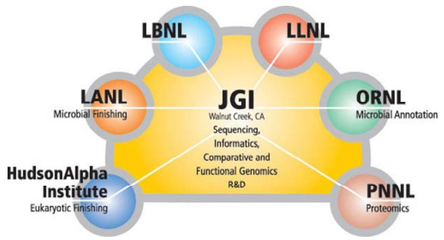
DNA Sequencing of Plants and Microbes for the Optimization of Biofuel Production

Krysta Biniek
Mentor: Lance Green

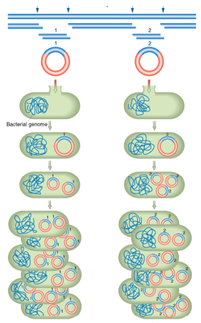
U.S. Department of Energy Joint Genome Institute DOE-JGI

The JGI unites the expertise of six institutions, including five national laboratories, to advance genomics in support of the DOE missions related to clean energy generation. Los Alamos National Laboratory (LANL) is the second largest partner institution of the JGI and specializes in high throughput genome finishing and analysis. The Wet Lab team in B-6 utilizes traditional Sanger sequencing, as well as the Roche 454 and Illumina Genome Analyzer next-generation sequencing platforms.



Sanger Sequencing

Sanger sequencing, invented by Fred Sanger in 1977, is the conventional method for whole genome shotgun sequencing.

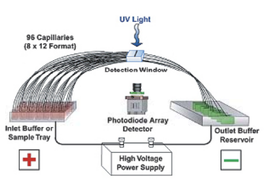


Plasmid Preparation
The DNA is sheared and the fragments are inserted into plasmids.

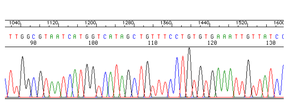
Bacterial Transformation
An electrical shock induces bacteria to incorporate the plasmids.

Cloning and Colony Picking
The bacterial cells are transferred to an agar plate where each cell replicates to form colonies. The colonies are then transferred to a nutrient-rich liquid medium for further growth.

Rolling Circle Amplification
The cells are lysed and the plasmid is amplified using RCA, a process similar to PCR.



Capillary Sequencing
The double stranded DNA is denatured into single strands using heat, and specific primers are annealed to the single strands. DNA polymerase elongates the primer, creating a complimentary strand of DNA by adding deoxynucleotides. A small fraction of the nucleotides, however, are dideoxynucleotides, which lack the 3' hydroxyl group. Thus, when a dideoxynucleotide is added, the chain cannot elongate any further. Because the place where the chain is terminated is random, if millions of chains undergo this process, there will be chains that stop at each nucleotide in the sequence.

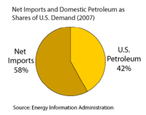


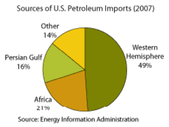
Each A, T, C, and G dideoxynucleotide is fluorescently marked with a different color and capillary electrophoresis is used to separate the DNA strands by size. The sequence is "read" by a camera as the labeled dideoxynucleotides flow past.

Acknowledgements
Mary Campbell, Karen Walston Davenport, Chris Detter, Joe Fawcett, Cheryl Gleasner, Lynne Goodwin, Lance Green, Kim McMurry, Linda Meincke, Chris Munk, Janna Nichols, Beverly Parson-Quintana, Judy Tesmer, Barbara Wendelberger, Patti Wills
Materials Design Institute

DNA Sequencing and Biofuels Production

Need for Biofuels
- U.S. imports 60% of petroleum consumed
  - Disruptions represent threat to national security and economic growth
- Environmental harm caused by burning of fossil fuels



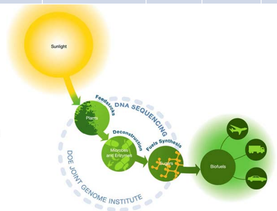


Biofuels derived from cellulosic biomass: Renewable, clean alternative source of energy

Switchgrass Statistics

Average Yield (U.S. tons/acre)	Amount of Fertilizer (lbs N/acre)	Irrigation Requirements (Lw/Le)	Total Cost	Energy Input	Reduction in Petroleum Use	GHG Emissions (as % of gasoline emissions)
5	140	0	$73/ton	8.5% of HHV	93%	13-18%

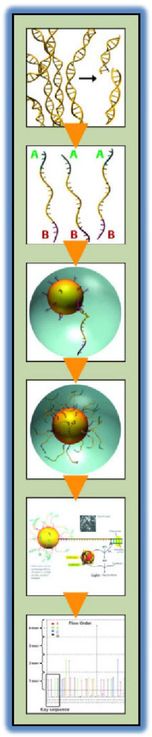
Plants
- Growth rate
- Cell wall composition
- Drought resistance
- Pest resistance



Microbes
- Pathways for cellulose and lignin deconstruction
- Improve efficiency of conversion

Roche 454 Sequencing

The 454 next-generation sequencing platform was released in 2005 and relies on the detection of light generated during DNA synthesis.



