Engineering Modular Synthetic Microbial Consortia for Sustainable Bioproduction From CO_2

David Carruthers

A dissertation submitted in partial fulllment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2020

Doctoral Committee:

Professor Xiaoxia Nina Lin, Chair
Professor Daniel C. Ducat
Professor Gregory A. Keoleian
Professor Henry Y. Wang
Professor Fei Wen

David N. Carruthers

dcarruth@umich.edu

ORCID iD: 0000-0002-8275-2278

ACKNOWLEDGMENTS

Firstly, I would like to sincerely thank my adviser, Professor Nina Lin, for supporting me and providing the opportunity to conduct high level academic research. Her expertise and guidance as an advisor enabled me to learn, hone my research skills, and ultimately have a very successful PhD. I am also grateful to my committee for insightful feedback that have enriched this work. Specifically, Dr. Daniel Ducat not only provided the *S. elongatus cscB* SPS strain (Chapter 2), but shared excellent advice for optimizing culture conditions.

I'd also like to thank past members of the Lin laboratory during my tenure at Michigan, including Dr. Tatyana Saleski, Dr. Scott Scholz, Dr. Corine Jackman, Dr. Steven Wang. Dr. Chun Wan, Dr. Jeremy Minty, Prof Chang Kyu Byun, who collectively provided a conducive and collaborative research environment. I am particularly grateful for Corine, Scott, and Tatyana who constantly provided encouragement and were excellent role models for success in the laboratory. I'd also like to thank current members of the laboratory, Adam Krieger, James Tan, and Erica Gardner, who have always been receptive to either listen to me complain about unsuccessful experiments or provide an ideation backboard for new directions.

I'd also like to thank my many academic collaborators outside of the University of Michigan including, Dr. Brett Barney provided the principle *A. vinelandii* strain for Chapter 2, AZBB163, as well as direct correspondence for culturing and cryopreservation. I am also grateful for Prof. James Liao (UCLA), who provided isobutanol production strains and plasmids (Chapter 3), and Prof. Maciek Antoniewicz, who provided the end-point ALE strains of *E. coli* K-12 (Chapter 3). My undergraduate lab, namely Prof. Steven Mayfield

and Dr. Ryan Georgianna, also inspired me to pursue graduate research and, ultimately, instilled a confidence that I could earn this degree.

Many collaborations throughout this PhD have led to successful research and academically enriching relationships. The LCA work (Chapter 4), for example, was a result of a collaboration between chemical engineers and ecologists. Collaborators included Prof. Bradley Cardinale, who imparted both a dedication to excellence and an ecological perspective to community engineering, and Dr. Casey Godwin and Dr. Dave Hietala who made major contributions to our algal biofuel LCA.

I'm also grateful for members of the Burns Lab, namely Prof. Mark Burns, Dr. Zach Pritchard, Dr. Martin de Beer, Dr. Sarah Mena, Alyse Krausz, and Brian Johnson for being excellent lab partners by sharing equipment from time to time. I'd also like to recognize members of the Chemical Engineering Department administration, especially Kelly Raichovich, Susan Hamlin, Barbara Perry, Michael Africa, and Sandy Swisher, who have all been willing to aid on all things technical throughout my time in the chemical engineering program.

And, of course, I'd like to thank my Ann Arbor family (in no particular order): Shreyas Vangala, Benjamin Kunstman, Dr. Benjamin Lee, Jefferson Sanchez, Michael King and his family, Avi Bregman, Kari Paine, Andrew Harrison, Dr. Phil Spinosa, Dr. Joe Quinn, Dr. Thom Logan, Kathy Kosinski, Dr. Martin de Beer, No Thai, the Blue Lep Trivia team, the MCycling Club team, and the many other friends with whom I have shared unforgettable memories over the last six years. Specifically, I'd like to thank Martin for all the adventures, the good times, and his cleverness in the laboratory; Kathy for literally anything dog, library, or meme related; Phil for teaching me how to live-laugh-love; Ben for fighting for the good in the world; Shreyas for being an absolutely outstanding human; and Jeff for always pushing himself and truly living life to the fullest.

