Synthetic metagenomics: converting digital information back to biology

Researchers at the US Department of Energy's Joint Genome Institute (DoE JGI) have developed synthetic metagenomic processes supported by high throughput screening technologies, to translate the vast amounts of "digital" data produced by next-generation sequencing into tangible observable biology.

Introduction

Next-generation sequencing technologies have resulted in the identification of millions of novel predicted genes which cover the full spectrum of known pathways and functional activities. These include large numbers of novel biocatalysts for potential applications in energy and environmental technologies, and potential uses in basic research and applied biotechnology. A major challenge is converting the vast amount of "digital" information (sequencing, bioinformatics, cheminformatics) to observable, functional biology (enzyme activity, cell death, protein expression, etc.).

The DoE JGI have developed a synthetic metagenomics process to overcome this challenge. Genes and pathways are synthesized in a template independent manner, thereby converting digital information into biochemical molecules in a rapid cost effective fashion.

A major bottleneck for commercial production of cellulosic biofuel is the lack of sufficiently active and robust enzymes for the conversion of biomass to glucose. Ionic liquids can be used as a pre-treatment to dissolve cellulose which increases availability of fermentable sugars such as glucose. In order to find suitable, industrially relevant enzyme candidates, synthetic metagenomic approaches were used. For a specific industrial application designed at the DoE Joint Bioenergy Institute (JBEI), enzymes with the following criteria were required: (i) functional enzymatic activity at 70°C, (ii) stability at pH 4.5 and (iii) resistance to ionic liquids. At the inception of the project, there were no known enzymes to exhibit sufficient activity under the required conditions.

Benefits

- **Simultaneously detect colonies and quantify fluorescent markers—use phenotypic selection to pick colonies**

- **Automate workflow from plating and streaking to picking**

- **Record and track sample histories from sample spreading, to picking, replication and re-arraying**

- **Accurately pick up to 30,000 selected colonies per day with > 98% efficiency**
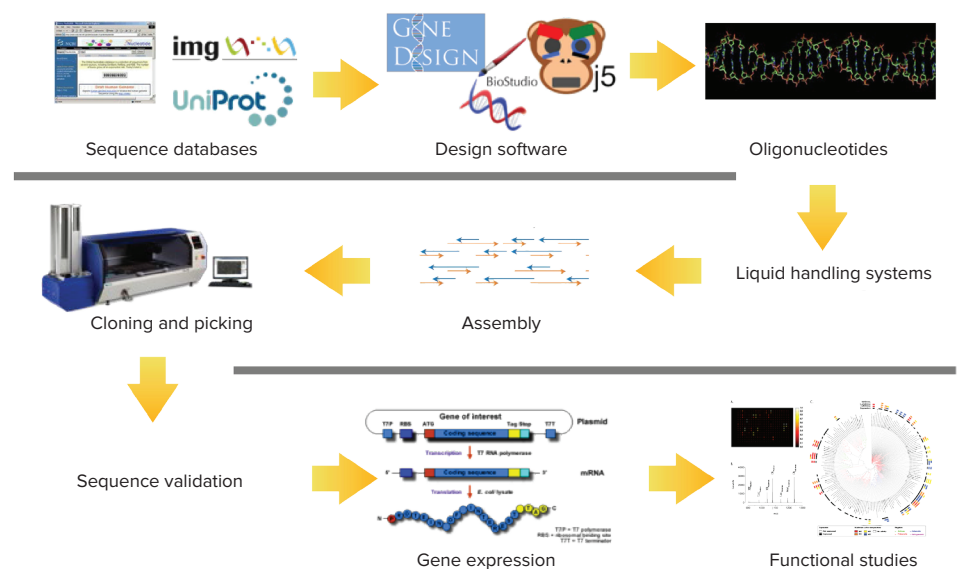
Figure 1. Synthetic metagenomics pipeline: a process adapted by DoE JGI for metagenomic studies. Typical synthesis schema for metagenomic studies. Using a QPix 460 System in the cloning and picking phases drastically shortens workflow timelines and increases efficiency by automating a labor-intensive, error-prone manual task. The fully automated workflow can accurately process thousands of colonies per week, from plating to picking, with full data tracking reducing the cost per base pair.

