Design tools for synthetic biology

Nathan J. Hillson
njhillson@lbl.gov

2016 JGI User Meeting – GenTech Workshop
March 22, 2016
**DNA Synthesis Science CSP Project**

**Synthesis call (~400 Kbp/project)**
- Biannual CSP: Feb and Jul
- Large CSP: Aug (w/ sequencing)
- : Apr (JGI-EMSL)
- BRCs: Any time

**Review (2-3 months)**
- User Agreement
- Approval by the DOE & JGI Director
- Internal Review ~3 Weeks
- Scientific Review ~3 weeks
- Technical Review ~3 weeks

**Production**
- Initiation call: Defining the scope of work
- Sequence Data Mining & Analyses
- Biosafety/Security Screening: BLiSS
- Construct Design: SPL DIVA
- Synthesis Assembly
- Transcript & Metabolite Analyses

**USERS**
4 Mbp/Yr in total

---

Simple Sequence Data Mining & Analyses
Synthetic Biology – Biosecurity Issues

DNA Synthesis

>DNA_construct_seq
ATGAAGAGAAAGAACAAAGTACT
ATCTATCCTGCTTACCCTGTGC
TTATAATTTTCGACAAACCTCAGTG
AACATGAGTTTTTGCTGTAAGCTAC
The HH&S Guidance

1) Customer Screening – establish the legitimacy of customers
2) Sequence Screening – identify when “sequences of concern” are ordered
3) Follow-up screening – verify the legitimacy of the end-use

EU Export Control List

- Commerce Control List (CCL)
  - Maintained by the US Export Administration

- Select Agents and Toxins
  - Maintained by the CDC
  - And USDA
Identifying “Sequences of Concern”

Technical recommendations for identification:

- Identify sequences that are 200 bp or longer
- Screen both DNA strands, and all six-frame translations
- Use sequence alignment methods against a public sequence database
- Use a “Best Match” approach to determine whether a query sequence is derived from or encodes a Select Agent or Toxin or, a sequence from a CCL-listed item
- Screening should identify sequences that are “unique” to Select Agents and Toxins and that “house-keeping genes” that maintain normal cellular physiology should be excluded
1) BLAST the query sequence (to be synthesized) using blastn (nr) and blastx (nt)

`ATGAAGAGAAGAAGCTGCTTACCC...........TGCAAAAGAGTTTTTTCGCTGAAGCTAC`

This yields a collection of alignments for each DNA strand, and all six frame translations (amino acids)
2) Discard all alignments that are under 200 nucleotides or 66 amino acids in length.

ATGAAGAGAAAGACCTGCTTACCC...........TGCAAAGAGTTTTTGCTGAAGCTAC
3) Screen remaining alignments against the blacklist database

ATGAAGAGAAAGAAACCTGCTTACCC............TGCAAAGAGTTTTTGCTGAAGCTAC

Matches to sequences of concern are identified
4) Calculate the span(s) of the query sequence corresponding to these “hits”

ATGAAGAGAAAGAACCTGCTTACCC........TGCAAAAGAGTTTTTGCTGAAGCTAC
5) For each 200 bp window along the hit span, align with all overlapping alignments

ATGAAGAGAAAGAACCTGCTTACCC...........TGCAAAGAGTTTTTGCTGAAGCTAC

For each 200 bp window along the hit span, BLAST the window query sequence against each alignment that overlaps the window

6) This gives us a % identity for each local alignment for the window, which we use to assign “Best Matches” for each window.

7) Each window is then assigned a “Status”
BLiSS Window Status Assignment

**Statuses:**
- **Failed**
  - Best match to an entity on the **Select Agents and Toxins List**
- **Controlled**
  - Best match to an entity on the **Commerce Control List**
- **Passed**
  - Not a best match to any sequence of concern

**Controlled Sequences**
- Consult with our Export Control Department

**Failed Sequences**
- Ask the user that submitted the sequence verify the legitimacy of the end-use
- Consult with our local FBI WMD coordinator

Failed sequences are not synthesized unless they have been cleared by the FBI
Statistics: 9192 Sequences Screened

Status

- 0.6% Select Agents and Toxins List (Failed)
- 2% Commerce Control List (Controlled)
- Passed 97%
DNA Synthesis Science CSP Project

