FINAL TECHNICAL REPORT

DOE Award Number: DE-FC36-07GO17058

Recipient: Verenium Biofuels, San Diego, CA; Jennings, LA, Cambridge, MA

Project Title: Improvements in Ethanologenic *Escherichia coli* and *Klebsiella oxytoca*

Project Director: David Nunn, Ph.D.

Consortium Members: University of Florida, Gainesville, FL; Massachusetts Institute of Technology, Cambridge, MA; Genomatica, San Diego, CA

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Executive Summary

The current Verenium cellulosic ethanol process is based on the dilute-acid pretreatment of a biomass feedstock, followed by a two-stage fermentation of the pentose sugar-containing hydrolysate by a genetically modified ethanologenic *Escherichia coli* strain and a separate simultaneous saccharification-fermentation (SSF) of the cellulosic fraction by a genetically modified ethanologenic *Klebsiella oxytoca* strain and a fungal enzyme cocktail. In order to reduce unit operations and produce a fermentation beer with higher ethanol concentrations to reduce distillation costs, we have proposed to develop a simultaneous saccharification co-fermentation (SScF) process, where the fermentation of the pentose-containing hydrolysate and cellulosic fraction occurs within the same fermentation vessel. In order to accomplish this goal, improvements in the ethanologens must be made to address a number of issues that arise, including improved hydrolysate tolerance, co-fermentation of the pentose and hexose sugars and increased ethanol tolerance. Using a variety of approaches, including transcriptomics, strain adaptation, metagenomics and directed evolution, this work describes the efforts of a team of scientists from Verenium, University of Florida, Massachusetts Institute of Technology and Genomatica to improve the *E. coli* and *K. oxytoca* ethanologens to meet these requirements.

Task A – Project Management, Strain evaluation, SScF development, and scale up.
Project Location: San Diego, CA and Jennings, LA
Written by: David Nunn

Subtask A.1 – Project Management

1. **Planned activity:** The PI will manage the integration of work with subcontractors to ensure timely information flow, resources and timelines are within scope and budget, and milestones are accomplished. Communications will involve monthly teleconferences and quarterly in-person meetings at Verenium or other convenient location. The PI will be responsible for progress and financial reporting to DOE.

2. **Actual accomplishments:** Integration of work performed by subcontractors, MIT, UF and Genomatica with Verenium was carried out by regularly scheduled monthly teleconferences. Standard operating procedures for pure sugar and hydrolysate fermentations for the E. coli ethanologens trains were transferred to MIT and UF. A revision to Table A, describing modifications to intermediate stage gate targets, was submitted and approved

3. **Explanation of Variance:** Quarterly in-person meetings did not take place although in-person meetings to discuss results and plans did occur semi-regularly with Genomatica and on two occasions during the course of the grant with the University of Florida team.

Subtask A.2 – Biomass Sample Production

1. **Planned activity:** Verenium will operate its 2-ton per day pilot plant to produce representative biomass hemicellulose hydrolysates and lignocellulosic fibers using dilute acid over a range of temperatures from 150 to 190°C. The high and low temperature conditions will be performed at equivalent modified severity factor settings. Therefore, the project will evaluate the influence of temperature on the toxicity of the respective hydrolysates. The modified severity factor will be in the range of 2.5 to 2.8. Verenium will provide subcontractors sufficient materials of the appropriate hydrolysate on a periodic basis to ensure consistent results among the subcontractors. The fiber solids will be separated from the hydrolysates and retained by Verenium for SSF and SScF process.

2. **Actual accomplishments:** Hemicellulose hydrolysates and lignocellulosic fibers were produced across a range of severity factors, with temperature ranges between 150 to 190, varying pressures and acid concentrations. Both energy cane and bagasse hydrolysates were produced. These materials were used for in-house evaluation of the *E. coli* C5 and *K. oxytoca* C6 ethanologens. In addition, in two lots, hemicellulose hydrolysates, prepared from bagasse, were shipped to UF and MIT subcontractors in sufficient quantity to carry out their respective subtasks. Protocols

for over-liming and neutralization of the hydrolysates were provided to standardize conditions for fermentation across the work team.

As per subtask A.2.3, compositional analysis was carried out on energy cane-derived hydrolysate, produced under a range of severities and is shown below. Eighteen runs were carried out with sugars, organic acids and aldehydes determine by HPLC/GC or MS. Amounts are indicated in g/l unless indicated otherwise.

Run #	Temp (°C)	Time (min)	Acid (%)	Log Severity	pH	Glucose (g/L)	Xylose (g/L)	Galactose (g/L)	Arabinose (g/L)	Mannose (g/L)	Cellobiose (g/L)	Total Sugar (g/L)
1	160	6.5	1.4	2.73	0.73	14.8	53.0	2.7	5.3	1.6	1.2	77.5
2	170	5	1.8	3.02	0.52	29.9	29.6	2.2	4.3	1.6	ND	67.5
3	160	6.5	1.4	2.73	0.64	15.5	53.7	2.6	5.1	1.5	1.1	78.5
4	150	5	1.8	2.43	0.64	12.8	74.7	3.3	6.2	1.2	1.6	98.1
5	170	8	1.8	3.22	0.61	30.5	26.9	1.8	3.9	1.3	1.5	64.5
6	170	8	1	2.96	0.83	20.3	40.8	2.0	4.3	1.5	0.9	68.9
7	150	5	1	2.17	0.93	6.3	51.5	1.8	4.1	0.9	2.1	64.6
8	150	8	1	2.38	0.83	10.2	72.0	3.1	6.0	1.4	2.1	92.5
9	170	5	1	2.76	0.87	19.5	64.3	2.2	5.3	1.5	1.2	92.7
10	150	8	1.8	2.63	0.73	10.7	51.0	2.5	5.0	1.3	1.0	70.5
11	170	8	0.5	2.66	1.33	16.1	71.0	3.4	5.9	1.6	1.8	97.9
12	170	3	1.8	2.79	0.64	25.2	40.5	1.8	4.7	1.3	1.1	73.5
13	140	8	1.8	2.34	0.72	10.3	74.2	3.0	6.2	1.2	1.4	94.9
14	150	11	1	2.51	0.9	12.6	79.3	3.3	6.4	1.4	1.7	103.1
15	150	5	2.2	2.51	0.56	13.9	71.3	2.6	6.1	1.1	1.0	94.9
16	175	5	1	2.91	0.93	31.2	55.2	2.5	5.9	2.0	1.4	96.8
17	150.6	7.45	1.8	2.62	0.5	14.5	80.2	3.7	7.0	1.5	1.2	106.8
18	154	5	1.2	2.37	0.79	12.8	67.5	2.7	5.5	1.3	1.5	89.7

Run #	Temp (°C)	Time (min)	Acid (%)	Log Severity	pH	Formic (g/L)	Acetic (g/L)	Levulenic (g/L)	Gallic acid (PPM)	Furoic acid (PPM)	4-Hydroxy benzoic acid (PPM)	Trace Acids (PPM)	Total Acids (g/L)
1	160	6.5	1.4	2.73	0.73	0.36	15.83	0.48	1.80	14.85	3.22	19.87	16.68
2	170	5	1.8	3.02	0.52	0.54	14.13	2.47	2.00	28.05	4.75	34.80	17.14
3	160	6.5	1.4	2.73	0.64	0.00	14.43	0.47	0.00	14.75	2.98	17.73	14.90
4	150	5	1.8	2.43	0.64	0.07	16.17	0.18	0.00	8.00	3.22	11.22	16.42
5	170	8	1.8	3.22	0.61	0.81	13.76	2.79	0.00	33.00	5.05	38.05	17.36
6	170	8	1	2.96	0.83	0.00	13.67	1.00	0.00	24.80	6.06	30.86	14.67
7	150	5	1	2.17	0.93	0.00	11.15	0.04	0.00	8.22	3.07	11.29	11.19
8	150	8	1	2.38	0.83	0.00	13.13	0.15	0.00	9.88	4.18	14.06	13.28
9	170	5	1	2.76	0.87	0.00	15.52	0.46	0.00	20.00	5.38	25.38	15.98
10	150	8	1.8	2.63	0.73	0.11	13.29	0.19	0.00	11.35	3.47	14.82	13.59
11	170	8	0.5	2.66	1.33	0.00	13.87	0.30	0.00	18.00	5.02	23.02	14.17
12	170	3	1.8	2.79	0.64	0.00	13.21	1.81	0.00	23.85	6.44	30.29	15.02
13	140	8	1.8	2.34	0.72	0.06	16.18	0.16	0.00	7.09	4.48	11.57	16.40
14	150	11	1	2.51	0.9	0.09	16.42	0.27	0.00	11.04	4.18	15.22	16.79
15	150	5	2.2	2.51	0.56	0.00	17.51	0.26	1.99	10.11	4.64	16.73	17.77
16	175	5	1	2.91	0.93	0.00	17.31	1.47	0.00	29.70	6.91	36.61	18.78
17	150.6	7.5	1.8	2.62	0.5	0.00	15.54	0.28	0.00	12.20	3.75	15.95	15.82
18	154	5	1.2	2.37	0.79	0.00	16.98	0.27	0.00	12.40	5.84	18.24	17.25

