Final Report **November 27, 2017**
Award Number: DE-SC0012627
Recipient Institution: Boston University
Project Title: A high-throughput pipeline for mapping inter-species interactions and metabolic synergy relevant to next-generation biofuel production
Principal Investigator: Segre, Daniel
Team Members: Christopher Marx (University of Idaho); Trent Northen (LBNL)

1. Executive Summary

The goal of our project was to implement a pipeline for the systematic, computationally-driven study and optimization of microbial interactions and their effect on lignocellulose degradation and biofuel production. We specifically sought to design and construct artificial microbial consortia that could collectively degrade lignocellulose from plant biomass, and produce precursors of energy-rich biofuels. This project fits into the bigger picture goal of helping identify a sustainable strategy for the production of energy-rich biofuels that would satisfy the existing energy constraints and demand of our society. Based on the observation that complex natural microbial communities tend to be metabolically efficient and ecologically robust, we pursued the study of a microbial system in which the desired engineering function is achieved through division of labor across multiple microbial species. Our approach was aimed at bypassing the complexity of natural communities by establishing a rational approach to design small synthetic microbial consortia. Towards this goal, we combined multiple approaches, including computer modeling of ecosystem-level microbial metabolism, mass spectrometry of metabolites, genetic engineering, and experimental evolution.

The microbial production of biofuels from lignocellulose is a complex, multi-step process. Microbial consortia are an ideal approach to consolidated bioprocessing: a community of microorganisms performs a wide variety of functions more efficiently and is more resilient to environmental perturbations than a microbial monoculture. Each organism we chose for this project addresses a specific challenge: lignin degradation (*Pseudomonas putida*); (hemi)cellulose degradation (*Cellulomonas fimi*); lignin degradation product demethoxylation (Methylobacterium spp); generation of biofuel lipid precursors (*Yarrowia lipolytica*). These organisms are genetically tractable, aerobic, and have been used in biotechnological applications. Throughout the project, we have used mass spectrometry to characterize and measure the metabolic inputs and outputs of each of these consortium members, providing valuable information for model refinement, and enabling the establishment of metabolism-mediated interactions. In addition to lignocellulose degradation, we have started addressing the challenge of removing metabolites (e.g. formaldehyde) produced by the demethoxylation of lignin monomers, which can otherwise inhibit microbial growth due to their toxicity. On the computational side, we have implemented genome-scale models for all consortium members, based on KBase reconstructions and literature curation, and we studied small consortia and their properties.

Overall, our project has identified a complex landscape of interactions types and metabolic processes relevant to community-level functions, illustrating the challenges and opportunities of microbial community engineering for the transformation of biomass into bioproducts.

2. Comparison of actual accomplishments with the goals and objectives of the project

The original goals of our project were: (i) To characterize the growth and metabolic input-output properties of the consortia members, using mass spectrometry and other approaches, and to use these data for improving genome scale metabolic network models, and for exploring possible partnerships based on metabolite exchange among strains; (ii) To perform evolution experiments and transcriptional analysis of constituent members and mutualistic pairs, towards optimal growth and maximally efficient conversion. This would inform models, leading to better predictions of metabolic flow and objectives, to be used in subsequent community-level models. (iii) To integrate synthetic ecosystem dynamic modeling and optimization for consolidated bioprocessing. In this phase, experimental measurements of the dynamics and metabolic input/output properties of selected microbial communities would be compared with dynamic community-level flux balance modeling, allowing us to assess and optimize efficiency and stability of the multi-species microbial system as a whole.

As illustrated below, several goals initially set for our projects were achieved, including the metabolic input/output characterization of different community members, the experimental and in silico implementation of prototype consortia, based on the exchange of metabolites resulting from ligno-cellulose degradation, and the development of mathematical and computational tools for the study of cross-feeding in microbial consortia. Initial challenges in identifying appropriate media for co-culture and analytical measurements, and the limited level of gene annotation for some of the organisms chosen, slowed down progress towards testing and optimizing a complete artificial community capable of producing biofuel precursors from biomass. However, lessons learned from our work, including numerous results still in the pipeline for publication, will be highly valuable towards progress in future extensions of this endeavor.