Most of all, I'd like to thank my wonderful girlfriend Genesis Lopez for her love, encouragement, patience, unbeatable cooking, and keeping me sane throughout this journey

and especially throughout this quarantine.

Lastly, I'd like to thank my actual family for encouraging and supporting me to pursue my research interests at Michigan despite being thousands of miles from home. Those video calls, Football games, and Michigan trips provided much needed perspective and respite from what was often an all-consuming, stressful endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ACRONYMS AND ABBREVIATIONS

ACP	Aqueous co-product
afdw	Ash free dry weight
AHL	N-acyl homoserine lactones
AHM	Algal HTL Module
amp	Ampicillin
ALE	Adaptive laboratory evolution
APD	Algal process description
BNF	Biological nitrogen fixation
bp	Base pair
CCS	Carbon capture and sequestration
cDNA	Complementary DNA
CFP	Cyan fluorescent protein
CIChE	Chemically inducible chromosomal evolution
CIDI	Combustion ignited direct ignition
CHG	Catalytic hydrothermal gasification
Chl-a	Chlorophyll-a
cm	Chloramphenicol
CO_2eq	Carbon dioxide equivalent
CRIM	Conditional-replication, integration, and modular
CRISPR	Clustered regularly interspaced short palindromic repeats
csc	Chromosomal sucrose catabolism
DHBA	Dihydroxybenzoic acid
DI	Deionized
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
EROI	Energy return on energy invested
FCM	Flow cytometry
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ISA	Ionic strength adjuster
ISE	Ion selective electrode
gDNA	Genomic DNA
GFP	Green fluorescence protein
GHG	Greenhouse gas
GWP	Global warming potential
GREET	Greenhouse gases, Regulated Emissions, and Energy in Transportation

HPLC	High-performance liquid chromatography
HTL	Hydrothermal liquefaction
ISO	International Organization for Standardization
kan	Kanamycin
LB	Lysogeny broth
LCA	Life cycle assessment
LCI	Life cycle inventory
LED	Light emitting diode
M9IPG	M9 isobutanol production medium with glucose
M9G20	M9 media with 20 g/L glucose
M9S20	M9 media with 20 g/L sucrose
MBTU	Million british thermal units
ME	Mosaic end
mNG	mNeonGreen
NREU	Nonrenewable energy usage
OD	Optical density
ODE	Ordinary differential equation
P1	Classical bacteriophage P1
PAR	Photosynthetic active radiation
PCA	Protocatechuic acid
PCR	Polymerase chain reaction
PGA	poly-β-1,6-N-acetyl-D-glucosamine
PGPR	Plant growth promoting rhizobacteria
OHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutyrate
PLA	Polylactic acid
PTS	Phosphotransferase
qPCR	Quantitative polymerase chain reaction
QS	Quorum sensing
RBS	Ribosome binding site
Registry	Registry of Standard Biological parts
SnoCAP	Syntrophic co-culture amplification of production
SOC	Super optimal broth with catabolite repression
TE	Tris-EDTA buffer
WI	Water intensity
YFP	Yellow fluorescent protein
YPD	Yeast extract, peptone, dextrose media

ABSTRACT

Engineering of synthetic microbial consortia has emerged as a new and powerful biotechnology platform with enormous potential for the production of biobased commodity chemicals. In this dissertation, I have designed, constructed, and optimized a tripartite system in which three microbes of differentiated specializations can convert sunlight, carbon dioxide, and atmospheric nitrogen into desired molecules or materials. Specifically, *Synechococcus elongatus*, a photosynthetic cyanobacterium that exports sucrose, and *Azotobacter vinelandii*, a nitrogen-fixing bacterium that secretes ammonia, form a symbiotic foundation hypothesized to support a third producer specialist. The tripartite consortia were implemented using a novel experimental set-up for continuous culture and extensive optimization was carried out with insights and guidance from computational modeling of the system dynamics. As a clear and strong proof of concept, I demonstrated various realizations of this tripartite platform, employing producer specialist strains ranging from model microorganism *Escherichia coli* to widely used industrial chassis such as *Corynebacterium glutamicum* and *Bacillus subtilis*. This versatile and modular technology platform offers potential for bioproduction without environmentally or monetarily expensive nutrient inputs thereby a pathway towards sustainable manufacturing of a wide range of bio-products.