Run #	Temp (°C)	Time (min)	Acid (%)	Log Severity	pH	Furfural (g/L)	5-HMF (g/L)	4-Hydroxy benzaldehyde (PPM)	Syringaldehyde (PPM)	Trace Aldehydes (PPM)	Total Aldehydes (g/L)
1	160	6.5	1.4	2.73	0.73	5.7	0.27	57.75	18.75	76.50	5.97
2	170	5	1.8	3.02	0.52	10.7	1.04	84.10	27.20	111.30	11.74
3	160	6.5	1.4	2.73	0.64	6.4	0.28	71.35	22.95	94.30	6.68
4	150	5	1.8	2.43	0.64	3.8	0.11	43.80	16.45	60.25	3.91
5	170	8	1.8	3.22	0.61	11.6	1.19	91.20	27.45	118.65	12.79
6	170	8	1	2.96	0.83	9.1	0.80	99.95	37.20	137.15	9.90
7	150	5	1	2.17	0.93	3.2	0.04	40.80	13.85	54.65	3.24
8	150	8	1	2.38	0.83	2.7	0.07	49.75	18.95	68.70	2.77
9	170	5	1	2.76	0.87	5.5	0.52	80.70	29.90	110.60	6.02
10	150	8	1.8	2.63	0.73	5.7	0.13	49.90	15.75	65.65	5.83
11	170	8	0.5	2.66	1.33	4.4	0.37	71.10	24.50	95.60	4.77
12	170	3	1.8	2.79	0.64	7.1	0.72	80.85	24.85	105.70	7.82
13	140	8	1.8	2.34	0.72	4.5	0.06	30.15	13.85	44.00	4.56
14	150	11	1	2.51	0.9	4.0	0.12	44.40	19.05	63.45	4.12
15	150	5	2.2	2.51	0.56	4.7	0.14	51.70	17.15	68.85	4.84
16	175	5	1	2.91	0.93	6.7	1.09	102.50	38.65	141.15	7.79
17	150.6	7.45	1.8	2.62	0.5	4.7	0.15	53.20	20.25	73.45	4.85
18	154	5	1.2	2.37	0.79	4.6	0.18	63.15	26.10	89.25	4.78

3. **Explanation of Variance:** Only materials derived from bagasse was provided to subcontractors.

Subtask A.3 – Demonstration of Baseline Fermentations

1. **Planned activity:** Verenium will establish baseline fermentations of current *E. coli* strain and *K. oxytoca* strain using pure sugars plus inhibitors and representative sugarcane and energy cane materials produced in Task A.2. This task will benchmark the current state of the fermentation technology for further improvements in tolerance to ethanol concentration and inhibitors.

2. **Actual accomplishments:** Completed as part of the "2 Month" audit. The fermentations completed within satisfactory limits with respect to time, rates, yields, and titers. Results presented in 1Q and 2Q-2008 reports.

3. **Explanation of Variance:** None

Subtask A.4 – Continuous culture adaptation

1. **Planned activity:** Verenium will carry out adaptation experiments using continuous cultivation techniques. This will compliment the techniques used by sub-contractors and has not been done previously with these strains. Continuous cultures will be

established using industrial medium components plus representative sugarcane bagasse and energy cane hemicellulose dilute acid hydrolysates. Cultivations will be run in turbidostat mode. Cultures will be evaluated form periodic sampling of the continuous culture. Evaluations will be made of improved tolerance to inhibitors (including ethanol) by plating on solid media containing elevated levels of inhibitors and in batch fermentations.

2. **Actual accomplishments:** None

3. **Explanation of Variance:** Due to concerns about applicability of this method for generating strains relevant to a fed-batch process this activity was removed from the project after consultation and approval of DOE

Subtask A.5 – SScF Process Demonstration

1. **Planned activity:** In order to achieve 8 – 10% ethanol concentrations in from lignocellulosic biomass, Verenium believes that the hemicellulose sugars must be co-fermented with the cellulosic fibers in the SSF-type reaction. The primary goal of the preceding tasks (and sub-contracts) has been to develop strains of *E. coli* and *K. oxytoca* that can achieve the target ethanol concentrations in the presence of inhibitory dilute acid hydrolysates.

2. **Actual accomplishments:** The ethanologenic *K. oxytoca* strain was tested initially for the ability to tolerate additions of hydrolysate to a clean sugar fermentation and found to be tolerant of <15% (v/v) additions of hydrolysate, short of the 30% target that would be required for an SScF process. An initial trial of fermentation with the ethanologenic *E. coli* of both hydrolysate sugars and fiber was carried out at 300 mL bench scale in batch mode. While the results indicated that the strain was able to use the glucose efficiently, the xylose was not used, indicating a diauxie, and the fiber was poorly converted, due to the relatively high pH of the fermentation and the inability to effectively use cellobiose. For this reason, SScF efforts were put on hold to address the strain deficiencies.

3. **Explanation of Variance:** Because of the need to improve the performance of the *E. coli* ethanologen in utilization of xylose and cellobiose, this activity was not completed within the course of this grant.

Subtask A.6 – Lab-scale demonstration of Improvements

1. **Planned activity:** In order to bench-mark improvements made to strains in Task A4 and B4, as well as by the sub-contractors, Verenium will evaluate the improved strains in 3 – 5 L bench-scale fermentations using representative bagasse and energy cane hydrolysates. These benchmark fermentations will determine if improvements are made and which approach is achieving the greatest improvement.

2. **Actual accomplishments:** All strains generated through the course of this work were evaluated and benchmarked to previous strains at multiple scales of fermentations, from 300ml fleaker fermentations to 3 liter BioFlo fermentations. Each

fermentation was evaluated for performance by measuring sugar consumption, ethanol production, ethanol yield, and ethanol concentration.

3. **Explanation of Variance:** None

Subtask A.7 – Pilot Plant Scale-up

1. **Planned activity:** Contingent upon passing the 18 month Stage Gate review, Verenium will demonstrate the improved strains and procedures developed in Tasks 4 – 6 at pilot scale using sugarcane bagasse and energy cane.

2. **Actual accomplishments:** None

3. **Explanation of Variance:** As the decision at the 18 month stage gate review was to bring the project to a close, this subtask was not begun.

.

Task B. Characterization of hydrolysate tolerance and identification of potential "tolerance" genes.

Objective: To identify genes, both endogenous to host strains and from environmental metagenomic libraries, whose expression increases tolerance to components of hydrolysates and are candidates for incorporation into improved ethanologen strains.

Strategy and rationale: Although the most ideal method to assess improvements in hydrolysate tolerance of ethanologen strains would be in a representative lab-scale C5 or C6 fermentation, this does not allow for the high throughput identification of genes whose expression could lead to improved hydrolysate tolerance.

For this reason we propose to develop a surrogate ethanologen host / microtiter plate-based growth assay to determine the relative tolerance to hydrolysates components and identify genes whose expression leads to increased hydrolysate tolerance. For both *E. coli* and *Klebsiella oxytoca*, strains will be generated that have only the ethanol fermentation pathway intact, directly linking growth to ethanol production. A dilution-titration strategy will be used to develop a microtiter plate assay for assessing tolerance to individual hydrolysate components within a complex mixture. Once minimum inhibitory concentrations are determined for each relevant component, small insert expression libraries, derived from each host strain or from environmental metagenomic DNA will be introduced and pools of clones screened in the microtiter format to identify clones that confer increased tolerance to hydrolysates. Pools of clones conferring increased resistance will be broken out to identify individual genes that may then serve as targets for expression amplification to improve hydrolysate and process performance.