The first study target involved a family of enzymes known as GH1s that mediate the conversion of cellobiose into glucose. Phylogenetic-guided approaches enabled the selection of 200 genes that captured the maximum functional diversity from the thousands of sequences available. Of the 200 genes, 180 genes were synthesized and cloned into an expression vector using the QPix 460 System (Figure 2).
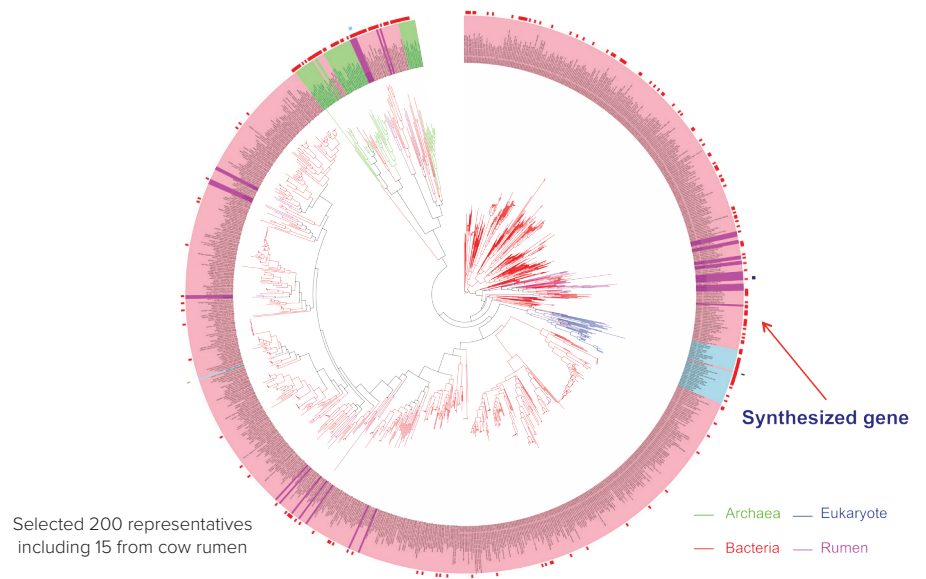
The high-throughput, automated workflow developed at the DoE JGI involves parallel transformation, plating and picking multiple clones from each gene-variant for sequence verification. This is made possible by the QPix 460 System, which offers high throughput plating, streaking and colony picking on a single platform. Once the sequence was verified, desired clones for each gene were transferred into protein production strains.

The cloned GH1 protein production strains were analysed for soluble protein expression. 65% of clones produced soluble protein which was further characterized under the required conditions described above. For example, as an early quality control, a number of enzymes were characterized for activity at different temperatures, revealing candidates with thermostable properties (Figure 3).

To screen the ~15,000 reactions required to evaluate all selection criteria, an ultra high throughput technology was required. Nanostructure-initiator mass spectrometry (NIMS), a high throughput technology based of mass spectrometry[1] was used for this purpose.

The enzyme-substrate reactions were spotted onto a solid support and ionized by a laser.

Enzyme activities were determined by quantifying the ratios of substrates to products (Figure 4). From this large dataset, the top 20 performing enzymes were selected for in depth characterization, five of which were active under the industrially relevant conditions described above. This demonstrates the use of synthetic metagenomics to rapidly identify enzymes with unique properties starting from predicted proteins in public sequence databases.

