Sequencing and Gene Mining the Largest Collection of Industrially used Acetone-Butanol-Ethanol (ABE) Fermentation Strains

Steven D. Brown[1]* (Steve.Brown@LanzaTech.com), Rasmus O. Jensen,[1] James Winkler,[1] Shilpa Nagaraju,[1] Wayne Mitchell,[1] Vinicio Reynoso,[1] Dawn M. Klingeman,[2] Zamin K. Yang[2], Nicole Shapiro,[3] Natalia Ivanova,[3] Frederik Schulz,[3] Tanja Woyke,[3] David Paez-Espino,[3] Nigel Mouncey,[3] Sean D. Simpson,[1] Michael Köpke,[1] and **Michael C. Jewett[4]**

[1]LanzaTech Inc, Skokie, IL; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]DOE Joint Genome Institute, Berkeley, CA; [4]Northwestern University, Evanston, IL

http://www.lanzatech.com; http://jewettlab.northwestern.edu

Project Goals: The Clostridium Foundry for Biosystems Design (*c*BioFAB) is developing an integrated framework that includes computational modeling, cell-free technologies, system-level omics data, high-throughput anaerobic strain construction and cultivation to rapidly model, design, and predictably engineer industrial clostridia strains to manufacture a variety of fuels (e.g., butanol) and building block chemicals. *Clostridium autoethanogenum* is a model acetogen used in commercial scale gas fermentation to produce bioethanol and it is being further developed by *c*BioFAB and related projects.

Clostridium autoethanogenum can use a wide range of CO, CO_2 and hydrogen containing gases for carbon and energy via the Wood-Ljungdahl pathway. To date, over 11 M gallons (>42 M liters) of fuel ethanol have been produced, avoiding nearly 60k metric tons of CO_2 emissions. A genetic toolbox exists that includes systems for heterologous expression via plasmid and chromosomal integration, gene deletion, CRISPR systems, validated genetic parts and codon adaptation algorithms.

The Acetone-Butanol-Ethanol (ABE) fermentation was one of the first large-scale industrial chemical production processes, widely deployed for the first part of the 20th Century and until 1983 in South Africa (1). The ABE process is a sugar-based fermentation using species in genus *Clostridium,* but not *C. autoethanogenum*. Acetone and butanol are non-native products for *C. autoethanogenum.* In collaboration with the US Department of Energy Joint Genome Institute (JGI), an historic collection of commercial solventogenic clostridia strains have had their genome sequences determined to survey ABE gene variants for heterologous expression, to potentially identify industrial robustness traits, regulatory networks and to investigate CRISPR systems and relationships between bacteriophage and prophage and immunity.

We have generated 273 new ABE genome sequences, with most being *Clostridium beijerenckii* (179) and *Clostridium saccharobutylicum* (75) strains and other species including *Clostridium acetobutylicum, Clostridium butyricum, Clostridium saccharoperbutylacetonicum* and *Clostridium tetanomorphum*. There are 208 strains where ten or fewer contigs represented the

genome sequences and majority of remaining genome assembled from Illumina only data. The 16S rRNA gene count ranged from 1 to 23 copies. Each genome was scanned for genes in the clostridial acetone fermentation pathway, including thiolases (*thlA*), acetoacetate:butyrate/acetate CoA transferases (*ctfAB*), and acetoacetate decarboxylases (*adc*), after which identical sequences were dereplicated with CD-HIT (100% clustering threshold) to identify the unique candidate genes. A total of 30 (from 508 gene calls) thiolases, 32 (506) CtfA, 30 (506) CtfB, and 13 (272) Adc-like unique proteins were identified, which are being screened via cell-free protein synthesis (CFPS) and through strain characterization. In a later study, the collection was mined for unique gene variants for the butanol pathway with 52 thiolases (*thlA*), 40 3-oxoacyl-CoA reductases (*hbd*), 49 3-hydroxyacyl-CoA dehydratases (*crt*), 43 trans-enoyl-CoA reductases (*ter*), 9 acyl-CoA/aldehyde reductases (*acr/ald*) and 236 phosphobutyrylases/butyrate kinases (*ptb-buk*) sequences identified. These gene variants (*thlA-hbd-crt-ter*) are being screened, along with termination enzymes (*acr/ald-ptb-buk*), via cell-free protein synthesis to identify the most promising candidates for *in vivo* butanol production.

About a third of the sequenced strains in collection have a complete CRISPR system with arrays and *cas* operons. All CRISPR systems across the different species are Type-IB, meaning they require multi-subunit effector complex or CASCADE proteins and a 5' protospacer acquisition motif (PAM) for interference. Except for five *C. beijerinckii* strains, all other CRISPR carrying strains have indications of phage and/or prophage infecttion. Of the putative infected, ten strains have spacers targeting the integrated phage and/or prophage. One *C. tetanomorphum* strain has 12 spacers targeting putative phage and/or prophage within its genome with a likely PAM of "CCN".

The new genome resources will facilitate biofuel and chemical production for a variety of biocatalysts and will enable broader biosystems design.

References

1. Jones DT and Woods DR (1986). Acetone-butanol fermentation revisited. *Microbiol Rev.* 50: 484–524.

We acknowledge the U.S. Department of Energy Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program for funding of this project under Contract No. DE-SC0018249. The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI) is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.