Subtask B.1. Generation of Host Strains.

1. **Planned activity:** Using a conventional marker exchange strategy, gene-specific deletions will be constructed in both the *E. coli* and *Klebsiella* strains to eliminate all functional fermentation pathways other than that of the introduced ethanol pathway from *Zymomonas mobilis* (*pdc, adhA, adhB*).

2. **Actual accomplishments:** Completed for both E. coli and K. oxytoca ethanologen strains. For *E. coli*, the deletions included the *frd* operon (to eliminate succinate production) and *ldhA* (to eliminate lactate production). In addition, an *mgsA* deletion was made to eliminate the production of methylglyoxal. For *K. oxytoca*, the *budAB* operon was deleted (to eliminate butanediol production) in *ldhA*. An *mgsA* deletion was also made to eliminate the production of methylglyoxal.

3. **Explanation of Variance:** None

Subtask B.2. Microtiter plate assay for hydrolysate tolerance.

Subtask B.2.1. Assay for C5 fermentation conditions.

1. **Planned activity:** To develop a surrogate microtiter assay for the C5 fermentation conditions, a "dropout" titration strategy will be used. In this strategy a real hydrolysate will first be over-limed, as per the method for preparing hydrolysate for large-scale fermentation, and then analyzed for composition of known hydrolysate components, including sugars, furfurals, organic acids, inorganic acids and lignin-derived polyphenolics. Standards and analytical methods have been developed for each. This material will then be diluted by titration into a "mock" hydrolysate, containing representative concentrations of all components but the component to be tested. In this manner the effects of any given component can be assessed while retaining the synergistic or antagonistic contribution of all other components. Trace elements and other components of the C5 fermentation medium will be added to maintain a constant concentration across plate. Alternative pHs may be tested, with validated buffers added to maintain a constant pH. Incubations of each of the strains to be tested will be carried out using a combined incubator/plate reader and OD_{600} measurements carried out throughout the experiment. Plates will be sealed with a gas-impermeable membrane to ensure microaerobic conditions.

2. **Actual accomplishments:** Completed with the following modifications. Initially, a "dropout" titration strategy was proposed to develop a screen/selection for identifying over-expressing clones, using a defined mixture of materials to generate a "mock" hydrolysate, minus each of the components for testing. Due to the complexity and variability of the hydrolysates, it was decided that two parallel approaches were most appropriate, one using single inhibitor compounds, and the other using a complete hydrolysate. Incubations of the strains were carried out using a combined incubator/plate reader and OD_{600} measurements were performed throughout the scoping experiments.

3. **Explanation of Variance:** None

Subtask B.2.1. Assay for C6 fermentation conditions.

1. **Planned activity:** To develop a surrogate assay for the C6 fermentation, a similar strategy will be used with the following variations. Hydrolysate will not be over-limed but will be used as representative of the material from washed cake and the diluent will be a medium composed of trace elements and media components containing either 5 or 10% glucose.

2. **Actual accomplishments:** Completed with the following modifications. Instead of using hydrolysates, washed C6 cakes are instead being predigested with enzymes and the resultant sugar syrup, contained released inhibitory compounds is being used to titrate the sensitivities of the individual strains. We have found this to be a more representative challenge condition than using C5 hydrolysates as the levels of exposure to these compounds will be minimal due to the washing steps that are part of the Verenium C6 process.

3. **Explanation of Variance:** None

Subtask B.3. Identify tolerance genes from genomic and metagenomic libraries.

Subtask B.3.1. Generation of libraries

1. **Planned activity:** Small insert libraries the genomes of *Klebsiella* M5A1, *E. coli* B, W, and K12 strains will be constructed. Genomic DNA will be prepared and randomly sheared to generate 2-5kb fragments of DNA for ligation into Verenium expression vectors to construct genomic libraries of > 10^5 individual clones. This will insure complete representation of each genome and the expression of the entire complement of coding sequences from a regulated vector promoter.

2. **Actual accomplishments:** Completed

3. **Explanation of Variance:** None

Subtask B.3.2. Selection of metagenomic small insert expression libraries.

1. **Planned activity:** More than 2000 metagenomic libraries, prepared from environmental genomic DNA from a wide variety of ecotypes from around the world have been constructed by Verenium. A subset will be selected from those likely to be exposed to components found in biomass hydrolysates.

2. **Actual accomplishments:** Completed, with five representative metagenomic libraries selected for introduction into the Verenium ethanologens and testing for increased tolerance to hydrolysate inhibitors.

3. **Explanation of Variance:** None

Subtask B.3.3. Transformation of libraries into ethanologen strains.

1. **Planned activity:** *E. coli* and *K. oxytoca* ethanologen strains, generated as described in Task B.1 will be individually transformed with the small insert libraries and pools of transformants will be generated. Numbers of individual clones per pool will depend on the size of each representative library and the number of cells inoculated per microtiter dish well. For example, if each well is inoculated to a starting OD_{600} of 0.1, this would contain ~ 2 x 10^7 cells per well. If a library is composed of 10^5 (for genomic library) to 10^6 (for metagenomic library) and a pool size of 1000 clones is selected per well, a single microtiter dish of 96 wells would be sufficient to screen a genomic library and ten plates would be sufficient to screen a metagenomic library.

2. **Actual accomplishments:** Completed for *E. coli* ethanologen strain with pool sized sufficient to screen selected libraries. We were unable to transfect the *K. oxytoca* strain with an efficiency required for obtaining a sufficient pool size for subsequent screening. Efforts were made to modify the *K. oxytoca* strain to improve transfection frequencies but were unsuccessful.

3. **Explanation of Variance:** As per above, unable to generate sufficient K. oxytoca pool of transfected clones for further screening.

Subtask B.3.4. Selection of "tolerant" clones.

1. **Planned activity:** Pools of transformed hosts will be inoculated into microtiter plates containing minimal inhibitory concentrations of each of the components to be tested, within the context of "mock" hydrolysate and incubated to select for clones able to grow in the presence of the inhibitory component. Controls containing only vector DNA will be inoculated to determine background level resistance and possible generation of spontaneous resistant mutants.

2. **Actual accomplishments:** The selection of tolerant clones from metagenomic libraries was carried out initially with three of the five identified metagenomic libraries. Selections for resistance to over-limed hydrolysate, furfural and ethanol were carried out using the Bioscreen plate selection method described above. While a number of clones were identified from initial selections, none of these were able to reproducible confer the corresponding resistances after breakout, sequencing and confirmation.

   The selection of tolerant clones from the overexpression libraries however, was able to identify a large number of genes that were confirmed as conferring resistance to the originally selected inhibitor after breakout, sequencing and confirmation. These included genes coding for efflux pumps, DNA repair proteins and DNA recombination genes. These genes were then tested for the ability to confer hydrolysate tolerance and improve performance of the *E. coli* ethanologen after integration into the chromosome, as per Subtask

B.5.

3. **Explanation of Variance:** See above.

Subtask B.3.5. Breakout of clones.

1. **Planned activity:** Pools from which clones are able to grow in the presence of hydrolysate will be broken out and individual clones tested for conferring tolerance.

2. **Actual accomplishments:** Completed, as routine part of workflow for confirmation of "tolerance" clones.

3. **Explanation of Variance:** None

Subtask B.3.6. Confirmation of clones.

1. **Planned activity:** Plasmid DNA will be prepared from isolated clones and retransformed to ensure that the tolerance phenotype is conferred by cloned DNA.

2. **Actual accomplishments:** Completed, as routine part of workflow for confirmation of "tolerance" clones.

3. **Explanation of Variance:** None

Subtask B.3.7. Sequencing and annotation.

1. **Planned activity:** Plasmids that are verified by retransformation will be sequenced to determine the nature and number of independent genes conferring resistance to each individual inhibitor.

2. **Actual accomplishments:** Completed, as routine part of workflow for confirmation of "tolerance" clones.

3. **Explanation of Variance:** None

Subtask B.4. Validation of clones in lab-scale fermentations.