**Synthesis call (~400 Kbp/project)**
- Biannual CSP: Feb and Jul
- Large CSP: Aug (w/ sequencing)
- Apr (JGI-EMSL)
- BRCs: Any time

**Review (2-3 months)**
- User Agreement
- Approval by the DOE & JGI Director
- Internal Review: ~3 Weeks
- Scientific Review: ~3 weeks
- Technical Review: ~3 weeks

**Production**
- Initiation call: Defining the scope of work
- Sequence Data Mining & Analyses
- Biosafety/Security Screening: BLiSS
- Construct Design: SPL, DIVA
- Synthesis Assembly
- Transcript & Metabolite Analyses

**USERS**
- 4 Mbp/Yr in total
FY 2015

User
Sequence Information

DOE JGI Synthesis Group

Design

Refactoring
Operon Structure

Back-Translate
Codon Optimize

Partition

Order sequences

Modify

Vendor

Verify
Violations?

YES

NO

Synthesis
What are common Synthesis Criteria?

- **Length (bp)**
- **GC% Content**
  - Global vs. Local (“Windowed”)
  - GC% Flux
- **Repeats and Repeat Coverage**
  - Global vs. Local (“Windowed”)
  - Tandem vs. Interspersed
  - Direct vs. Inverted
  - Exact vs. Mutated
- **Restriction Sites**

Example: Gen9

Improved Design for Synthesis and Assembly

User
Sequence Information

DOE JGI Synthesis Group

Design
- Refactoring
- Operon Structure
- Back-Translate
- Codon Optimize

Verify
Violations?
- YES Modify
- NO Partition

Order sequences

Vendor
Improved Design for Synthesis and Assembly

User
Sequence Information

DOE JGI Synthesis Group
Design
Refactoring
Operon Structure

Sequence Polishing Library (SPL)

Order sequences

Vendor

Verify Partitions?
YES
NO
Modify

Back-Translate
Codon Optimize
Enabling Customizable and Automated Workflows

- SPL Modules

**JUGGLER**
- Back-Translate
- Codon Optimize

**POLISHER**
- Verify
- Modify

**PARTITIONER**
- Partition
Work in Progress: SPL Web User-Interface

Sequence Information

Check against synthesis constraints (e.g. GC%, repeats)?

Check against pattern sequences (e.g. restriction sites)?

Suggest Modifications?

Codon Replacement Strategy:

Random

Codon Usage Table:

| Codon | UUU 22.2| UCU 15.9| UCC 13.8| UCG 13.0 | AAA 3.9| AAC 22.4| GCA 11.4| CCC 10.5| CGU 51.1| CUX 11.4| CGC 23.9| ACA 22.4| GCU 15.4| GCC 24.2| GUA 32.8| GUC 30.8| GUG 28.4| CAU 12.8| CAC 9.4| CCA 8.6| CCG 15.8| CUA 13.8| CUC 15.9| UUA 21.3| UUA 21.3| UUG 21.3| UAG 21.3| UAA 21.3| UAC 21.3 |
|       | (35946) | (25565) | (22136) | (20904) | (6257) | (36724) | (18366) | (16869) | (82300) | (16836) | (38504) | (36724) | (47324) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) | (25749) |

Enter Codon Usage Table  Upload Codon Usage Table

Save Code of Usage Table?
DNA Synthesis Science CSP Project

Synthesis call (~400 Kbp/project)
- Biannual CSP: Feb and Jul
- Large CSP: Aug (w/ sequencing)
  - : Apr (JGI-EMSL)
- BRCs: Any time

Review (2-3 months)
- User Agreement
- Approval by the DOE & JGI Director
- Internal Review ~3 Weeks
- Scientific Review ~3 weeks
- Technical Review ~3 weeks

Production
- Initiation call: Defining the scope of work
- Sequence Data Mining & Analyses
- Biosafety/Security Screening: BLiSS
- Construct Design: SPL DIVA
- Synthesis Assembly
- Transcript & Metabolite Analyses

USERS
- 4 Mbp/Yr in total

Simple Sequence Data Mining & Analyses

Approval by the DOE & JGI Director

Internal Review ~3 Weeks

Scientific Review ~3 weeks

Technical Review ~3 weeks

Synthesis call (~400 Kbp/project)
Acknowledgements

DOE JGI DNA Synthesis Science Program

Lisa Simirenko
BLiSS

Ernst Oberortner
SPL