As an important component of the effort of engineering the tripartite system described above, I also carried out genetic modifications of *E. coli* K-12, the most widely used microbial chassis in synthetic biology, to enable efficient utilization of sucrose. A multigene *csc* operon encoding non-PTS sucrose catabolism was randomly transposed into *E. coli* K-12 using Tn5 transposase. Isolates from the transposon library yielded a range of growth rates

on sucrose, including some that were comparable to that of *E. coli* K-12 on glucose. Narrowness of the growth rate distributions, improved gene expression conferring faster growth compared to that of plasmids, and enhanced growth rate upon transduction into strains that underwent adaptive laboratory evolution indicate that efficient *csc* expression is attainable and not limiting to cellular growth. Transduction of a *csc* fast-growth locus into an isobutanol production strain also yielded high titer with significant sustainability benefits. This work demonstrated that random integration is a viable and effective strategy for optimizing heterologous expression within the context of cellular metabolism for certain desirable phenotypes.

In the last part of my thesis, through life cycle assessment, I investigated multi-species algal polycultures, which are different yet related CO_2-fixing microbial communities. Experimental studies have previously shown that algal polycultures can be designed to enhance biomass production, stability, and nutrient recycling compared to monocultures. However, it remains unclear whether these impacts of biodiversity make polycultures more sustainable than monocultures. I have conducted a comparative life cycle assessment which showed that when algae were grown in outdoor experimental ponds, certain bicultures improved the energy return on investment and greenhouse gas emissions substantially, compared to the best monoculture. Bicultures outperformed monocultures by performing multiple functions simultaneously (*e.g.*, improved stability, nutrient efficiency, biocrude characteristics), which outweighed the higher productivity attainable by a monoculture. These results demonstrated that algal polycultures with optimized multi-functionality lead to enhanced life cycle metrics, highlighting the significant potential of ecological engineering for enabling future environmentally sustainable algal bio-refineries.

Collectively, this dissertation demonstrates how CO_2-fixing microbial communities may be engineered to enhance sustainability metrics compared to monocultures. By successfully engineering more sustainable bioproduction platforms, we move closer to a society with lower dependence on petrochemicals.

Chapter 1: Background and Motivation

1.1 A brief overview of microbial consortia

1.1.1 Natural microbial consortia

Natural microbial consortia are ubiquitous in Earth's microbiomes and fundamental to the field of ecology. Engineering microbial communities requires a fundamental understanding of how organisms have evolved, interact, and function within the natural environment. Environmental niches are characterized by a subset of abiotic conditions that include, among others, temperature, pH, resource availability, oxygenation, and irradiance. Abiotic conditions are complemented by biotic factors attributed to constitutive members of the community. Microbial communities propagate, adapt, and coevolve in proximity for many generations within an environmental niche to produce complex interaction networks. The networks themselves consist of an assemblage of pairwise relationships between microbial species that ultimately yield ecosystem function and survival. Communities can occupy familiar environments like marine, soil, and the plant rhizosphere as well as extreme environments like hydrothermal vents and acid mine drainage. Networks are inherently complex and often form syntrophic systems in which organisms are codependent for survival and require metabolite exchange due to extreme niche partitioning [1].

The rhizosphere is a canonical example of a natural microbial system. It spans trophic levels and includes both the root epidermal layers and the surrounding microbial community in the soil. The relationship between microbes and the plant itself is an example of symbiosis. The rhizosphere comprises hundreds or thousands of different microbial species

with collective cell populations often exceeding $10^{10}-10^{12}$ cells per gram of soil [2,3]. Species like diazotrophs and mycorrhizal fungi impart substantial benefit to the plant. Diazotrophs specifically are capable of biological nitrogen fixation (BNF) of atmospheric nitrogen (N_2 gas) for plant usage as NH_4^+ or NO_3^- (Fig. 1.1).