3. Summary of project activities

3.1. Exometabolomic Analysis of Cross-Feeding Metabolites

The Trent lab used exometabolite profiling to follow the resource processing by a microbial co-culture of the bacterial cellulose degrader *Cellulomonas fimi*, and the oleaginous yeast *Yarrowia lipolytica*. We characterized the substrate preferences of the two strains on compounds typically found in lignocellulose hydrolysates. This allowed prediction that specific sugars resulting from hemicellulose polysaccharide degradation by *C. fimi* may serve as a cross-feeding metabolites to *Y. lipolytica* in co-culture. We also showed that products of ionic liquid-treated switchgrass lignocellulose degradation by *C. fimi* were channeled to *Y. lipolytica* in a co-culture. Additionally, we observed metabolites, such as shikimic acid accumulating in the co-culture supernatants, suggesting the potential for producing interesting co-products. Insights gained from characterizing the exometabolite profiles of individual and co-cultures of the two strains can help to refine this interaction, and guide strategies for making this an industrially viable co-culture to produce valuable products from lignocellulose material.

3.2. Genome-scale modeling of community members

Members of the Segrè lab, using input from team collaborators, took advantage of the establishment of a defined minimal medium (modified MPIPES) on which each organism in our consortium can grow, to construct and gap-fill KBase-derived genome-scale models for *Cellulomonas fimi* and *Rhodococcus jostii*. We also modified the metabolite and reaction nomenclature in manually curated and published *Yarrowia lipolytica* (iMK735),

Methylobacterium extorquens (iRP911), and *Pseudomonas putida* (iJN746) genome scale models in order to normalize the naming schemes between them and the KBase reconstructions. This was an essential step towards establishing ecosystem-level stoichiometric models. All models, at every round of improvement were checked for basic consistency and validity, including verifying that no growth is observed in absence of known essential compounds, that the models include appropriate maintenance reactions to simulate non-metabolic energy demands, and that all reactions are mass balanced. We compared simulations and experiments for *Yarrowia lipolytica* in rich medium (yeast nitrogen base, YNB) with a variety of carbon sources (glucose, arabinose, mannose, galactose, and xylose) in order to assess how well *Y. lipolytica* can utilize lignocellulose saccharide degradation products. It was experimentally observed that *Y. lipolytica* will grow optimally in YNB+glucose and YNB+mannose. This was not initially matched by computational simulations, prompting us to perform additional model curation. Further experimental results not easily matched by the model include a diauxic shift that *Y. lipolytica* seems to undergo, consuming the carbon sources in a sequential pattern. Ongoing efforts to address this issue include the use of capacity constraints on the total flux, previously shown to lead to carbon source sequential utilization. Towards the implementation of *in silico* co-culture experiments of consortium members, we identified potential cross-feeding metabolites using dynamic flux balance analysis. We focused initially on *Cellulomonas fimi* and *Yarrowia lipolytica*. We recapitulated some of the metabolites identified using exometabolomic characterization. Specifically, we found that *C. fimi* consumed leucine, valine, and succinic acid in both our simulation and experiments in rich medium (basal salt medium, BSM, with cellulose), and that *Y. lipolytica* produced guanine, alanine, valine, and succinic acid in minimal medium (yeast nitrogen base, YNB, with glucose). These are potential cross-feeding metabolites that may mediate cross-feeding interactions between *C. fimi* and *Y. lipolytica*.