1. **Planned activity:** Because growth of the engineered strains has been used as a surrogate indicator of tolerance to hydrolysate during ethanol production, plasmids that have identified by their ability to confer increased hydrolysate tolerance in the microtiter plate format will be introduced into the most current respective C5 or C6 strain for testing of improved performance in lab-scale fermentation.

2. **Actual accomplishments:** All clones obtained as described above, were tested

in shake flask cultures for demonstrable improvement in growth and ethanol production. Only those clones which showed improvement were carried forward to next task.

3. **Variance:** None

Subtask B.5. Chromosomal expression of tolerance genes and functional expression.

Subtask B.5.1. Introduction of homologous tolerance genes.

1. **Planned activity:** Introduction of individual homologous genes into chromosome and up-regulation of gene expression followed by evaluation of constructed strains under original conditions used to evaluate increased inhibitor tolerance. These strains will also be tested under more relevant fermentation conditions to determine effectiveness.

2. **Actual accomplishments:** Nine of eleven genes isolated from the *E. coli* overexpression library were introduced and up-regulated by introduction of a suite of constitutive promoters. Performance was tested for growth in the presence of inhibitor compounds and all nine showed demonstrable difference over wild-type control for growth and production of ethanol.

3. **Explanation of Variance:** None

Subtask B.5.2. Introduction of heterologous tolerance genes.

1. **Planned activity:** Introduction of individual heterologous genes into chromosome and up-regulation of gene expression followed by evaluation of constructed strains under original conditions used to evaluate increased inhibitor tolerance. These strains will also be tested under more relevant fermentation conditions to determine effectiveness.

2. **Actual accomplishments:** Five heterologous genes were obtained from the expression *K. oxytoca* libraries in the *E. coli* ethanologen. Each were introduced into the chromosome of the *E. coli* ethanologen at the site of the deleted *ldhA* gene and regulated by a suite of constitutive promoters. Evaluation of fermentation performance of these strains compared to the wild-type control. Three of the five genes demonstrated improved performance when highly expressed.

3. **Explanation of Variance: None**

Subtask B.5.3. Testing of strains in lab-scale fermentations.

1. **Planned activity:** Testing of improved strains in 3L fermentations for improved performance.

2. **Actual accomplishments:** Three strains containing the overexpressed heterologous genes were tested in 3L Bioflo fermentations for improved performance vs. a wild-type control. A demonstrable improvement in fermentation performance was seen for two of the three tolerance genes but further characterizations were not carried out.

3. **Explanation of Variance:** Due to termination of the program, further validation of these clones was not carried out.

Subtask B.5.4. Combination of tolerance genes in final strains.

1. **Planned activity:** Introduce tolerance genes identified in above activities into a single strain and evaluate performance against benchmark strain.

2. **Actual accomplishments:** None

3. **Explanation of Variance:** As the decision at the 18 month stage gate review was to bring the project to a close, this subtask was not begun.

Patents: Two applications have been prepared describing the introduction of heterologous genes into the *E. coli* ethanologen to improve performance.

Publications / Presentations: None

Project Location: University of Florida

Written by: Peter Turner/Lonnie Ingram

Task number: C. Adaptation to inhibitors and analysis of RNA

C.1. Construction of ethanologenic *E. coli* that lack requirement for complex nutrients

1. **Planned Activities**: The goal was to construct an ethanologenic strain of E. coli able to grow well in a defined minimal medium, and to identify the genetic changes responsible for improved growth and ethanol production.

2. **Actual Accomplishments**: The ethanologenic strain LY180 was isolated by a combination of metabolic evolution, and deletion and insertion of defined genes (Yomano, 2008). The resulting strain grows well in AM1, a simple and inexpensive minimal medium,and ferments sugars derived from hemicellulose very efficiently. Deletion of the *mgsA* gene (methylglyoxal synthase) was found to improve simultaneous utilization of different sugars in LY180 (Yomano, 2009). A large portion of the genomic sequence of the parent strain KO11 has been determined. RNA expression levels were determined using microarrays for strain LY180 in comparison with the ultimate parent KO11 and for the intermediate strains LY160 and LY168. We were not able to determine which genetic changes in the LY180 genome were beneficial. However, analysis of the metabolic pathways perturbed in LY180 from the RNA expression data suggested that iron metabolism is particularly important in improved growth of LY180.

3. **Explanation of Variance**: None.

C.2. Construction of ethanologenic *K. oxytoca* without need for complex nutrients

1. **Planned Activities**: In a strategy that paralleled our work with *E. coli*, we constructed an ethanologenic *K. oxytoca* strain that expressed the *Z. mobilis adhA*, *adhB* and *pdc* genes, and then selected for variants with improved growth in minimal medium by metabolic evolution. Finally we had intended to define which mutations contributed to improved performance in minimal medium.

2. **Actual Accomplishments**: Serial transfers of *K. oxytoca* strain XZ530 generated derivatives that grow strongly in AM1 with 9% xylose. Selections for increased resistance to the inhibitors furfural and hydroxymethyl furfural (HMF) have been successful (see sections C.5). We were not able to identify genomic changes that contributed to improved growth during the period that the work was funded.

3. **Explanation of Variance**: The lack of a fully-assembled genomic sequence for K. oxytoca M5a1 hindered progress in identifying genomic changes in *K. oxytoca*, as did the lack of a commercially available gene expression microarray for *K. oxytoca*.

C.3. Improvements in ethanol titer during fermentation

1. **Planned Activities:** The goal was to develop strains capable of producing 8% w/v (80 g/L) ethanol.

2. **Actual Accomplishments:** We were able to isolate strain LY180, which is able to grow and ferment in AM1 medium containing up to 15% xylose initially. The highest ethanol titer we have obtained has been 67 g/L, short of the target of 80 g/L. The obvious strategies to boost ethanol production were tried, such as increasing the initial xylose level above 15%. However, growth is inhibited when the xylose concentration is greater than 15%, presumably because of osmotic stress, despite the presence of the osmoprotectant betaine. We found that spiking AM1-14% xylose cultures mid-way through fermentation with an additional 3% xylose increased the final ethanol titer obtained. We tried fermentation at 30ºC instead of 37ºC, which slows fermentation but results in a slightly higher final yield of ethanol. Incubation at 35ºC was compared with 37ºC and with 30ºC, and gave results very similar to those obtained at 37ºC. We tried supplementing AM1 medium with various supplements, including iron, magnesium, and trace metals. A combination of additional $MgSO_4$ and Fe-EDTA was most beneficial and gave a slight increase in ethanol titer to approximately 70 g/L. Experiments with added ethanol were done to determine the ability of fermentation to contiune in the presence of high ethanol levels. LY180 was grown for 48 h in AM1-14% xylose until approximately 40 g/L ethanol had been produced. The fleakers were then spiked with 0 to 40 g/L additional ethanol. Ethanol values were measured immediately after this addition (at 48h); and after an additional 24 h of incubation (72h). When the total ethanol level was at 80g/L (at 48h), almost no additional ethanol was produced by fermentation, despite the presence of residual unfermented xylose. At lower total ethanol levels (<75 g/L EtOH), fermentation continued so that the final yield at 72h was greater than the ethanol level at 48h. Microarray RNA expression data was generated for strains KO11, LY160, LY168, and LY180, all grown in AM1-14% xylose. The data did not indicate clearly which pathways in LY180 are altered to cause improved ethanol production. The microarray data suggested that LY180 has evolved more efficient ways to take up and metabolize iron, consistent with our supplementation experiments showing that added Fe-EDTA can improve both cell growth and ethanol production.

3. **Explanation of Variance:** Direct selections for improved ethanol tolerance by transferring ethanologenic strains in the presence of added EtOH were not productive. We may have reached an intrinsic limit in the ability of *E. coli* to continue fermenting when a high level (~80 g/L) of ethanol is present in the medium.

C.4. Improvement in resistance to acetate during xylose fermentation

1. **Planned Activities:** The goal was to develop strains that are able to ferment in the presence of 1.0% (10 g/L) acetate.

2. **Actual Accomplishments:** We obtained acetate-resistant strains from two sources. Strain BJ103, an *E. coli* derivative, was isolated by selecting for continued growth in the presence of increasing concentrations of acetate. BJ103 is able to grow in AM1-9% xylose in the presence of 15 g/L acetate. A second acetate-resistant strain, MM105, was obtained by selection for growth with increasing concentrations of hydrolysate (see section C.7). MM105 was found to have significant resistance to acetate, a natural

component of hydrolysate. RNAs were prepared from both BJ103 and MM105 grown with and without acetate, and compared with the parent strains EMFR17 and LY180, respectively, grown under the same conditions. Analysis of the resulting data has yielded a short list of genes whose expression is perturbed in a similar way in both BJ103 and MM105.