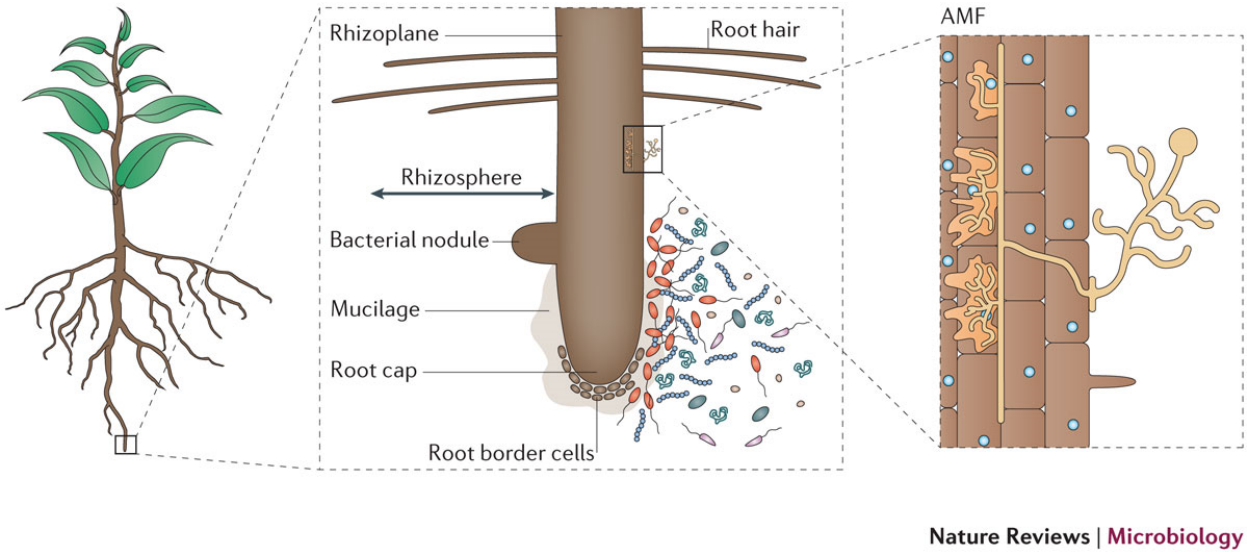


Figure 1.1: A cartoon representation of the rhizosphere and, more closely, the root microbiome. It consists of a mixed population of bacteria and fungi. The inset shows an arbuscular mycorrhiza fungus that penetrates several layers of the plant root surface. Reprinted by permission from Springer Nature: Springer Nature Reviews Microbiology (Ref: [4]), Copyright (2013).

The rhizosphere includes a subclassification of organisms called plant growth promoting rhizobacteria (PGPR). PGPR are multifunctional and while some species like *Azotobacter* sp. are diazotrophic, others enhance plant growth through siderophore production, phosphate solubilization, production of growth factors, and enhanced general stability of a community [5]. In return, plant roots can release as much as 10-40% of all photosynthetically fixed carbon as usable substrate for soil bacteria [6]. Without diazotrophic organisms or supplemented nitrogen, plant growth is impossible. Furthermore, photosynthesis and BNF are energetically intensive and catalytically complex given that oxygen irreversibly damages nitrogenase, the principle enzyme in BNF. While some organisms (*e.g.*, *Anabaena* sp., purple nonsulfur bacteria) are capable of both functions, they do so through energy intensive cell differentiation or complex intracellular regulation. The partitioning of carbon and nitrogen

fixation between different organisms capitalizes on individual resource availability in the rhizosphere. Division of labor in the rhizosphere enables far higher community production as well as community stability and functional resilience [7].

While microbial community function is often well described, the nature of the interactions between constitutive members is less obvious. Interaction networks are based on fundamental ecological principles help to mathematically discern how community function arises. Plants and PGPR, for example, maintain a mutualistic symbiosis as each organism provides a net benefit to the other. Within the classification of PGPR, however, two species consuming a mutual substrate are in direct competition because nutrient consumption by one organism negatively affects its availability to the other. Interactions also include predation, commensalism, amensalism, and neutralism as described in Table 1.1.

Table 1.1: Pairwise microbial interactions. Microbes have variable interactions that can positively, negatively, or neutrally affect one another.