3.3. Testing a 3-species microbial consortium on a defined minimal medium designed to mimic lignocellulose

Efforts in the Marx lab focused largely on extending knowledge gained about individual organisms towards constructing communities that include lignin toxic byproduct utilization. A main consortium pursued consists of *Pseudomonas putida*, *Cellulomonas fimi*, and *Methylobacterium extorquens* PA1; the medium is based on the PIPES-buffered medium from Delaney et al. (2013) and contains vanillic acid, xylose, cellobiose, and sometimes methanol as the carbon substrates. The focus has been on the aerobic degradation of lignocellulose-derived substrates in a controlled small-scale system; principles learned here may later be applied to systems with more complex substrates and with a focus on product generation. Experiments have been primarily in well-aerated batch culture for periods of several days, and measurements made include optical density, viable cell counts of each species (by plating and counting colonies on agar medium), substrate consumption (by targeted HPLC and GC-MS assays), and formaldehyde production/consumption (by Nash assay (Nash 1953)).

Major findings include: (a) Conflicts between species: This medium is designed so that species do not compete for carbon, and we have detected no production of growth-inhibiting factors by any species. However, *P. putida* competes with other species for available iron and methionine. Also, the production of formaldehyde by *P. putida* as a by-product of vanillic acid consumption can transiently inhibit *C. fimi* growth. (b) "Cooperation" between species: *C. fimi* is sensitive to vanillic acid, formaldehyde, and high concentrations of iron: all three compounds inhibit growth of this species. Sequestration of iron by *P. putida*, removal of vanillic acid by growth *P. putida*, and consumption of formaldehyde by both *P. putida* and *M. extorquens*, are all necessary before *C.*

fimi can consume xylose and cellobiose in the medium. Also, a methionine-overproducing strain of *M. extorquens* (see below) may be introduced to supply *C. fimi* with this necessary amino acid. (c) Better characterization of formaldehyde production/consumption dynamics, and formaldehyde tolerance levels of all organisms: As we found earlier, *P. putida* growing on vanillic acid produces formaldehyde which reach levels in the medium that inhibit the growth of *C. fimi*. We have recently confirmed that although *C. fimi* does not grow at these concentrations, it can survive the transient exposure without significant loss of viability, and resumes growth once the formaldehyde is gone. Introduction of *M. extorquens* can speed up formaldehyde consumption, possibly allowing *C. fimi* to resume growth earlier.

3.4. Further development of *M. extorquens* relevant for the consortium

Development of *M. extorquens* strain that produces methionine

Because *C. fimi* requires methionine to be externally supplied, we developed a strain *of M. extorquens* that excretes methionine. We did this by selecting for mutants resistant to methionine (as in Lawrence et al., 1968) and confirmed that the mutant *Methylobacterium* strain can promote *C. fimi* growth both in batch liquid culture and on agar plates. We have yet to quantify the excreted formaldehyde and/or confirm the genetic basis of the phenotype.

Development of *M. extorquens* strain expressing β-ketoadipate pathway: For contrast to the consortium described above (and especially given the difficulties that *P. putida* presents in coculture), we have been pursuing the development of a strain of *M. extorquens* that can fully metabolize lignin-derived aromatic compounds. Earlier, we reported the successful introduction of the *vanABK* gene cluster into *M. extorquens* on an expression plasmid, allowing the strain to convert vanillic acid into formaldehyde (which *M. extorquens* can use as a growth substrate) and protocatechuic acid (PCA) (which it cannot). During this past year, we have introduced into *M. extorquens* an expression plasmid containing several genes from β-ketoadipate pathway for PCA degradation, and demonstrated growth on PCA as a sole carbon substrate. Future plans include the evolution of the β-ketoadipate pathway-expressing *M. extorquens* strain to investigate which genomic modifications may be necessary for improved growth on PCA.

Ecology of *Methylobacterium* growth on lignin-derived aromatics: Our work on methoxy-substituted aromatics and *M. extorquens* described above led to the finding that a major clade of poorly-studied *Methylobacterium* species (such as *M. nodulans*, *M. platani*, and *M. variabile*) possess the genetic capacity to degrade vanillic acid and PCA. We are currently pursuing this finding further, by characterizing the growth of these *Methylobacterium* species on lignin-derived compounds; investigating the phylogeny of the *vanABK* and β-ketoadipate pathway genes in the genus; and mining metagenome data to elucidate the role of *Methylobacterium* in lignin degradation in the environment.