3. **Explanation of Variance:** There was not sufficient time to determine the impact of individual expression changes on acetate tolerance.

C.5. Improvement in resistance to furfural during xylose fermentation

1. **Planned Activities:** The goals were to isolate ethanologenic strains of *E. coli* and *K. oxytoca* resistant to 0.1% (1 g/L) furfural, and to identify mutations and changes in expression levels that are beneficial.

2. **Actual Accomplishments:** A series of strains has been derived from LY180 with increasing resistance to furfural. These were obtained by selecting for growth in the presence of increasing furfural concentrations. EMFR9, an early strain in the series, grows at concentrations up to 1.5 g/L furfural, and subsequent strains EMFR17 and EMFR35 have higher resistance to furfural. We have therefore exceeded our initial target of growth in the presence of 1 g/L furfural. Good progress has been made in determining the genetic mechanisms involved in tolerance to furfural. Surprisingly, down-regulation of the oxidoreductases that act on furfural is beneficial. Although YqhD and DkgA can both reduce furfural to the less toxic furfuryl alcohol, they do so by depleting the pool of available NADPH, and thereby reduce cellular processes that require NADPH. These processes include synthesis of the amino acids methionine and cysteine, amongst others, and assimilation of sulfur. The mutation that causes silencing of the adjacent *yqhD* (b3011) and *dkgA* (b3012) genes is caused by insertion of IS10 into *yqhC* (b3010), a previously uncharacterized gene. We have shown that YqhC acts as a positive regulator of the *yqhD-dkgA* operon. MM105 and MM205, *E. coli* strains that were selected as resistant to hydrolysate (section C.7), were shown to have mutations in *yqhD* and *yqhC*, respectively, and to have increased furfural tolerance in comparison with the parent strain. These results confirm the central importance of the *yqhC-yqhD-dkgA* region in furfural resistance. EMFR9 and subsequent strains also contain a mutation in the *ydhM* (*nemR*) gene, which results in constitutive expression of the adjacent *nemA-gloA* operon. The *ydhM* mutation confers resistance to methylglyoxal, a natural toxic product that accumulates during unbalanced metabolism and serves to reduce growth under these conditions. Recently we have studied the gene expression changes that occurred in the development of EMFR17 and EMFR35 from EMFR9, and have identified candidate genes likely to be involved in the increased level of furfural resistance seen with EMFR17 and EMFR35. We found that resistance to furfural confers cross-resistance to the related compound 5-hydroxymethyl furfural (HMF), a second important inhibitor found in hemicellulose hydrolysate. The mechanisms for detoxifying furfural and HMF appear to be similar but not identical. Selections for *E. coli* strains resistant to HMF have been carried out, yielding strains that grow in the presence of 2.75 g/L HMF. In parallel with the *E. coli* selections, we have selected for derivatives of *K. oxytoca* (BJ157) that are resistant to 1.75 g/L furfural, and are in the process of determining which genetic changes are responsible for increased resistance.

3. **Explanation of Variance:** None

C.6. Improvement in resistance to desalted hydrolysate

1. **Planned Activities:** The original goal was to isolate strains of ethanologenic *E. coli* and *K. oxytoca* able to grow in desalted hemicellulose hydrolysate.

2. **Actual Accomplishments:** We obtained ethanologenic *E. coli* strains that grow well in NaOH-neutralized hydrolysate that has not been desalted (see section C7), and the proposed work with desalted hydrolysate therefore became unnecessary.

3. **Explanation of Variance:** None

C.7. Improvement in resistance to NaOH-neutralized hydrolysate

1. **Planned Activities:** The objective was to develop strains capable of growth in full-strength NaOH-neutralized hydrolysate.

2. **Actual Accomplishments:** Long-term selections for growth in the presence of increasing hydrolysate has yielded two *E. coli* strains: MM105 and MM205. MM105 grows with 50% hydrolysate (Hz) without added complex nutrients, and is the best of all the strains we have tested for growth in hydrolysate. We have constructed a genetically improved version of MM105 by removing extraneous DNA (FRT sites) that was introduced into the sites of gene deletions. The resulting strain is lacking many of the "scars" that were introduced into the genome in previous constructions. Transfers of this derivative of MM105 (LY197) in hydrolysate is continuing. Characterization of MM105 has shown that this strain has resistance to furfural and to HMF (by mutation in *yqhD*), and to acetate, as well as giving good growth in the presence of a high hydrolysate concentration. RNA was isolated from MM105 in the presence of acetate, furfural, HMF, and hydrolysate, and compared with RNA of the parent strain LY180 grown under similar conditions. We have identified a number of changes in gene expression that might be involved in improved growth in hydrolysate, but have not had time to test the effect of individual mutations.

3. **Explanation of Variance:** Hydrolysate-resistant strains of *Klebsiella oxytoca* have not been isolated by direct selection, but the lessons learned from hydrolysate-resistant *E. coli* should enable us to transfer beneficial mutations into the closely-related *K. oxytoca*.

C.8. Combining mutations

1. **Planned Activities:** Initially the goal was to combine beneficial mutations in both E. coli and K. oxytoca.

2. **Actual Accomplishments:** We were successful in constructing a derivative of MM105 that contained mutations in *yqhC*, *yqhD*, and *ydhM* (MM105 Δ*yqhC yqhD*$^{-}$ Δ*ydhM*). However, this mutant did not show improved growth in furfural in the Bioscreen C growth

curve machine compared with the parent strain MM105, which has a single mutation in *yqhD*. Growth of the mutants in hydrolysate was also not improved in comparison with MM105. Improved furfural and hydrolysate resistance may require mutation in an additional gene or genes as well as the mutations in the *yqhC-yqhD-dkgA* and *ydhM-nemA-gloA* regions. We are currently trying to identify beneficial mutations in the EMFR9/EMFR17/EMFR35 series besides the known ones that result in reduced expression of the yqhD-dkgA genes.

3. **Explanation of Variance:** None

C.9. Combining selections

1. **Planned Activities:** The objective was to develop strains with better performance in hydrolysate by combining selection protocols.

2. **Actual Accomplishments:** Combined selections became somewhat unnecessary as strain MM105 already combines resistance to acetate and hydrolysate, and has some resistance to furfural and hydroxymethyl furfural.

3. **Explanation of Variance:** We decided that combining mutations was a more efficient strategy than combining selections.

Summary

We met the targets for resistance of *E. coli* to acetate and furfural, and substantial progress has been made toward understanding the mechanisms of furfural and HMF resistance. Many of the lessons learned with *E. coli* are applicable to the closely related *K. oxytoca*. The goal of developing a strain able to ferment cellulosic hydrolysate was essentially achieved with strain MM105, which combines many of the desired features of strains that have been selected to date, including resistance to acetate, furfural, and hydrolysate. We came close to the target of 80 g/L EtOH, although continued work in this area is still required.

Publications / Presentations:

Yomano LP, York SW, Zhou S, Shanmugam KT, Ingram LO (2008). Re-engineering *Escherichia coli* for ethanol production. Biotechnol Lett 30:2097-2103.
http://www.springerlink.com/content/1220142t112n0v31/

Yomano LP, York SW, Shanmugam KT, Ingram LO (2009). Deletion of methylglyoxal synthase gene (*mgsA*) increased sugar co-metabolism in ethanol-producing *Escherichia coli*. Biotechnol Lett 31:1389-1398.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2721133/?tool=pubmed

Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO (2009). Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-resistant ethanologenic *Escherichia coli*. Appl Environ Microbiol 75:4315-4323.
http://aem.asm.org/cgi/content/full/75/13/4315?view=long&pmid=19429550

Miller EN, Jarboe LR, Turner PC, Pharkya P, Yomano LP, York SW, Nunn D, Shanmugam KT, Ingram LO (2009). Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. Appl Environ Microbiol 75:6132-6141. http://aem.asm.org/cgi/content/full/75/19/6132?view=long&pmid=19684179

Turner PC, Miller EN, Jarboe LR, Baggett CL, Shanmugam KT, Ingram LO. YqhC regulates transcription of the adjacent *Escherichia coli* genes *yqhD* and *dkgA* that are involved in furfural tolerance. Submitted to Appl Environ Microbiol.