Type of interaction	**Signs**
Mutualism	+/+
Competition	-/-
Predation	+/-
Commensalism	+/0
Neutralism	0/0
Amensalism	0/-

Recent advances in genomic sequencing can help elucidate constitutive members of a microbial community. The most common method is sequencing of the 16S ribosomal RNA gene. As a highly conserved region, 16S sequences reliably enable taxonomic cataloging of organisms by genus. Whole genome sequencing and cross referencing of genes within large genomic databases has also facilitated predictions of how an organism might interact within a network. This is especially important in the rhizosphere due to recalcitrant or nonculturable organisms that compose as much as 80-90% of the community [8]. Many mycorrhizal fungi,

for example, are obligate symbionts and cannot be grown independent of plant hosts. By collectively describing and mapping these interactions as well as by cross referencing genetic sequences with known gene function, ecologists can elucidate "hubs" or critical organisms for community growth or stability. Sequencing can also enable postulation of metabolite release or uptake. An isolated network of hubs might include several dominant interactions (Fig. 1.2A) within the broader context of community interactions (Fig. 1.2B).

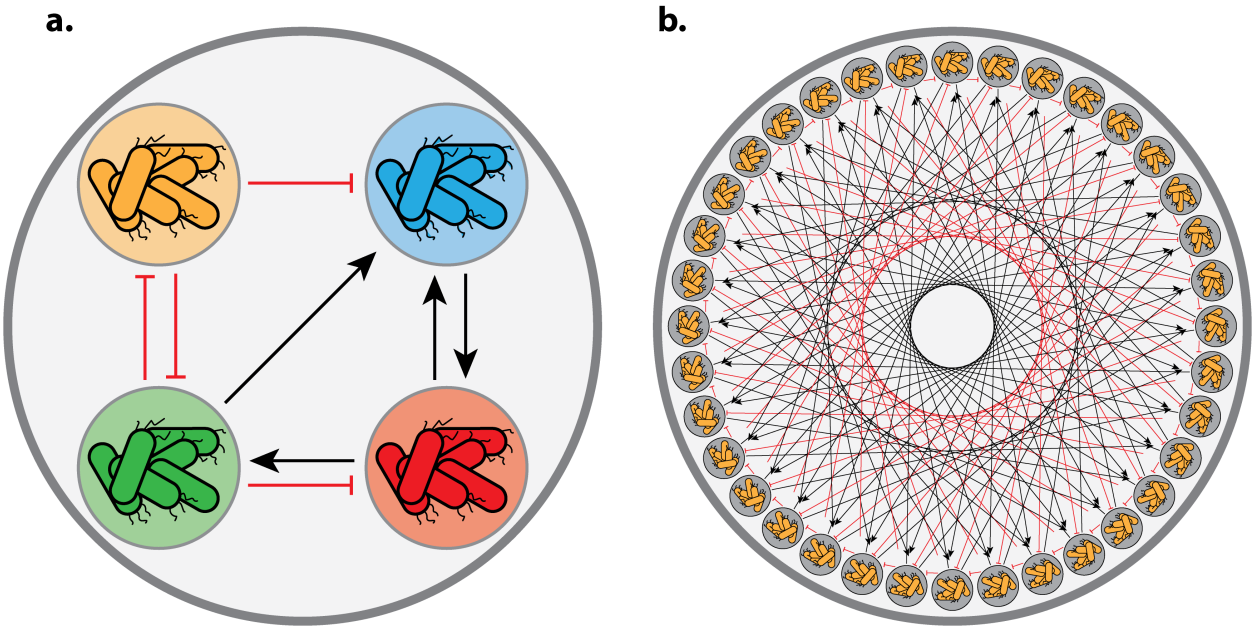


Figure 1.2: Diagram of hypothetical microbial interaction networks. (a) A simple microbial network with mutualism (blue/red), competition (yellow/green), predation (green/red), commensalism (blue/green), neutralism (yellow/red), and amensalism (yellow/blue). (b) Depiction of interactions between a hypothetical 36-member network. Interactions were tessellated for simplicity.