3.5. Mathematical and computational tools for synthetic ecology

As the indispensable role of natural microbial communities in many aspects of life on Earth is uncovered, the bottom-up engineering of synthetic microbial consortia with novel functions is increasingly considered an attractive alternative to engineering single-species systems. As described in detail in a J. Mol. Biol. review published by the Segrè lab, recent work on synthetic microbial communities can help identify open challenges and opportunities in the application of synthetic microbial ecology towards environmental sustainability. The growth of this field is tightly coupled to the development of mathematical approaches, ranging from phenomenological to mechanistic, to decipher the principles that govern the function, dynamics and evolution of microbial ecosystems. Future progress will require more efficient computational algorithms, a better integration of empirical methods and model-driven analysis and improved gene function

annotation. We also suggest that it would be very valuable to generate standardized libraries of well-characterized organisms to be used as building blocks of synthetic communities.

3.6. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities

Mutual dependencies among microorganisms through the exchange of metabolites are ubiquitous in microbial ecosystems. While similar interactions have been engineered in laboratory systems, the evolutionary rise and maintenance of these interactions constitutes an unresolved puzzle. The Black Queen Hypothesis suggests that in communities with essential functions that are costly to focal cells, or "producers", but are unavoidably leaky and partially available to the broader community, metabolic dependencies could arise through adaptive gene loss: in such communities organisms benefit from losing their own capacity to produce a costly metabolite (thus becoming "non-producers"). This could give rise to an obligate dependency of non-producers on producers. However, little is known about the conditions under which these dependencies would be established. A limited number of theoretical studies have explored this question using ecological models. Similarly, other studies have used evolutionary game theory, and concepts from economics to better understand inter-species dependencies in microbial communities. While these approaches have provided valuable phenomenological insight into the general principles of metabolic interdependencies, they often do not take into account the specific details of the organisms, pathways, and molecules involved: behind the biosynthesis, leakiness, and utilization of these metabolites, is a complex network of biochemical reactions, which may significantly vary across different environmental conditions, metabolites, and organisms. A powerful avenue to address this gap is the use of systems biology methods, such as genome-scale network models of metabolism. These models take into account the full metabolic circuitry of a cell and provide quantitative predictions of its growth capacity and metabolic fluxes. In a recently published paper, we proposed a hybrid modeling approach that combines the theoretical insight of evolutionary game theory with the organism-specific-detailed analysis of cell-wide metabolic networks. We demonstrated how this strategy allows one to map the landscape of possible inter-species interactions, for which genome-scale metabolic models provide unique mechanistic insights. Specifically, we use microbial fitness values estimated by metabolic models to infer evolutionarily stable interactions in multi-species microbial "games". We first validated our approach using a well-characterized yeast cheater-cooperator system. We next performed over 80,000 in silico experiments to infer how metabolic interdependencies mediated by amino acid leakage in *Escherichia coli* vary across 189 amino acid pairs. While most pairs display shared patterns of inter-species interactions, multiple deviations were caused by pleiotropy and epistasis in metabolism. Furthermore, simulated invasion experiments revealed possible paths to obligate cross-feeding. This study provided genomically driven insight into the rise of ecological interactions, with implications for microbiome research and synthetic ecology. We expect that similar effects, to be further explored, will be relevant for establishing stable inter-species interactions in engineered consortia for biofuel production.

4. Products developed under the award

4.1. Research Articles

Joshua E. Goldford*, Nanxi Lu*, Djordje Bajic, Sylvie Estrela, Mikhail Tikhonov, Alicia Sanchez-Gorostiaga, Daniel Segrè, Pankaj Mehta, Alvaro Sanchez: *Emergent Simplicity in Microbial Community Assembly*, BioRxiv (2017), https://doi.org/10.1101/205831.