Miller EN, Turner PC, Jarboe LR, Ingram LO. Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing *Escherichia coli* LY180. Submitted to Biotechnology Letters.

Final Progress Report

Project Location: Massachusetts Institute of Technology

Reporting Period: Final Report
Date of Report: December 10, 2009
Written by: Christie Peebles and Brian Pereira

Task number: D.1. Library Construction

1. **Planned Activities**: We planned to construct 10-12 libraries of regulators of transcription, RNA degradation, and protein translation. A summary of the libraries constructed during this time can be found below.

2. **Actual Accomplishments**: Previous evidence suggesting that artificial transcription factors can be used to improve complex phenotypes (e.g. see **Nat Biotechnol (2003) 21:1208-14)** as well as our own research (e.g. see **Science (2006) 314:1565-8** and **Proc Natl Acad Sci U S A (2008) 105:2319-24**)) have served as the basis for library construction. In principle, targets have been chosen that have a high potential of globally altering the intracellular environment. So far, libraries of rpoD (the principal sigma factor of *E. coli*) and rpoA (a subunit of the RNA polymerase with regulatory functions) have been constructed by error-prone PCR (epPCR) and used to improve phenotypes. These targets were selected because there is substantial evidence in the literature that describes their structure and physiological roles.

   The experimental procedure (epPCR) is standard and relatively easy to troubleshoot. Phenotypic diversity quantification (as outlined in task D.2) has been used to direct the construction of novel rpoA libraries with improved properties. In particular, mutagenesis has been focused to different regions of the alpha subunit in order to reduce the search space and increase the probability of isolating improved mutants. This way, 6 libraries of rpoA have been constructed, in addition to rpoD libraries that were optimized. In addition, saturation libraries with varying mutation rates on surface-exposed amino acids in the C-terminal domain of rpoA have been constructed by site-directed mutagenesis protocols. One such library was the αCTD*t, which conferred the highest degree of diversity of the rpoA libraries.

   Similar to the optimization of rpoA libraries, we focused on building new rpoD libraries. Targeting mutations to particular DNA-binding amino acids in sigma D proved ineffective, but we still believed that reducing the sequence space as a means of optimizing the *rpoD* libraries was possible. One way of attaining this goal would have been choosing a different set of amino acids to target for mutagenesis. Because it was not obvious which residues to target for a new design, we instead chose to select the regions that contained the surface amino acids that were previously selected (spanning a couple of hundred base pairs each).

   The libraries of the region of sigma D that binds the -10 promoter hexamer, which we denoted -10*L and -10*H according to their mutation frequency, were constructed by epPCR of a base pair sequence between amino acids 422 and 456; this region includes

all of region 2.4 and parts of regions 2.3 and 3.0 (see supplementary material of (Campbell *et al.*, 2002)). The libraries of the region of sigma D that binds the -35 promoter hexamer, denoted -35*L and -35*H, were similarly constructed by targeting the region from amino acid 546 and until the stop codon; this region contains most of region 4.1 and all of 4.2 (same source as above). Fig. D1 illustrates the design of these four libraries schematically.

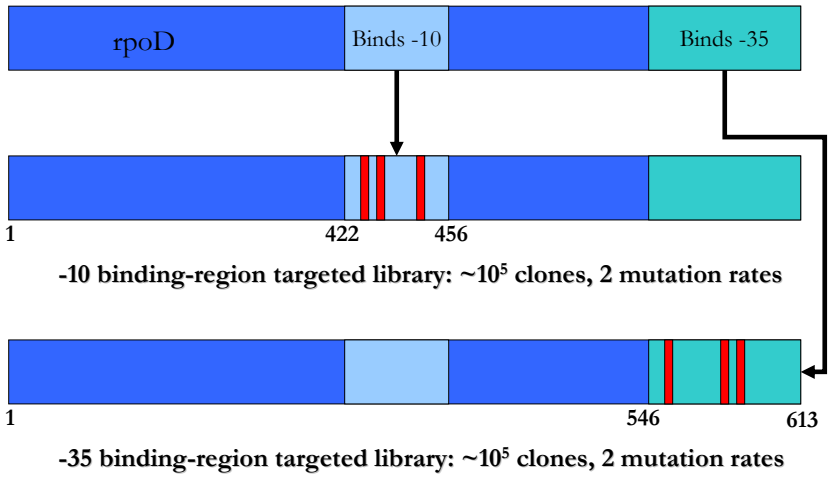

Figure D1. Design of the libraries targeted to the 2.4 and 4.2 regions of sigma D. Generalized mutations are indicated in red for schematic purposes only; they are not intended to show the actual location or frequency of mutations in our libraries.

The concept of global transcription machinery engineering (gTME) is being extended to the processes of RNA stability/degradation and translation. Three additional genes were selected for testing their potential for improving complex phenotypes. The first two genes, the endonuclease RNase E (rne) and the RNA pyrophosphohydrolase (rppH) control the stability of large subsets of mRNAs (e.g. see **Annu Rev Microbiol 2007, 61:71-87** and **Science (2008) 451:355-8**). The third gene, the ribosomal protein S1 (rpsA), is involved in translation of most if not all mRNAs (e.g. see **J Mol Biol 1998, 280:561-9**). Both, rne and rpsA are highly autoregulated with the involvement of the 5'-untranslated region (5'-UTR) of the respective mRNA. All 3 genes have been cloned in a low-copy plasmid under the control of the trc promoter and for both, rpsA and rne a second construct was made where the gene is controlled by its own promoter and 5'-UTR. Several libraries were constructed for all 3 regulator genes (Figure D2). Fragments 1, 2 and 3 of rne and rpsA were mutagenized with a mutagenesis rate of ~2-4 mutations per fragment (size of each fragment: ~500 bp). Furthermore, rne was divided into two parts, 1-3 and 4-6. Both fragments and the whole rpsA gene were mutated at two different mutation rates: 2 and 8 mutations/kb. The rppH gene with a size of approx. 500 bp was not divided into fragments and the whole gene was mutated again at two different mutation rates of approx. 3 and 6 mutations per gene.

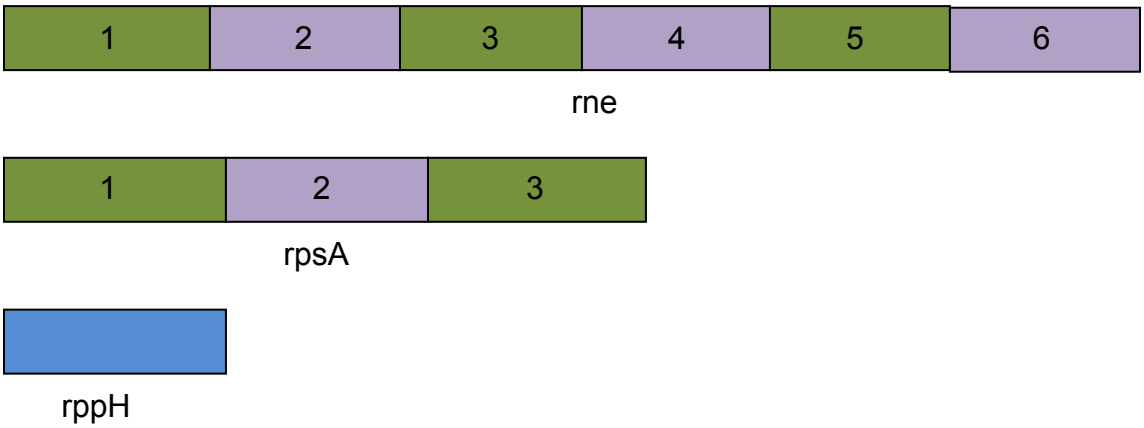

Figure D2. Fragments for epPCR mutagenesis of regulators involved in RNA stability and translation.

We have recently worked to recreate and improve the libraries discussed above. We have also worked on creating whole-gene and regional libraries of rpoS. The rpoS gene codes for sigma factor 38 which is important in regulating expression in response to stresses. At this point, these new libraries have not been completed.

