to select" box). A checkbox "Sequences are circular" is checked. At the bottom, a slider indicates the "ideal read range: from 150bp after primer annealing to 800bp".

available
primers



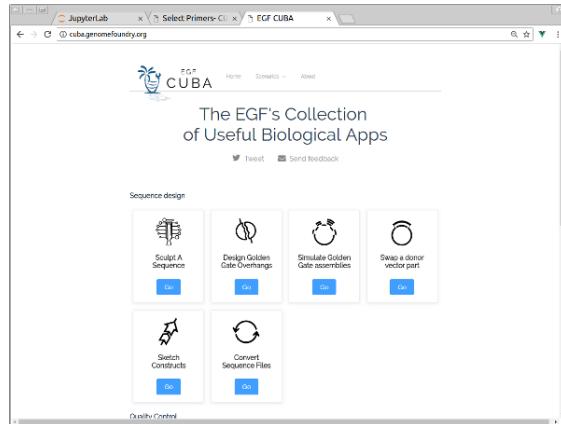
optimal sanger sequencing plan



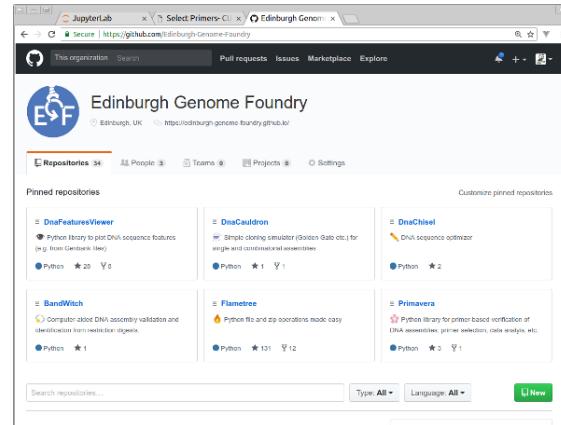
Conclusion

- ▶ The path to DNA assembly is dark and full of terrors.
One mistake can cost weeks and plenty of money.
- ▶ Advances in molecular biology and growing software investment should solve many issues in the next years.
- ▶ Next time you design or assemble DNA, have a look at the latest software... or call the Foundry !

Conclusion: links



Public web apps:
cuba.genomefoundry.org



Computational libraries on Github:
[Edinburgh-Genome-Foundry](https://github.com/Edinburgh-Genome-Foundry)

Contact:
valentin.zulkower@ed.ac.uk