Microbial communities have enormous potential for biotechnological innovation through remediation of wastewater, sludge, sediment, and soil or generalized bioproduction [9]. Continuing with the example of the rhizosphere, a thorough understanding of PGPR interactions enables tailoring of microbial communities as "biofertilizers" or "rhizoremediators" that can promote sustainable agriculture or remove contaminants from soil, respectively [5]. While evolution has provided robust microbial communities, the explosion of synthetic biology advances over the last several decades have significantly lengthened the reach of genetic modification. By capitalizing upon natural microbial functions and by applying our collec-

tive knowledge of genetic circuits wrought by recent advances in molecular biology, we can engineer novel and complex microbial consortia that address pressing global issues.

1.1.2 Synthetic biology

Synthetic biology applies engineering principles to rewire biological systems and has been well complemented by advances in genotyping and recombinant DNA technologies. The field is perhaps best exemplified by the BioBrick system. BioBrick is a Registry of Standard Biological parts (Registry) that are compatible and interchangeable. The Registry contains over 20,000 parts, namely promoters, terminators, ribosome binding sites (RBS), protein coding sequences (reporters, regulators, selection markers, etc.), and vectors, usually plasmids. Each part may also vary in strength. Strong promoters, for example, promote RNA polymerase binding for transcription. RBS strength in bacteria can be directly measured by similarity to the Shine-Dalgarno sequence (AGGAGG) [10]. Components may be added or rearranged using any of a myriad of biological strategies. The BioBrick system is designed principally for standard cloning with endonuclease digestion and subsequent "blunt" or "sticky" end ligation. Though not supported by the BioBrick system, other cloning technologies include golden gate or Gibson assembly, which uses lengthy homology and backbone amplification for scarless integration into a specific site on a plasmid. BioBrick vectors have five key characteristics: 1) a cloning site for heterologous gene expression, 2) a selection marker, 3) copy number, 4) transcriptional terminators and translational stop codons, and 5) verified primer annealing sites for screening [11]. Synthetic biologists can easily build circuits of varying function and production (*e.g.*, fluorescence level) using the "bricks" from the Registry.

The BioBrick framework serves a basis for investigation of episomal genetic circuits. The modular approach has been applied to study expression cascades, or genetic expression linked by inducing or repressing chemicals that can be reproducibly created and mathematically modeled [12]. One such study developed a "repressilator", a synthetic biological

oscillator in which three cyclic repressors inhibit each other to spontaneously yield oscillations in expression of a green fluorescence protein (GFP) reporter [13]. The repressilator was also used in a quorum sensing microbial system. Quorum sensing (QS) is a unique biological phenomenon in which an autoinducer is secreted and, after reaching a threshold, genetic expression is regulated and induced by that chemical at the population level. The specific signaling molecule differs between cell populations, though the most well studied class are N-acyl homoserine lactones (AHL) that bind directly to transcription factors as in the luxI/luxR system [14]. *LuxI* produces the autoinducer lactone while *luxR* encodes a receptor and transcription factor. When the autoinducer binds to luxR the *lux* operon is induced and bioluminescent luciferase is produced [14]. By pairing the repressilator with the QS system, the authors enhanced synchronization or predictable generational oscillatory behavior of the system at the population level [15].

Another work constructed a synthetic predator-prey relationship using two analogous QS systems (luxI/luxR and lasI/lasR) in two strains of *E. coli* as indicated in Fig. 1.3. In this system *ccdB* and *ccdA* genes were expressed in two different organizations under luxI and lasI inducible promoters, respectively [16]. Expression of *ccdB* produces a toxic protein (ccdB) that inhibits DNA gyrase, resulting in strand breakage [16]. Conversely, *ccdA* produces an antitoxin (ccdA) that binds to and inactivates ccdB protein [16]. The genetic architecture enabled predictable tuning of predator-prey dynamics based on relative component strength (*i.e.*, strong/weak promoter, strong/weak RBS, etc.) using prototypical nonlinear Lotka-Volterra dynamics with high precision. Collectively, these works established that microbial circuits can be extrapolated as simplified models of complex biological phenomena, like to circadian rhythm in the repressilator system or to maintenance and generation of biodiversity in the predator-prey *E. coli* system. Furthermore, such microbial communities can be readily engineered to investigate natural phenomena.

