REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
16-05-2014	R&D Status Report (Quarterly)	17-02-2014 to 16-05-2014

4. TITLE AND SUBTITLE
Rapid Parallel Screening for Strain Optimization

5a. CONTRACT NUMBER
HR0011-12-C-0062

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
NA

6. AUTHOR(S)
Chuck Merryman

Tony Yee

5d. PROJECT NUMBER
NA

5e. TASK NUMBER
NA

5f. WORK UNIT NUMBER
NA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
J. Craig Venter Institute
9704 Medical Center Drive
Rockville, MD 20850

8. PERFORMING ORGANIZATION REPORT NUMBER
HR0011-12-C-0062.8

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DARPA, MTO
Prog: Living Foundries ATGC
675 N Randolph St
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NA

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES
NA

14. ABSTRACT

Progress has been made on identifying biosensors that will be used to report on the fermentation yields of industrially relevant biological compounds. Screening of the desired chemicals, sequencing and annotation of isolated microbes was completed previously. Construction and optimization of reporter systems for testing candidate transcription factors was completed previously. Similarly, we have optimized the screening process so that empirical verification of transcription factor candidates is less labor intensive than originally envisioned. During this period we have finalized the identification of biodegradation gene clusters likely to be controlled by the desired transcription factors. Nearly 40 high quality candidates have been identified. Five DNA sequences representing transcription factors and their corresponding operators have been assembled and we will complete assembly of remaining candidates in batches of five. Although we need to verify genotypes, we will begin assaying responsiveness of the reporter to exogenously added inducers within two weeks. Upon identification of a suitable and well-behaved transcription factor, the biosynthetic pathway for the corresponding biochemical will be installed and we will begin work on overproducing the molecule.

15. SUBJECT TERMS
NA

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Tony Yee
UU	UU	UU	UU	5	**19b. TELEPHONE NUMBER** *(include area code)* 301 795 7133

J. Craig Venter™
INSTITUTE

Rapid parallel screening for strain optimization

Report Title:	Quarterly R&D Status Report
Report Number:	HR0011-12-C-0062.8
Reporting Period:	17FEB14 to 16MAY14
Contract No.:	HR0011-12-C-0062
Performing Organization:	J. Craig Venter Institute 9704 medical Center Drive Rockville, MD 20850 USA
Principal Investigator:	Chuck Merryman

Abstract

Progress has been made on identifying biosensors that will be used to report on the fermentation yields of industrially relevant biological compounds. Screening of the desired chemicals, sequencing and annotation of isolated microbes was completed previously. Construction and optimization of reporter systems for testing candidate transcription factors was completed previously. Similarly, we have optimized the screening process so that empirical verification of transcription factor candidates is less labor intensive than originally envisioned. During this period we have finalized the identification of biodegradation gene clusters likely to be controlled by the desired transcription factors. Nearly 40 high quality candidates have been identified. Five DNA sequences representing transcription factors and their corresponding operators have been assembled and we will complete assembly of remaining candidates in batches of five. Although we need to verify genotypes, we will begin assaying responsiveness of the reporter to exogenously added inducers within two weeks. Upon identification of a suitable and well-behaved transcription factor, the biosynthetic pathway for the corresponding biochemical will be installed and we will begin work on overproducing the molecule.

Sponsored by
Defense Advanced Research Projects Agency
Microsystems Technology Office (MTO)
Program: Living Foundries: Advanced Tools and Capabilities for Generalizable Platforms (ATCG)
Issued by DARPA/CMO under Contract No: HR0011-12-C-0062

Table of Contents:

Summary

In total, 108 compounds have been used for enrichment culture and 85 compounds produced colonies (when used as the sole source of carbon and energy). These have been sequenced and the genomes annotated. Analysis resulted in the identification of 38 genomes with candidate transcription factors that likely respond to one of the 108 chemicals. Constructs for cloning and evaluating transcription factors (over an improved dynamic range) were completed previously. We are now in a position to empirically identify transcription factors that can report on compound biosynthesis. Refactored constructs are now being produced and we will begin screening them after their genotypes are assessed.