Sample Fragmentation
Genomic DNA samples are fractionated into small, 300- to 800-basepair fragments.

Library Preparation
Short adaptors (A and B) are added to each fragment. The adaptors are used for purification, amplification, and sequencing steps.

Capture Bead Attachment
The single-stranded DNA library is immobilized onto DNA Capture Beads. The bead-bound library is emulsified with amplification reagents in a water-in-oil mixture resulting in microreactors containing just one bead with one unique sample-library fragment.

emPCR Amplification
Each fragment is amplified within its own microreactor. Amplification of the entire fragment collection is done in parallel; for each fragment, this results in a copy number of several million per bead.

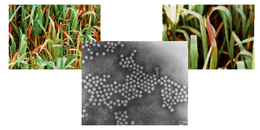
Sequencing by Synthesis
The clonally amplified fragments are enriched and loaded onto a PicoTiterPlate (PTP) device for sequencing, with one bead per well. After addition of sequencing enzymes, the Genome Sequencer FLX Instrument flows individual nucleotides in a fixed order across the wells. Addition of one(or more) nucleotide(s) complementary to the template strand results in a chemiluminescent signal recorded by a CCD camera.

Data Analysis
The combination of signal intensity and positional information generated across the PTP device allows the software to determine the sequence of more than 1,000,000 individual reads per 10-hour instrument run simultaneously.

Virus Interactions Project

Goals
- Determine how viruses may influence biofuel production from grass species communities
- Evaluate any spillover influence of biofuels plantings on virus and vector populations in nearby plantings of other susceptible crop species, such as wheat, oats, and maize

- *Isolate and analyze virus-like particles (VLPs) that may represent unidentified viruses*
  - *Use new technique of sequencing viruses on Roche 454*



Experimental Procedure
- VLPs isolated from *Andropogongerardii* (big bluestem), *Schizchyriumscoparium* (little bluestem), and *Panicum virgatum*(switchgrass)
- Receive amplified, barcodedgDNA fragments
- Pool and prepare barcoded material to run on the 454
- Deconvolute barcodes and generate draft assembly

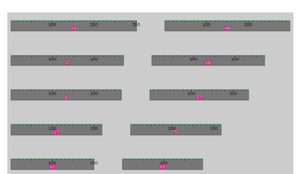
Results

Electrophoresis Gel
- Sample had smaller than optimal fragments
  - Average length: 237 bp
  - Desired length: 500-800 bp

Sequencing Run

	ATB Sample	Desired Values
Raw Wells	425,967	-
Key Pass Wells	420,653	-
Passed Filter Wells	207,405	-
% Dot + Mix	17.43	< 20
% Short	33.27	< 20
% Passed Filter	49.31	> 60
Length Average	182	400 - 500
Total Bases	37,738,504	60,000,000 – 110,000,000

Finishing
- Attempted with 454 Newbler assembly software
- 34 contigs
  - 5-200 reads
  - 100-500 bp in length



BLAST
- BLASTed against nucleotide database at NCBI
- 31 contigs matched bacterial ribosomal RNA
- 3 contigs had no matches (poor quality)

Discussion
No viruses were identified in the sample. It is hypothesized that due to the random amplification of the sample, the viral RNA, if any was present, was drowned out by the abundant bacterial ribosomal RNA. In order to complete the project, this hypothesis must be verified and, if proven correct, a method for isolating viral RNA will need to be devised.

Conclusions
Although the goal of identifying VLPs from biofuel crops was not achieved, the new technique of sequencing RNA on the Roche 454 proved successful. The Virus Interactions Project is just one project in the vast body of research that is being conducted on biofuels by the DOE-JGI. Biofuels have the potential to end our dependence on foreign oil and halt global warming, but the first step in achieving this goal is to exploit the information rooted in the genetic code of plants and microbes in order to optimize the process of growing biomass and converting it to fuel.

LAUR 09-06704