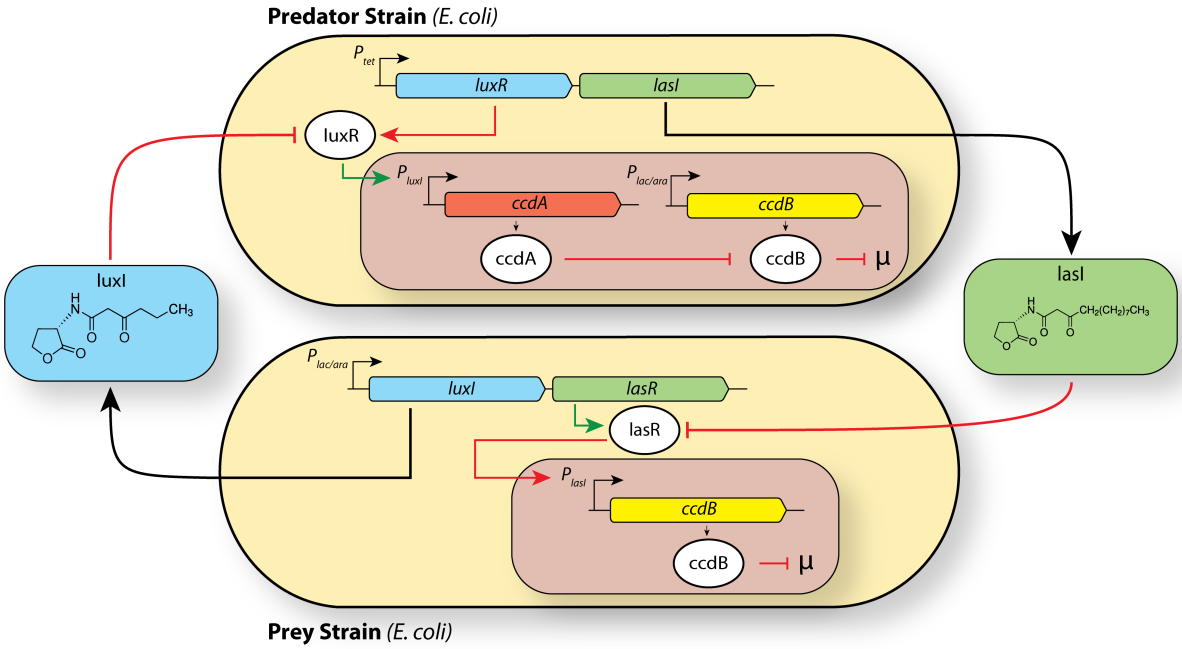


Figure 1.3: Depiction of the microbial predator-prey model adapted from [16]. The predator and prey are related by AHL cross feeding. The predator, for example expresses *lasI*. *LasI* produces an AHL that binds to lasR in the prey and induces cell death by expression of *ccdB*. On the other hand, the prey expresses *luxI*, which binds to luxR and induces *ccdA* expression, which eliminates ccdB protein and allows growth. The result is that high cell concentrations of predator yield high concentrations of lasI and ultimately kill the prey. Low cell concentrations of prey yield low concentrations of luxI such that ccdA is not expressed and the predator dies. Low predator concentration yields low lasI, lending to prey regrowth. Arrows colored red for cell death/inhibition (*i.e.*, lasI binds to lasR, enabling *ccdB* expression and cell death), or green for cell growth. The system is identical in principle to the ecological

1.1.3 Synthetic microbial consortia

More recently, metabolic engineering has enabled manipulation of strains for biochemical production by capitalizing on well-defined molecular circuits and have led to the development of vast libraries of biochemical production schemes. This process began with the fermentation of *Clostridium acteobutylicum* for production of acetone as a byproduct of acetone-n-butanol-ethanol fermentation and has since matured into chemical microbial cell factories [17, 18]. Traditionally, the focus of synthetic biology has been to engineer specific microbes or "superbugs" that can complete complex heterologous processes in monoculture. Major breakthroughs have included the introduction of recombinant DNA for heterologous

production of human insulin and artemisinin, an antimalarial drug, in *E. coli* [19].