Introduction

The overall goal in this contract is to link cell-based production to cell survival and thereby make the engineering of new microbial strains that produce industrially relevant biochemicals routine. Recent synthetic biology techniques can make billions of variant cells. Although, many potentially informative mutants are easily made, product yield can only be determined in a few of these. The majority of industrially relevant biomolecules are not chromophores, naturally discernible, or otherwise easily detected. Nevertheless, genetic circuits are capable of linking chemical production to discernible signals such as growth or color intensity. Such a system would allow numerous mutants and mutant combinations to be examined quickly. Genetic circuits needed to screen mutant populations in parallel rely upon the availability of an appropriate biosensor that activates a reporter gene in a product dependent fashion. These are not routinely available. In this project, genes for two-component and one-component signaling systems (that respond to industrially relevant biomolecules) are identified. To demonstrate that such sensors can be used to maximize product yield, one sensing system will be further engineered. We will reformat this sensor so that it drives expression of a reporter such as an antibiotic resistance marker. This sensor/resistance cassette, and a biosynthetic pathway capable of producing the molecule to which the sensor responds, will be placed within a heterologous host that does not have an overlapping pathway. Basal synthesis of the targeted chemical (by the orthogonal biosynthetic pathway) activates the sensor and increases transcription of the resistance marker (i.e. reporter). In other words, the fermentation product is also the sensor ligand and thus, biosynthesis drives production of the reporter and a discernable cell phenotype. Targeted, genome-wide and barcoded alterations to the host genome will then be installed. Variants with better and better chemical production are selected by virtue of increased reporter activity.

Methods, Assumptions and Procedures

To identify candidate transcription factors for experimental evaluation we previously processed 13 sequenced genomes and have now completed the task. A BLAST database of all biodegradation gene clusters that we could identify in sequencing repositories was initially made. Best 'hits' for the experimentally isolated colonies were identified by querying the database with each genomes annotated open reading frames (ORFs). A stringent cutoff score was used (1e-80) so that only ORFs very similar to experimentally investigated biodegradation enzymes were labeled. Potential degradation pathways were collected by parsing the output and selecting regions where two or more putative degradation enzymes occurred within a sliding window of 10 ORFs and were co-located with a one component, two component, or TonB sensor system. This approach mimics the gene-mapping algorithm developed for marking biosynthetic gene clusters; however, in this case, catabolic genes are used to map potential degradation pathways. One assumption that limits the effectiveness of this approach is that the genes encoding a catabolic cluster are co-located (both naturally and with respect to the 10's or 100's of assembled sequence fragments that represent a 'next-generation' genome). Typically, catabolic enzymes for a particular pathway are co-located. Likewise, incomplete assembly of genomes (i.e. fragmentation) typically occurs at long repetitive sequences such as ribosomal operons and insertion elements, etc. Repetitive regions are rare within the boundaries of catabolic clusters. Nevertheless, half of the genomes did not ultimately yield a candidate catabolic cluster likely able to utilize the corresponding chemical. We will explore other funding opportunities to eliminate and more fully understand the nature of this shortcoming.

Although manual inspection of candidate gene clusters was adequate for processing the 13 genomes from the preliminary phase, a brute-force approach proved too cumbersome with the full dataset. This issue was addressed by focusing on pathways that were appropriate for a specified chemical. All carbon sources must ultimately supply material to central metabolism. Thus, the structure of an initial carbon source constrains probable intermediates as they are transformed and funneled into central metabolism. We thus used elucidated degradation pathways and predictions from the University of Minnesota Biocatalysis/Biodegradation Database to identify whether our target chemicals likely utilized known steps (i.e. enzymes) downstream of initial processing. Each genome was then probed with enzymes from the protocatechuate, benzoate, catechol, or other reference pathway as dictated by the structure of the parent compound. This yielded 1-3 high-quality candidate clusters for nearly 40 chemicals. In the next phase, these candidates will be assembled into the reporter system and experimentally validated.