Lubbe A, Bowen BP, Northen T., *Exometabolomic Analysis of Cross-Feeding Metabolites*. Metabolites. 2017 Oct 4;7(4). pii: E50. doi: 10.3390/metabo7040050. PMID: 28976938

Zomorrodi AR, Segrè D., *Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities*. Nature Communications 2017 Nov 16;8(1):1563. doi: 10.1038/s41467-017-01407-5. PMID: 29146901

Reznik E, Christodoulou D, Goldford JE, Briars E, Sauer U, Segrè D, Noor E., *Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity*. Cell Reports, 2017 Sep 12;20(11):2666-2677. doi: 10.1016/j.celrep.2017.08.066. PMID: 28903046

Goldford JE, Hartman H, Smith TF, Segrè D., *Remnants of an Ancient Metabolism without Phosphate*. Cell. 2017 Mar 9;168(6):1126-1134.e9. doi: 10.1016/j.cell.2017.02.001. Epub 2017 Mar 2. PMID: 28262353

Zhao Q, Stettner AI, Reznik E, Paschalidis ICh, Segrè D., *Mapping the landscape of metabolic goals of a cell*. Genome Biology, 2016 May 23;17(1):109. doi: 10.1186/s13059-016-0968-2. PMID: 27215445

Brian R. Granger*, Yi-Chien Chang*, Yan Wang, Charles DeLisi, Daniel Segrè, Zhenjun Hu: *Visualization of Metabolic Interaction Networks in Microbial Communities Using VisANT 5.0*, PLOS Computational Biology (2016) 12(4): e1004875. doi:10.1371/journal.pcbi.1004875.

Lorenzo Castelli, Raffaele Pesenti and Daniel Segrè: *The cell as a decision-making unit*, IEEE Life Sciences Letters (2016), DOI: 10.1109/LLS.2016.2644648.

Qi Zhao, Daniel Segrè and Ioannis Ch. Paschalidis: *Optimal allocation of metabolic functions among organisms in a microbial ecosystem* (2016), IEEE 55th Conference on Decision and Control (CDC), DOI: 10.1109/CDC.2016.7799357

4.2. Review Articles

Ali R. Zomorrodi and Daniel Segrè: *Synthetic ecology of microbes: mathematical models and applications*, Journal of Molecular Biology (2016) Vol. 428, Issue 5, Part B, Pages 837–861.

4.3. Posters

Meghan Thommes, Andrea Lubbe, Jessica Lee, Arion Stettner, Ilija Dukovski, Alyssa Baugh, Nicholas Shevalier, Joshua Wirtz, Sergey Stolyar, Christopher Marx, Trent Northen, and Daniel Segrè. Designing a Microbial Community for Production of Biofuel from Lignocellulose. 2016 Genomic Sciences Program Annual PI Meeting. Tysons, VA. March 6-9, 2016. 120-121.

Jessica A. Lee, Nicholas Shevalier, and Christopher J. Marx, July 2015: Gordon Research Conference on Microbial Population Biology (Andover, NH) Optimizing Methylobacterium extorquens for the role of detoxifier in a microbial biofuel consortium.

Nicholas Shevalier, Jessica A. Lee, and Christopher J. Marx, November 2015: University of Idaho College of Science Student Research Exposition (Moscow, ID)
Metabolic capabilities to degrade lignin-derived compounds of leaf and root Methylobacterium strains.

Alyssa Baugh, Tomislav Ticak, Jessica A. Lee, and Christopher J. Marx: May 2016: Inland Northwest Genomics Research Symposium (Moscow, ID), Optimizing vanillic acid demethoxylation in Methylobacterium extorquens via genetic engineering

Andrea Lubbe, Meghan Thommes, Arion Stettner, Jessica Lee, Christopher Marx, Trent Northen and Daniel Segrè: Integrating exometabolomics and metabolic modeling of microbial interactions towards consolidated bioprocessing for next generation biofuel production, 11th International Conference of the Metabolomics Society, June 29th 2015 to July 2nd 2015, Burlingame, CA.

