Reading & Writing Genomes Goals

• 2nd Generation BI/O: Reading & Writing
  Engineer cancer- & virus-resistant genomes

• Personal Genomes – Integration tasks
  - Personal Genomes: Environments, Traits,
  - Stem cells, Microbiome/Immunome
  - Synthesis for Causality (CEGS)
# How & why genome engineering for chemicals?

<table>
<thead>
<tr>
<th>Engineering</th>
<th># genes</th>
<th>Scale</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic</td>
<td>1-2</td>
<td>Plasmid</td>
<td>Protein drugs</td>
</tr>
<tr>
<td>Genome</td>
<td>1-2</td>
<td>Chromosome</td>
<td>Mutant models</td>
</tr>
<tr>
<td>Metabolic</td>
<td>2-30</td>
<td>Pathway</td>
<td>Chemical production</td>
</tr>
<tr>
<td>Code</td>
<td>300-all</td>
<td>Genome</td>
<td>Multi-virus resistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Safety, new AA</td>
</tr>
</tbody>
</table>
Engineering Life

Interoperable parts
Hierarchical designs

CAD
Cost effectiveness
Standards

Isolation
Testing

Redundant systems
Surveillance
Licensing

Evolution
How & why genome engineering for chemicals?

$/$kg

- Imiglucerase
- M13 DNA
- Salmon DNA
- White silk
- Casein
- Raw cocoons
- Algae biomass
- Cellulases
- Oil
- Wood
- Cement
**Clostridium phytofermentas Cellulase**

**Tolonen et al.**

Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367mmi_6890 1..16
Bio-petroleum from microbes flotiation -- not distillation

Fatty acid derived

Gasoline & diesel for current engines & infrastructure

2010 Presidential Green Chemistry Challenge Award
Comparative + Structural Genomics Success

Microbial Biosynthesis of Alkanes.
Schirmer, et al. Science 2010
Joule: fuel from phototrophs

• Achievements:
  – 40% of theoretical maximum productivity of organism

• Ethanol @ 10,000 gal/acre/year (target 25,000)
  – Discovery of unique genes and pathways for hydrocarbon diesel production
  – First-ever secretion of diesel from a phototroph
  – Ability to partition majority of carbon to product via carbon switch
Reading & Writing Genomes:
First semi-synthetic plasmid 1978: $10/b
CGI Human diploid genome 2009: $1500 / 6Gb
(=7 logs/30y mostly since 2005)

• Human insulin
• Human growth hormone
• Alpha-interferon
• G-CSF
• TPA
• GM-CSF
• Gamma-interferon
• IL-2
• Erythropoietin
• Hepatitis B vaccine

NAR 1978
Sutcliffe & Church
(BR: Bolivar & Rodriguez)

(Amgen, Biogen, Genentech, etc)
40,000-fold in 4 years

(Moore’s law) 1.5x/yr for electronics vs 10x/yr for DNA Sequencing

>20 years ahead of the 1970-2004 exponential

$/genome (30X)

2000: $3 billion

2004-7: $70M (ABI)

2008 $50K (GA)

2010 $1500 (CGI)
Moore’s law = 1.5x/yr
vs 10x/yr

1st-generation
Gene synthesis
vs
2nd-generation
Sequencing
& DNA synthesis

24Mb 30kb/$
<table>
<thead>
<tr>
<th>2nd-generation sequencing technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Illumina-GA</td>
</tr>
<tr>
<td>2. AB-SOLiD</td>
</tr>
<tr>
<td>3. CGI</td>
</tr>
<tr>
<td>4. Polonator</td>
</tr>
<tr>
<td>5. Roche-454</td>
</tr>
<tr>
<td>6. Helicos</td>
</tr>
<tr>
<td>7. Ion Torrent</td>
</tr>
<tr>
<td>8. Pacific Bio</td>
</tr>
<tr>
<td>9. Intelligent Bio</td>
</tr>
<tr>
<td>10. GnuBio</td>
</tr>
<tr>
<td>11. Halcyon</td>
</tr>
<tr>
<td>13. LightSpeed</td>
</tr>
<tr>
<td>14. Bionanomatrix</td>
</tr>
<tr>
<td>15. OxfordNanopore</td>
</tr>
<tr>
<td>16. Visigen</td>
</tr>
<tr>
<td>17. ZS Genetics</td>
</tr>
<tr>
<td>18. Nabsys</td>
</tr>
<tr>
<td>19. GE Global</td>
</tr>
<tr>
<td>20. IBM</td>
</tr>
<tr>
<td>21. Electronic Biosci</td>
</tr>
</tbody>
</table>
2nd-Gen Gene Synthesis: chips

$500 per 15Mbp

8K Xeotron Photo-Generated Acid
12K Combimatrix Electrolytic
120K Roche, Febit Photolabile 5'protection
244K Agilent Ink-jet standard reagents

