Synthetic Biology II

March 1, 2017
RECAP: Building Synthetic Gene Networks and Systems

- Construction of small gene networks from well-characterized biological parts, guided by models

Bistable Toggle Switch

Repressilator
Elowitz & Leibler Nature 403 (2000)

Feedback Loops
Freeman Nature 313 (2000)
& EMBO J 20 (2001)
Isaacs et al PNAS 100 (2003)

RECAP: What Can We GET From Engineering Biological Systems?

Payoffs for the Nation

<table>
<thead>
<tr>
<th>Within a Decade</th>
<th>Long Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop knowledge base for cost-effective cleanup strategies</td>
<td>Save billions of dollars in toxic waste cleanup and disposal</td>
</tr>
<tr>
<td>Understand earth’s natural carbon cycle and design strategies for enhanced carbon capture</td>
<td>Help stabilize atmospheric carbon dioxide to counter global warming</td>
</tr>
<tr>
<td>Increase biological sources of fuels and electricity</td>
<td>Contribute to U.S. energy security • Biohydrogen-based industry in place</td>
</tr>
</tbody>
</table>

http://doegenomestolife.org/
Today’s Overview

1. Synthesis and assembly of DNA
2. Diversity and selection in artificial constructs
3. Scaffolds: Minimal genomes
4. In vitro synthetic biology
The chemistry for solid phase synthesis of nucleic acids is incredibly efficient (>99%).

Oligos are very cheap – cost is comparable to shipping of the product.

Commercial sources can provide oligos up to 200 bp.

http://www.trilinkbiotech.com/tech/oligo_history.asp
DNA synthesis costs continue to drop much faster than Moore’s Law.

Note *useful* costs (e.g. finishing a genome) are higher than *raw* costs.

NHGRI; http://www.genome.gov/sequencingcosts/
DNA Synthesis

Synthesis has gotten cheaper and extends to huge constructs.

Approaching lengths of eukaryotic chromosomes.
Gene Synthesis

(a) 56 40-mers

(b) gene assembly: 55 cycles PCR

(c) add outside primers; gene amplification: 23 cycles PCR

(d) cut with Sfi I; clone in pUC vector

5' Sfi I 0.9 kb Ap^R Sfi I 5'

pUC322-Sfi

Tc^R

Stemmer, Gene 164 (1995)
Advances in gene assembly

Low-synthesis, high-throughput oligo syntheses – microarrays, microfluidics

Lots of computation to optimize hybridization and reduce mis-hybridization

Microfluidics confine very small amounts of synthesized oligos to small volumes such that the concentrations are workable

www.thermofisher.com

Kong, et al., NAR 35 (2007)
(1) **Selection**: When an assembled construct is required for survival, ruling out errors is easy (e.g. Stemmer assembly of *bla* gene, 1995)

(2) **Repair/Optimize**: MutS systems to identify and degrade mismatches for subsequent resynthesis; site-directed mutagenesis to fix errors found by sequencing; matched annealing temperatures, high-fidelity polymerases, etc.

(3) **Purification**: Hybridization-based retention of perfect matches, mismatch specific endonucleases

(4) **Sequencing**

Carr and Church, Nature Biotech., 27 (2009)
Genome Synthesis

Target: *Mycoplasma genitalium*
582,927 bp genome

Starting point:
- 101 cassettes ~5-7 kb
- outsourced cassette synthesis
- 80 bp overlaps
- inter-genic boundaries
- confirmed by sequencing

Genome Synthesis

Hierarchical assembly:
- 4 cassettes combined to form ~24kb assemblies
- Joined by in vitro recombination
- Cloned into BACs

Hierarchical assembly:
- BACs were purified and cut with NotI
- Same process repeated yield B and C assemblies
- Final assemblies problematic in E. coli
- Recombination in yeast used

Genome Synthesis

Sequencing of the synthetic genome matched the designed sequence exactly, but sequencing at earlier stages did reveal some errors – lots of sequencing!

Cost? (est. $40M !!!)

The size of the construction exceeded what could be done in E. coli – yeast was required for stability, but also provided the recombinational machinery.

The same group has since demonstrated single-step assembly of the same genome from 25 overlapping parts.

Metabolic engineering involves only 10-50 kb – assembly is far ahead of design/engineering.

Ok, the genome is harbored in yeast – now what?

Need to swap back into a viable bacterial chassis

As long as your genome is in yeast, you can take advantage of genetic tools in yeast.

Demonstration of a seamless deletion of a gene in the bacterial genome while in yeast.

Restriction-modification system was an issue

Genomes were introduced into donor cell by PEG-mediated transformation – low efficiency (1 cell per 150,000 transplanted)

What are the parts?

While the number of reported synthetic regulatory circuits has increased, the complexity of those circuits has leveled off.