Jessica A. Lee, Alyssa Baugh, Tomislav Ticak, Nicholas Shevalier, and Christopher J. Marx, August 2016: Gordon Research Conference on the Molecular Basis of Microbial One-Carbon Metabolism (Waterville Valley, NH). Formaldehyde is in the air: cross-feeding of a toxic metabolite in a lignocellulose-degrading community.

Meghan Thommes, Daniel Segrè. Computer-driven Design and Experimental Testing of a Synthetic Microbial Community. BMES. Minneapolis, MN. October 5-8, 2016.

4.4. Book Chapters

Lubbe A., Northen T. (2016) Exometabolomics for Linking Soil Carbon Dynamics to Microbial Communities. In: Beale D., Kouremenos K., Palombo E. (eds) Microbial Metabolomics. Springer, Cham, pp 119-145.

4.5. Computational tools (see next section for additional details)

- Improved version of COMETS: http://comets.bu.edu/
- Matlab Toolbox for COMETS: https://github.com/segrelab/comets-toolbox
- Inverse FBA (InvFBA): https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0968-2
- Optimization tool to identify Nash Equilibria, for application to Microbial Consortia (NashEq Finder): http://bit.ly/2qexiC6
- VisANT 5.0, Network visualization software, with newly added capacity to visualize metabolite-mediated interactions in microbial ecosystems: http://visant.bu.edu/

5. Information on theory and modeling approaches

5.1. Improvements and new tools for COMETS (Computation of Microbial Ecosystems in Time and Space)

One of the aims of our software development efforts has been to make COMETS available to a large community of users and potential developers. With that in mind we have created a web site dedicated to COMETS at http://comets.bu.edu. The users can download the software and follow the detailed and user-friendly installation instructions. A new self-extracting automatic installer

is provided for Windows users. This will enable users of wide range of computational skills and expertise to be able to approach COMETS for their specific needs. We have put together a detailed user's manual including a tutorial on running COMETS. The COMETS package release includes simple examples that the users can run with ease and later build upon according to their needs.

5.2. A Matlab Toolbox for COMETS

COMETS is written in Java. It has a graphic user interface, that allows users to design simple in silico experiments. However, complex in silico experiments (e.g. with multiple species, time-dependent boundary conditions and structured environments) have required until now that users generate input data manually or with ad hoc tools. We have recently completed the development of the first fully working version of the COMETS Matlab Toolbox (available at https://github.com/segrelab/comets-toolbox). This toolbox is a suite of Matlab scripts and functions designed to simplify the processes involved in generating COMETS layout files, executing simulations, and analyzing COMETS output. Metabolic models are handled via the integrated COBRA Toolbox (https://opencobra.github.io), a popular library for this task, and command syntax follows COBRA-like patterns to make the system easier for users to learn. Using the Toolbox, a user can quickly create a Layout data structure describing the cell culture dish in a simulated experiment. Functions are available to apply media components uniformly across the dish or at specific positions, and to designate the rate at which media components may be refreshed. Biomass representing each metabolic model can be added to the culture in a set of predefined or customized patterns. These approaches offer a significant improvement in speed and readability when compared to previous workflows without the Toolbox which often required the user to manually designate the position of components as series of coordinate positions within a text file. The COMETS Toolbox also allows execution of COMETS simulations and parsing COMETS's output within Matlab scripts. This facilitates experiments which compare results between related layouts by programmatically creating and executing a large number of simulations.

5.3. Genome-scale game theory approach

The paper by Zomorrodi and Segrè (Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities), published in Nature Communications in 2017, includes a detailed description of a computational approach to (i) compute a payoff matrix of fitness values for different interacting species, based on genome scale models, and (ii) subsequently estimate Nash Equilibria for a community based on this payoff matrix. As part of this work, we implemented a python software, called "NashEq Finder", that identifies all pure strategy Nash equilibria of a game with any number of players and strategies in one shot. The script and details on how to run it are available as supplementary material of the paper, at http://bit.ly/2qexiC6.