3. **Explanation of Variance**: Not applicable

Task number: D.2. Library Evaluation

1. **Planned Activities**: We planned to measure the phenotypic diversity of our libraries by using the developed intracellular pH assay as a metric.

2. **Actual Accomplishments**:
   After constructing the rpoD libraries, we used the pH_i assay for quantifying the divergence. We also quantified the divergence of a library of *rpoD* with high mutation frequency (rpoD*H) in order to compare the new libraries to the best sigma factor library available then. In addition, we included the αCTD*t in our analysis to judge the libraries in light of the library that showed the highest level of divergence so far. The results are shown in Figure D3.

   From Figure D3, we can observe that (i) localizing mutagenesis to regions in charge of DNA-binding is effective for imparting phenotypic diversity; (ii) such diversity is significantly greater than that of non-targeted mutagenesis to rpoD; (iii) such diversity compares favorably with that of previously-optimized *rpoA* libraries; (iv) mutagenesis of the region that binds the -35 promoter hexamer can introduce higher phenotypic diversity than that of the region that binds the -10 promoter hexamer; and (v) higher sequence diversity translates into higher phenotypic diversity for the case of the targeted libraries.

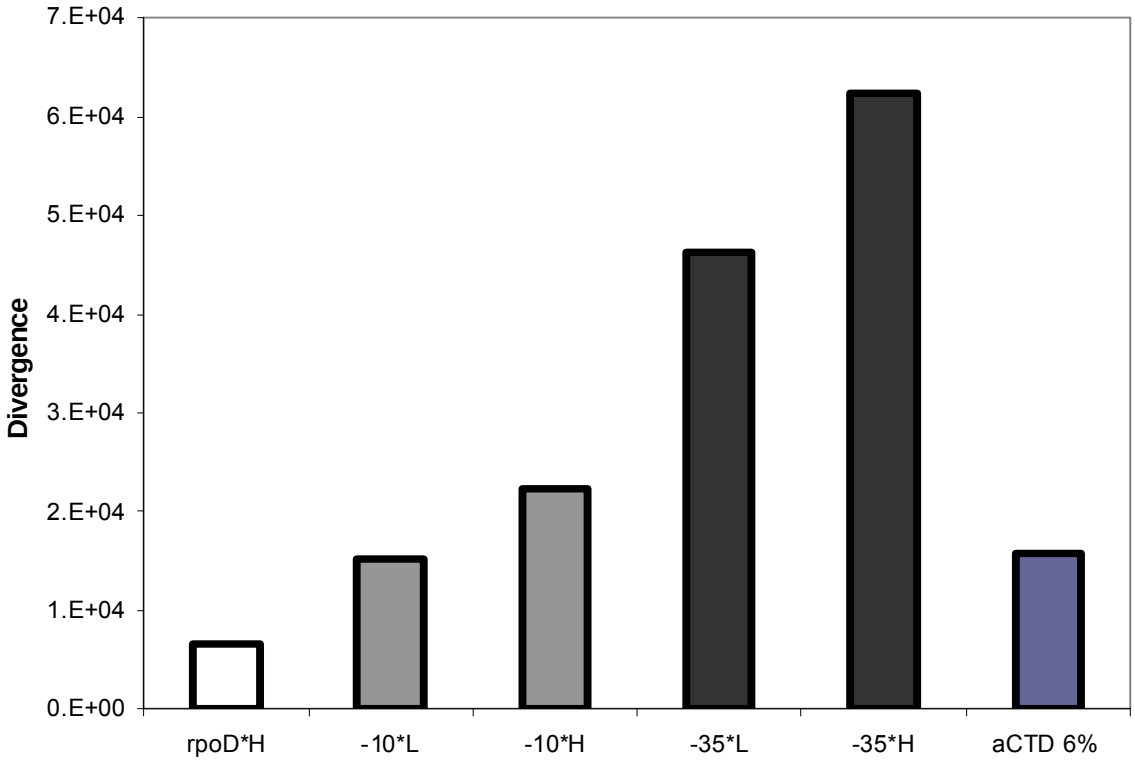

Figure D3. Divergence of the libraries targeted to the 2.4 and 4.2 regions of sigma D.
The divergence was calculated using pH_i in growing conditions, and contrasted to that of a rpoD library with high mutation frequency throughout the coding region (rpoD*H) and to that of the αCTD*t library that had showed highest divergence so far.

All libraries of the RNA stability/translation group show a lower diversity in our pH_i assay as compared to the rpoA*t18% library (Figure D4). Of the libraries quantified, the two rpsA libraries of subfragments 1 and 2 are the best with approx. 50% of the diversity of the rpoA*t18% library. Two technical factors might contribute to the lower diversity of the new libraries from the RNA stability/translation group. First, the library size of most libraries is smaller as for the rpoA*t18% library (10^3-10^4 as compared to 10^5). And second, all libraries were generated by epPCR with mutations covering a relatively large fraction of the genes. Therefore they are not highly focused to specific positions in the gene responsible for its activity like the optimized rpoA*t18% library. Optimization and improvement of the new libraries is still in progress to achieve larger library sizes. Based on the previous results special emphasis will be on the rpsA libraries.

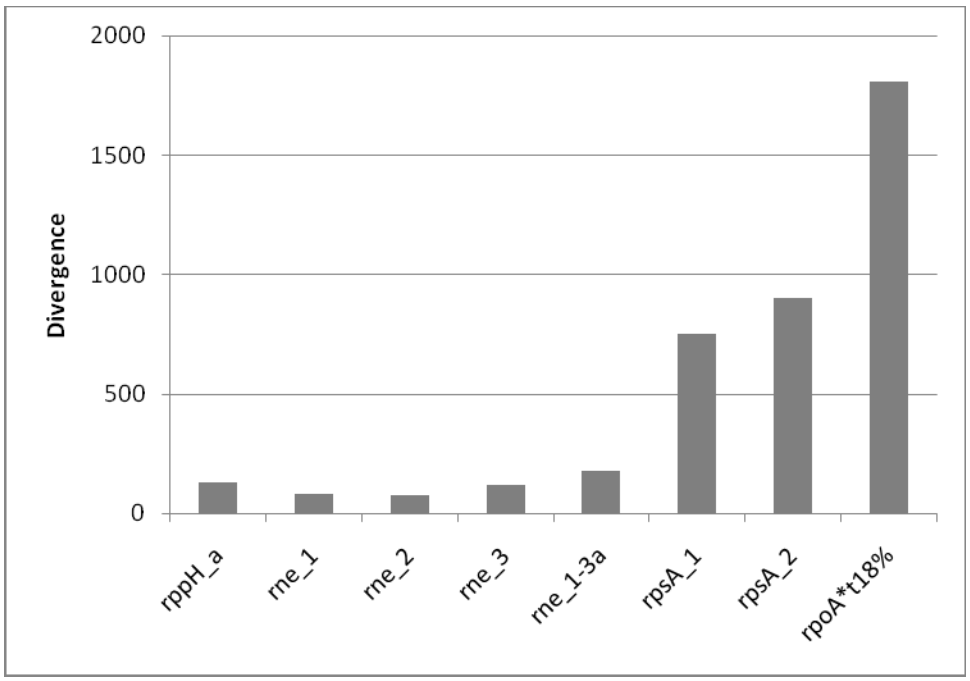

Figure D4. Diversity quantification of different libraries addressing RNA stability/translation.
All libraries are random mutagenesis libraries of the whole gene or subfragments thereof. As comparison the diversity of a rpoA library with focused mutations in the C-terminal domain (rpoA*t18%) is shown. rppH_a is a library of the whole gene of rppH. rne_1 to 3 and rne_1-3a are libraries of subfragments of rne. rpsA_1 and 2 are libraries of subfragments of rpsA.

The developed metric would be used to evaluate any newly constructed libraries.

3. **Explanation of Variance**: Not Applicable

Task number: D.3. Develop Selection Assays

1. **Planned Activities**: We planned to develop strain selection assays under stress induced by high ethanol or acetate concentrations, high osmotic pressure, low pH, and biomass hydrolysate.

2. **Actual Accomplishments**: The initial selection assays were developed during the first year of the project. We continued to fine tune these selection assays and search for methods that may more efficiently screen our libraries for the desired phenotypes. For example we recently implemented a microfluidic screen based on a system developed in the Stephanopoulos lab to select for strains that grew in the presence of 30 g/L ethanol in minimal media (AM1).