4 chip technologies: Amplify oligo pools with flanking universal primers

4 Paths to error correction
1. Hyb-Select: Tian et al. 2004 Nature
4 Paths to error correction

#4 : Bang
Nat Meth. 2008
Coalescence of 2\textsuperscript{nd} Generation of DNA reading & writing

**Write**
- Oligo elute & amplify
- Febit or LC

**Read**
- GS FLX Roche/454
- Polonator flowcell
- sort & select

**Select**
- pick & place
- multiwell plate
- Rolony Photorelease
- PCR

**Data**
- Agilent OLS
- Array synthesis

Oligonucleotide-Design
From open-access Sequencer to Bio-Fab

1. Select ‘perfect part’ sequences
2. Device characterization
3. Cell sorting (FACS)

- Digital micromirrors
- Flow-cell
- Billion beads or cells/run
- Photo-labile immobilization

Rich Terry
2nd Gen Assembly from 2.6 Mb raw oligos

Kosuri Eroshenko
2nd Gen Assembly from 2.6 Mb raw oligos

Kosuri Eroshenko
4 Protein/RNA-directed recombination strategies

1. Integrase/recombinase
   \(\lambda\) (Gateway)
   \(\phi C31, \text{Cre-lox}\)

2. Group II introns

3. Meganucleases
   Scel, Drel

4. Zn Finger Nucleases

4 DNA homology-directed strategies

#1: ds-Circle x Circle
2 step recA+ recombination
Select + counterselect
Link et al J. Bact 1997
(Open-access)

#2: ds-Linear x Circle
1 step 5’>3’ exo Reda/E b/T
Select
(GeneBridges license)

#3: ss-90mer x ds-Circle
Costantino & Court PNAS’03
Wang et al., Nature '09

#4: ss-Mb x ds-Circle conjugation
Multiplex Automated Genome Engineering (MAGE)

Allelic Replacement
- Strain: MG1655, ΔmutS, integrated λ-Red
- Highly complex oligo pools for multiplexed multi-loci modifications
- >4 billion bp of targeted genetic variation produced per day

Optimized Parameters
- Oligo length: 90mer
- Oligo 2ndary structure: <12 kcal/mol
- Oligo half-life: 5’ phosphothiol bps
- Oligo conc.: up to 50 uM
- Cycle time: 2 to 2.5 hrs
- up from 1E-4 to 25% efficiency per cycle

Costantino & Court PNAS 2003 ; Wang HH, et al Nature 2009
MAGE System Upgrade Jan-May 2010

Dimension: 4’ x 3’ x 2.5’

96-well Electroporator
Sample Storage
Motion 1
DNA Queue
Motion 2
Incubation
Waste
Expression Reagent
Electroporator Head Unit
Microfiltration Plate
Reagents: H2O, EtOH, Growth Media

Harris Wang
Accelerated Evolution 100K combinations

Lycopene (hydrocarbon): 20 genes up, 4 down, 2 new

Max yields in 3 days & improve growth rates

Harris Wang et al Nature 2009
Existing Sensors (select for new ligands)


12 Riboswitches: Adenine B12 FMN Guanine Glucosamine-6-phosphate Glycine di-GMP Lysine Molybdenum PreQ1 SAM SAH TPP theophylline 3-methylxanthine

http://pubs.acs.org/doi/abs/10.1021/ja048634j
In vivo coupled biosensors

tRNA-ribosome

Repressor ds-DNA  mRNA binding
In vivo coupled biosensors

Tetracycline repressor ds-DNA
Divide & Conquer Genome Engineering
Multi-virus resistance: stop codons: TAG / total

<table>
<thead>
<tr>
<th>Virus</th>
<th>Length (b)</th>
<th>Type</th>
<th>TAG / Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>φX-174</td>
<td>5,386</td>
<td>ss-DNA</td>
<td>0 / 9</td>
</tr>
<tr>
<td>M13</td>
<td>6,407</td>
<td>ss-DNA</td>
<td>1 / 10</td>
</tr>
<tr>
<td>MS2</td>
<td>3,569</td>
<td>ss-RNA</td>
<td>2 / 4</td>
</tr>
<tr>
<td>T7</td>
<td>39,937</td>
<td>ds-DNA</td>
<td>6 / 60</td>
</tr>
<tr>
<td>T4</td>
<td>168,903</td>
<td>ds-DNA</td>
<td>19 / 277</td>
</tr>
<tr>
<td>E.coli</td>
<td>4,639,675</td>
<td>ds-DNA</td>
<td>314 / 1,360,152</td>
</tr>
</tbody>
</table>

ncbi.nlm.nih.gov/nuccore/9626372 56718463 176120924 9627425 29366675 (7 tRNAs: RITSPGL)