Workhorses of synthetic biology:
  repressor-operator pairs: lac, tetR, λ cl
  signaling: AHL/LuxI/LuxR

BioBricks (bbf.openwetware.org)
  DNA-encoded biological functions meeting technical and legal standards
  Promoters, RBSs, protein domains, terminator, plasmid backbones
  Standard use of restriction sites facilitates fusing BioBricks
  You can buy a BioBricks assembly kit from NEB

You may have to make your own!
DNA Shuffling

Partially digest a gene (or homologous genes) with Dnase I
Gel purify reasonably-sized fragments
Treat these as overlapping oligos
Recombination for PCR – “sexual PCR”, “molecular breeding”

DNA family shuffling: subtilisin

Shuffled 26 homologous genes
Selected for four enhanced properties
Best parent not always the closest parent
Found improved progeny (for a well-optimized protein)
Multiple solutions for each property

Genome Shuffling

Progeny bacteria have more than 2 parents
Phenotype improvement greater than sequential random mutagenesis
Two cycles of genome shuffling provided improvements comparable to those requiring 20 cycles of sequential mutation
Shuffling accomplished by protoplast fusion

Multiplex automated genome engineering (MAGE)

Degenerate oligos are targeted to the lagging strand in DNA replication. Repeatedly introduced into a cell - >30% of the cell population modified every 2-2.5 hours.

Multiplex automated genome engineering (MAGE)

This process has been automated and applied to optimize the production of lycopene.

20 genes were simultaneously targeted with degenerate rbs sequence-containing oligos, and 4 genes from secondary pathways were targeted for inactivation.

Multiplex automated genome engineering (MAGE)

Six variants were sequenced after 35 MAGE cycles

5-fold increase in lycopene production

Mutations were consistently chosen

Key idea: Speed up directed evolution by tying each round to the life cycle of phage, not bacteria

You can couple GeneIII expression to a variety of desired functions:

- Protein-DNA
- Protein-protein
- Bond-forming catalysis
Phage Assisted Continuous Evolution (PACE)

Test Case: Change T7 RNAP to recognize T3 operator sequence
Scaffolds for synthetic biology

- What are desired properties for host organisms?
  - Well-studied (E. coli, yeast)
  - Easily manipulated genetically
  - Performs a function similar to an engineering goal
  - Provides/generates an important precursor
  - Small
Minimal genome: E. coli

Why try to reduce the genome of a bacterium?
- Less redundancy
- Remove functions unnecessary for engineering applications
- Fewer genes means fewer interactions, more isolated modules

Almost 100 segments in the K-12 E. coli genome missing from 5 other E. coli were deleted (contained ~900 genes)

Strain MDS43 lost 743 genes, with 708kb deleted (final size 3.9 Mb)

Posfai, et. al., Science 312 (2006)
Essential genes: M. genitalium

M. genitalium has the smallest genome of a culturable bacterium

Transposon mutagenesis identified 100 out of 482 genes as nonessential

Hypothetical genes of unknown function account for 28% of essential genes

Glass, et. al., PNAS 103 (2006)
1. Tools exist to engineer DNA construct from oligos to genomes
2. Decrease in cost for DNA synthesis and sequencing are driving progress
3. Genome manipulation is out ahead of design
4. True design is rare and hard – most success stories involve tweaking and rearranging pre-existing systems
5. Biological systems are uniquely capable of managing diversity and selection
In vitro transcription and translation systems allow for cell-free selections.

Emulsions sequester protein activity, linking genotype and phenotype.

Display techniques

Diversity generation:
- Error-prone PCR
- Degenerate cassettes
- DNA shuffling
- Recombination
- Somatic hypermutation

Phage display

Genotype-phenotype connection – protein is physically bound to its DNA or RNA: Phage, yeast cell surface, mRNA, ribosomes, B cells

DNA-directed combinatorial chemistry

To apply evolution to chemistry, you need to couple genotype to phenotype.

Synthesizing a molecule on a DNA molecule is one way to accomplish this.

Extends evolution and selection to unnatural molecules and polymers.

Synthesis is directed by DNA – identity of product can be deduced from DNA sequence, so less material needed for analysis.

DNA-directed combinatorial chemistry

"Translation" site

Capture beads to split library by hybridization

DNA-directed combinatorial chemistry

Combinatorial Drug Biosynthesis

Many important natural products are synthesized by modular assemblies - “assembly line enzymology”:
Non-ribosomal polypeptides
Polyketides

The chemical structure of the natural product can often be directly inferred from the variable domains in each modular assembly – combinatorial rearrangements of the modules would yield new natural product scaffolds

Menzella and Reeves, Curr. Opin. Micro., 10 (2007)
Combinatorial Drug Biosynthesis

Modules have been manipulated genetically, and much work is being done to understand the ‘hand-off’ of intermediates between modules.

A long-term goal is to combinatorialize (diversify) these modules and select for useful products.

Menzella and Reeves, Curr. Opin. Micro., 10 (2007)
Glycorandomization

Many natural products must be glycosylated to be fully active, or in some cases to be active at all.

This process has well-defined steps.

Diversifying the sugars that decorate a scaffold molecule is an attractive route towards modifying or improving the activity of the molecule.

This effort combines synthetic chemistry and synthetic biology.

Evolution of Promiscuous Glycosyltransferase

Five hard truths

1. Many of the parts are undefined
2. The circuitry is unpredictable
3. The complexity is unwieldy
4. Many parts are incompatible
5. Variability crashes the system