Selected 200 representatives including 15 from cow rumen

Synthesized gene

— Archaea   — Eukaryote
— Bacteria   — Rumen

Figure 2. Metagenomic phylogenetic analysis to select representative gene families. Tree diagram of the metagenome, including bacteria (red), eukaryotes (blue), cow rumen (purple) and archaea/"extreme prokaryotes" (green). Long branch length indicates increased genetic diversification, whereas short length does not. 200 representative samples from shorter branch areas were selected, including 15 from cow rumen to contribute to phylogeny-driven sample diversity.
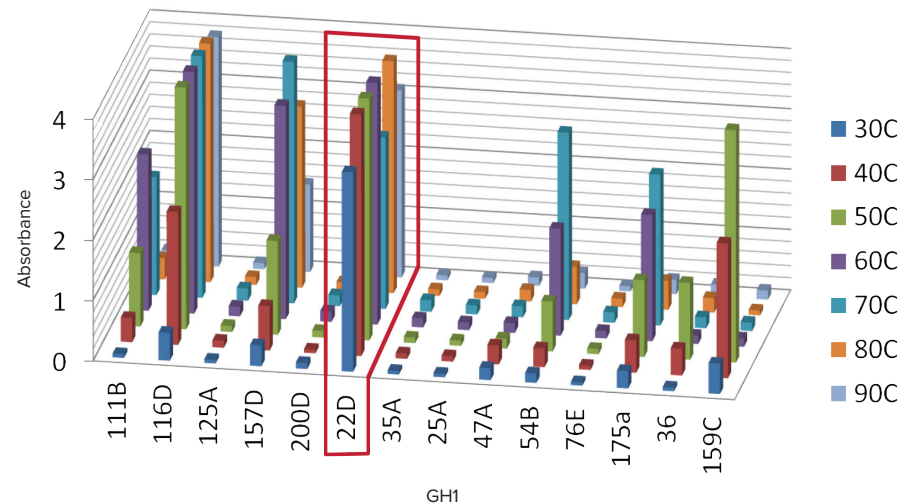
Figure 3. Family GH1 enzyme activity characterization revealing thermostable enzymes. Enzyme activity against colorimetric substrates as measured by optical density. Proteins are assayed for enzyme activity at seven different temperatures, from 30°C to 90°C. Optimal candidates (e.g. 22D) show little or no decrease in activity at 70°C.

Conclusions

This work demonstrates the process developed and utilized by the DoE JGI can successfully convert digital information into biologically relevant, functional enzymes exhibiting novel properties. 180 diverse proteins from the GH1 enzyme family were screened in a high-throughput, automated approach utilizing the QPix 460 System. The system has become an integral part of the automated workflow for efficient metagenomic library generation and management.

Furthermore, this process led to the identification of 5 novel enzyme candidates which exhibit activity against industrially relevant substrates at 70°C, pH 4.5 and in the presence of ionic liquids. Joint BioEnergy Institute (JBEI) is currently studying the potential application of these enzymes in an industrial setting to effectively scale up production of converting biomass into glucose.

References

1. A nanostructure-initiator mass spectrometry-based enzyme activity assay. Northen TR, Lee JC, Hoang L, Raymond J, Hwang DR, Yannone SM, Wong CH, Siuzdak G. *Proc Natl Acad Sci USA.* 2008 Mar 11;105(10):3678-83. doi: 10.1073/pnas.0712332105. Epub 2008 Mar 4. PMID:18319341 [PubMed – indexed for MEDLINE]
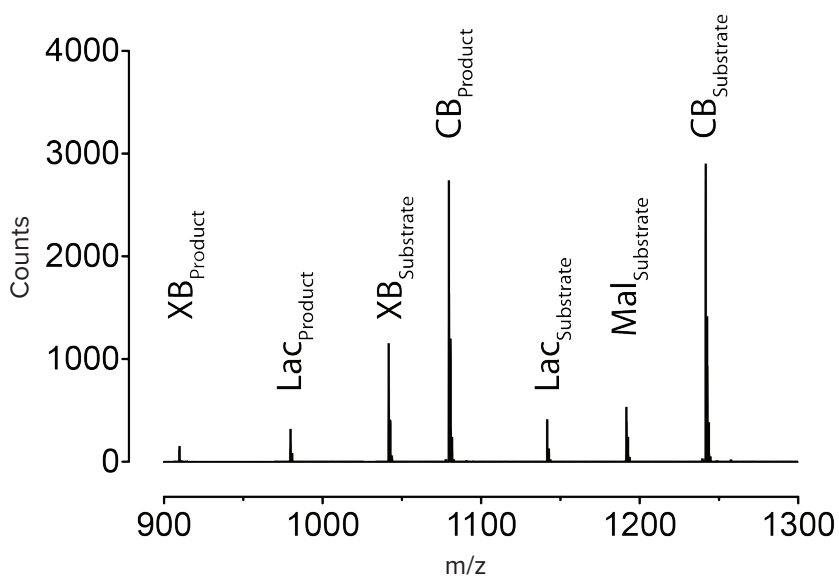
Figure 4. Ratio of substrate: product based on NIMS. NIMS analysis revealed the top 20 performing enzymes based on the optimal reactivity with substrates as measured by the ratio of substrates to products.

MOLECULAR DEVICES