3. **Explanation of Variance**: Not applicable

Task number: D.4. Elicit desirable phenotypes in *E. coli*

1. **Planned Activities**: Libraries of mutant regulators that show the greatest divergence as described in Task D.2 are to be transformed into *E. coli* and assayed under the various stress conditions described in Task D.3. Once effective mutant regulators are isolated, they will be re-transformed into a clean background and evaluated in terms of stress tolerance and fermentation performance.

2. **Actual Accomplishments**: We have screened the highest diversity rpoA (αCTD*t) and rpoD (-10*H and -35*H) libraries. Both libraries are targeted to specific areas of the protein. We have found a few rpoD mutants that exhibit increased survivability to overlimed hydrolysate with 50 g/L ethanol and a few rpoD mutants that have increased growth rates in 60 g/L ethanol. However, these mutants have not shown better ethanol yield or productivity in small scale bagasse hydrolysate fermentations than the control strain. The rpoA library has given a few mutants with better survivability in ethanol and better ethanol productivity and yield in the small scale bagasse hydrolysate fermentations, but this may be dependent on the specific batch of hydrolysate as reproducibility is a concern.

   We re-transformed these rpoD and rpoA mutants into clean *E. coli* background and reevaluated them. The enhanced phenotypes, with respect to ethanol survivability, were only exhibited by the original isolates and not by the re-transformed clones. This suggests that the isolates contain background genomic mutations that are responsible for improving the phenotype. To confirm this, we effected the loss of plasmid from some of the original isolates; preliminary tests on these strains indicate that loss of the plasmid does not negatively affect the phenotype.

3. **Explanation of Variance**: Not applicable.

Task number: D.5. Combine mutant regulators for multiple phenotypes

1. **Planned Activities**: We planned to evaluate the performance of co-expressing mutant regulators in eliciting tolerance to multiple stresses.

2. **Actual Accomplishments**: None to date. We are still screening mutant regulators for increased stress tolerance.

3. **Explanation of Variance**: Not applicable.

Task number: D.6. Evaluate fermentation characteristics of cells harboring mutant regulators

1. **Planned Activities**: We planned to evaluate mutant regulators for ethanol productivity and yield.

2. **Actual Accomplishments**: We have performed preliminary small scale evaluations of all the mutant regulators found in a screen of the rpoA and rpoD libraries. The best mutant produces ~50 g/L ethanol in 36 hrs and reaches a final ethanol concentration of

~60 g/L in over-limed hydrolysate, while the control produces ~36 g/L ethanol in 36 hrs and reaches a final ethanol concentration of ~50 g/L. These preliminary results show promise that an increase in stress tolerance will result in an increase in ethanol production.

3. **Explanation of Variance**: Not applicable.

Publications / Presentations: None.

Final Progress Report

Project Location: Genomatica
Written by: Mark Burk

Genomatica will contribute to the overall objective of creating improved ethanologenic strains by using *in silico* modeling to promote both product (ethanol) formation and carbon substrate availability in strains while ensuring that the ethanol-producing strains grow robustly. The following specific tasks were set forth to accomplish this task:

1. Construction of *Escherichia coli* KO11-RD1 and *Klebsiella oxytoca* genome-scale metabolic models
2. Model-facilitated analysis of gene expression data
3. Development of OptKnock strain design strategies for ethanol production in *Klebsiella oxytoca*
4. ^{13}C flux analysis to understand metabolic bottlenecks and establish mechanisms for attaining high ethanol yield and robust growth in *E. coli* and *K. oxytoca*.

Task E.1: Construction of *Escherichia coli* W and *Klebsiella oxytoca* genome-scale models

1. **Planned activity:** Construct a metabolic model for E. coli W and K. oxytoca based on the genome sequences provide by Verenium.

2. **Actual accomplishments:** Models have been constructed for *E. coli* W and *K. oxytoca* M5A1. Specifically, we have evaluated all the reactions and the genes in this model and determined the accuracy and completeness of all the gene-protein-reaction associations and predicted metabolites. The completed models were then evaluated to eliminate any energy or redox-generating cycles in the network. Finally, the models were interrogated to check if it could predict growth on glucose and xylose. These predictions were compared with the growth yields predicted by the thoroughly validated *E. coli* K12 model in the Genomatica model repository. Maps were then constructed to depict the network graphically.

3. **Explanation of Variance:** None

Task E.2: Analysis of gene expression data

1. **Planned activities**: Use the derived metabolic models to facilitate the evaluation of the gene expression data and provide hypotheses for experimental testing.

2. **Actual accomplishments**: We have used our models to do a number of strain comparisons in support of UF expression array work, including the comparisons *E.*

coli strains KO11, and the furfural resistant strains LY160, LY168 and LY180. The objective of this set of microarray studies is to determine the genotype changes that confer better growth abilities on each of the strains over the precursor from which it has been constructed, either by introducing additional genetic manipulations or by adaptive evolution. A global analysis of the differentially regulated genes in all the comparison datasets has been conducted and the genes have been tabulated. The four tables that compare each of the strains with its precursor and LY180 with KO11 were classified on the basis of functional groups. These were been sent to Dr. Ingram's group in Florida. This data analysis using our models has enabled us to come up with hypotheses for better growth of LY180 in minimal medium in contrast to KO11 and these are related to the lack of iron availability in KO11. We also analyzed another dataset aimed at understanding the tolerance of the acetate-resistant strain BJ103 as compared to its precursor, EMFR17. We have generated tables comparing the resistant strain with the precursor, both in the presence and absence of acetate. This information was passed on to Dr Ingram. Visualization of the data using our proprietary software, SimPheny, provided us information on some key central metabolism genes that are differentially regulated in the acetate resistant strain. We also observed cellular response to the presence of acid by upregulation of genes that encode for ammonia-producing enzymes.

3. **Variance:** None

Task E.3: Development of OptKnock strain designs for ethanol production

1. **Planned activities**: Develop OptKnock designs for growth-coupled ethanol production in K. oxytoca.

2. **Actual accomplishments**: We used our strain design algorithm, OptKnock, to identify the best designs for ethanol production in *K. oxytoca* with xylose and glucose as carbon substrates. A report outlining the designs and their predicted impact on ethanol production and growth was compiled. Briefly, three OptKnock runs were calculated for growth on glucose: 1) wild type *Klebsiella oxytoca* (strain BW21), 2) The strain described in 1 with acetate kinase (ACKr) and/or PTAr functionality removed, and 3) strain 1 with pyruvate formate lyase (PFLi) functionality removed. Key modeling assumptions include (i) anaerobic growth conditions; (ii) a glucose or xylose uptake rate of 10 mmol/gdw/hr; (iii) unlimited ammonia uptake, and (iv) a minimum non-growth-associated maintenance requirement of 3 mmol/gDCW/hr. The predicted knockouts strategies were targeted at elimination of byproduct formation and towards maximizing the availability of redox for ethanol formation. Some of these knockouts were straightforward, such as deletion of fumarate reductase and lactate dehydrogenase; others were non-intuitive. For example, we found that a deletion in transhydrogenase significantly increases the availability of NADH for ethanol formation. This is because transhydrogenase converts catabolic NADH to NADPH utilized for biosynthesis. We also found a way to channel flux through the Entner-Doudoroff pathway and away from glycolysis. This increased the predicted ethanol yield to more than 96% of the theoretical maximum of 0.51 g/g glucose. This method of carbon utilization, specifically glucose utilization, is very similar to what is observed in *Zymomonas mobilis,* an organism with high conversion yields of glucose to ethanol.

3. **Variance:** None

Task E.4: – Adaptation of E. coli and Klebsiella ethanologens to neutralized hydrolysate.

1. **Planned Activities:** Carry out adaptations E. coli and Klebsiella ethanologens to increased concentrations of neutralized hydrolysate using proprietary Genomatica Enhanced Evolutationary Engineering technology.

2. **Actual Accomplishments**: Two *E. coli* strains with improved tolerance to neutralized hydrolysate were isolated (improved from 30 to 60%) and are being tested in 3L and 10L scale to demonstrate improved performance at larger scale. No *Klebsiella* clones were obtained that demonstrated significant improvement.

3. **Explanation of Variance**: None

Publications / Presentations: None.