.

Results and Discussion

Screening of chemicals, processing of the resulting microbes, and construction of necessary plasmids etc., was completed previously. We also improved the dynamic range of the reporter system and decreased the labor necessary to characterize transcription-factor candidates by employing an automated system for recording results. Candidate transcription factors and their corresponding operons are now ready to be reformatted, assembled and tested for induction by an exogenously supplied effector. We have not identified any significant E. coli toxicity or utilization of our biochemical candidates. The full set of molecules remains accessible. The targets have been ranked for construction based on a combination of factors. Of primary concern was the confidence in the transcription factor. This was modulated by other potential issues such as the likelihood that the substrate would require transport. Construction of reporter constructs is ongoing but five will undergo the initial round of assays in the next couple of weeks. We expect to complete all constructs this quarter.

Conclusions

The results indicate that a chemical made by one organism is likely to be used as food by some other microbe. Bacteria typically utilize the most efficient carbon source available (glucose often being the preferred substrate). More exotic carbon sources are generally subject to catabolite repression and systems for their utilization are activated after preferred carbon sources are exhausted. Besides catabolite repression, sensors are often employed so that the appropriate degradation pathway for a non-preferred carbon source is activated. Our sequencing results have identified organisms rich in transcription-factor based sensors that are integrated with appropriate catabolic gene clusters. With the technology employed approximately 20% of a diverse set of target chemicals yields readily accessible biosensor candidates. With improvements in sequencing technology and declining costs, we suspect that the yield could improve in the near term. For example, the largest step-down in candidates worthy of promotion occurred during the identification of appropriate catabolic gene clusters. Recently, several publications and JCVI testing of the PACBIO next-generation sequencing platform indicated that assembling mostly unfragmented microbial genomes is now possible in a single run. Together with advances in RNAseq and constantly improving bioinformatics as more catabolic pathways are rigorously defined, future large-scale screens will likely improve significantly.

Statement of Work Task List:

- Task 1 (Phase I, Year 1, Months 0-3): Completed (please refer to report HR0011-12-C-2.1)
- Task 2 (Phase I, Year 1, Months 4-9): Completed. Sixty-five isolates have been sequenced.
- Task 3 (Phase I, Year 1, Months 10-12): Completed. Selected microbes have been sequenced, and annotated.
- Task 4, (Phase II, Year 2, months 13-18). Completed.
- Task 5 (Phase II, Year 2, Months 19-24): Initiated and optimized. Construction and testing of the reporter system has been completed and an automated process was produced during sequencing delays. We now expect to be able to process more than the 5-10 original candidates.

Planned Activities for the Next Reporting Period

During the next reporting period we will finish construction of all candidates and finish or nearly finish defining the reporter system for the metabolic engineering phase of the project. While we expect to be able to identify a readily usable sensor from our list of ~40 candidates, it may be necessary to further optimize the sensor before it is prudent to begin metabolic engineering.

Program Financial Status

In Process & Completed Tasks	Planned Expend	Actual Expend (Cumulative to Date)	% Budget Completion	At Completion	Latest Revised Estimate	Remarks
Task 1	$59,251	$59,251	100%	$59,251	$59,251	Completed
Task 2	$69,229	$69,229	100%	$69,229	$69,229	Completed
Task 3	$124,706	$124,706	100%	N/A	$124,706	Completed
Task 4	$255,817	$255,817	100%	N/A	$255,817	Completed
Task 5	$255,817	$218,901	86%	N/A	$255,817	In Progress
Cumulative	$764,820	$727,904	95%	N/A	$764,820	N/A

There is no management reserve or unallocated resources.

Based on the currently authorized work:

- Is current funding sufficient for the current fiscal year? Yes
- What is the next fiscal year funding requirement at current anticipated levels? The budgeted amount for Year 2 of the project is $396,905.25.
- Have you included in the report narrative any explanation of the above data and are they cross-referenced? Not applicable; current funding is sufficient for the current fiscal year.