Farren Isaacs et al
<table>
<thead>
<tr>
<th>2nd base</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phe 22705</td>
<td>Ser 11802</td>
<td>Tyr 16795</td>
<td>Cys 8797</td>
</tr>
<tr>
<td></td>
<td>Leu 18894</td>
<td>Ser 11512</td>
<td>GUA 22037</td>
<td>GCA 7016</td>
</tr>
<tr>
<td></td>
<td>Ser 9620</td>
<td>RF1 2765</td>
<td>RF2 1249</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VG A 12210</td>
<td>RF2 1249</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RF2 1249</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RF2 1249</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RF2 1249</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RF2 1249</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Safety features:** no functional DNA exchange
**multi-virus resistance**

*New translation code:* novel AA

**Isaacs Charalel Church Sun Wang Carr Jacobson Kong Sterling**
Next: Freeing 8/33 tRNAs & 1/3 RF (9/64 codons)

ATA(I), GTC(V), TCC(S), CCC(P), ACC(T), GCC(A), CGG(R), AGR(R), TAG(-)

Min # bp changed = 5733 + 21050 + 11802 + 7490 + 32265 + 35382 + 7401 + 2771 + 1496 + 314
= 125,704 bp (2.7 % of the genome)

Genomes Environments Traits

One in a life-time genome + yearly (to daily) tests
Bio-weather map: Allergens, Microbes, Viruses
PersonalGenomes.org

PERSONAL GENOME 3M alleles

Therapies

Immunome

Personal stem-cells
Epigenome (RNA, mC)

TRAITS (Phenome)

Microbiome

Chemicals
4 Alleles from 4 Genome Sequences

• Primary ciliary dyskinesia (lungs)

• Pyrimidine synthesis (face & limbs)

Logan & Heather Madsen


Exome Sequencing Identifies the Cause of a Mendelian Disorder. Ng et al. Nature Gen. 2010
1) First/only open access data
2) Avoid over-promising on de-identification
3) 100% on Exam to assure informed consent
   (*Educate pre-consent rather than post-discovery*)
4) Genome sequence + epigenome
5) Multi-traits: images, iPS-etc.RNA, microbe/VDJ
6) Cells available for personal functional genomics
7) IRB approval for 100,000 diverse volunteers
   501(c)(3)

   16,000 so far
Resting State Laterality

Date/Time: 2010-03-31 06:25:21

Bold runs: 14 15

GlobalLtLI: 0.408467
"If one hospital is doing it, you can be sure others will start, because patients will vote with their feet,"
-- Elizabeth Worthey
HMGC & Children's Hospital of Wisconsin

“All the age of 3, he had more than 100 separate surgeries …On the basis of [sequencing], the physician recommended a bone-marrow transplant in June 2010. By mid-July, the child was eating his first meal.”
Medical Genomics: Individually rare collectively common 10%

1540 genes are **highly predictive & medically actionable** (inherited & cancer) at ~$2K per gene.

**Very few of these are on DTC SNP chips.** Why?
PKU, Tay Sachs, Cystic Fibrosis, BRCA1/2, etc.

Pharmacogenomic drug/allele combinations: Herceptin, Iressa, ..

Also:
Ancestry, Forensics,
Social Networking,
Education, Research
Figure 5.4: **Summary of MYL2 A13T findings from clinical testing.** This variant has been seen in two families and three sporadic cases based on our enquiries to clinical testing laboratories that sequence the MYL2 gene. In one family MYL2 A13T has been seen concurrently with MYBPC3 E619K and in the other family with MYH7 N1327K.
Genomes Environments Traits

One in a life-time genome + yearly (to daily) tests
Bio-weather map: Allergens, Microbes, Viruses

PersonalGenomes.org

PERSONAL GENOME
3M alleles

PERSONAL GENOME
3M alleles

Therapies

Immunome

Personal stem-cells
Epigenome (RNA,mC)

TRAITS
(Phenome)

Microbiome

Chemicals
PGP Vaccination Immunome

Harvard/MIT: Vigneault, Laserson, Lieberman-Aiden, Church
Roche: Egholm, Simen
PGP Time Series Vaccine Experiment

Tracking human dynamic response to vaccination to 11 strains: Hepatitis A+B, Flu A/Brisbane/59/2007 (H1N1)-like, 10/2007 (H3N2)-like, B/Florida/4/2006-like virus
Polio, Yellow fever
Meningococcus
Typhoid, Tetanus
Diptheria, Pertussis
Collect samples at -14d, 0d, +1d, +3d, +7d, +14d, +21d, +28d
V D J usage – CDR3 size distribution

SR1+SR2+TR1

IMGT/LIGM

Count

CDR3 length (nucleotides)
PGP Vaccination Immunome
Self Organizing Map (SOM) clustering
Reading & Writing Genomes Goals

• 2nd Generation BI/O: Reading & Writing
  Engineer cancer- & virus-resistant genomes

• Personal Genomes – Integration tasks
  - Personal Genomes: Environments, Traits,
  - Stem cells, Microbiome/Immunome
  - Synthesis for Causality (CEGS)