However, expression or manipulation of complex pathways often yields reduced growth rates due to competitive impairments that adversely affect industrial scale-up. Single organisms, for example, often accumulate high levels of byproducts like acetate under industrial biofermentation conditions, which lowers efficiency and yield [20]. Such difficulties stem from intracellular competition between enzymes, robust and tightly wired metabolic pathways, and bottlenecks that are often highly conserved among many microbial species [21]. These drawbacks are characteristic of engineered superbugs. Metabolic flux and flux balance analyses have enabled in silico optimizations that approach the theoretical maxima of a pathway. Nonetheless, the fundamental challenge of improving overall pathway or process efficiency by reducing negative functional trade-offs remains.

A novel approach for overcoming these drawbacks is the development of synthetic microbial consortia, which function coherently to produce a chemical target. Microorganisms naturally exist in communities and often form symbioses as showcased in the human gut microbiome, where many specialized bacteria facilitate digestion, and as described in the previous section between BNF rhizobacteria, mycorrhizal fungi, and carbohydrate secreting plants. Natural microbial symbioses have flourished for millennia and extend between bacterial, algal, and fungal species [22]. Collectively, these symbiotic consortia have higher efficiency and increased resistance to both environmental variance and invasion than any single organism within the community. Synthetic microbial consortia capitalize upon differences in microbial metabolism as well as native metabolite production to partition a chemical pathway. Consortia can displace costs and associated environmental impacts of biochemical production by circumventing the constraints intrinsic to optimizing a superbug. It is accomplished by dividing strains by function to complete complex tasks simultaneously (*i.e.*, division of labor), often increasing overall productivity [23]. The use of model organisms within a consortia also enables an unparalleled level of control by allowing for spatial separation (*i.e.*, membrane, emulsion, or hydrogel), temporal separation (conditioned media

or inoculation time and ratio), genetic expression under inducible promoters, and variable environmental conditions (temperature, pH, and agitation).

A simple synthetic microbial consortium is exemplified by the cross-feeding of tryptophan and tyrosine between two auxotrophic strains of *E. coli* under arabinose and propionate inducible promoters, respectively [24]. The modified *E. coli* strains are codependent for an essential amino acid in an engineered symbiosis. Researchers also demonstrated that the consortium may be tuned by changing strain inoculation ratio and by adjusting chemical concentration of promoter to affect pathway expression level. Individual growth rates could be mathematically modeled using Monod kinetics as functions of inducing chemical and initial cell composition [24]. Krieger et al. also demonstrated control of fluorescently labeled *B. subtilis* (T_{opt} = 42 °C) and *E. coli* (T_{opt} = 37 °C) by temperature tuning, which produced different strain growth rates and thereby different community compositions over sequential passages.

Novel methodologies for high throughput screening of synthetic microbial communities have stemmed from recent advances in microfluidic technologies. Microfluidics has demonstrated effective screening of high production algal polycultures and human microbiome co-cultures [25, 26]. Beneficial phenotypes have also been screened through a novel platform called Syntrophic co-culture amplification of production phenotype (SnoCAP) [27]. SnoCAP uses a secretor strain that produces a specific molecule of interest and a sensor strain that fluoresces in the presence of the molecule. The result is a microfluidic platform in which a library of strains may be quickly screened in co-culture for high fluoresce and characterized at throughput rates 10^3 higher than in microwell plates. SnoCAP has now made rapid screening of enormous cell libraries possible.

Microbial communities are extremely versatile. Decades of research have produced high resolution models of microbial metabolism and compatible biological component libraries for circuit construction. Biologists can now capitalize upon the natural functional or environmental predispositions of select microbes to build novel communities, screen quickly for target

phenotypes, and tune community composition or metabolite exchange in the laboratory to construct synthetic microbial consortia.

1.1.4 Mathematical modeling of microbial growth and production

Mathematical modeling of microbial interactions is often imperative for engineering synthetic consortia or simply guiding genetic modification within a single organism. Fortunately, interaction networks can often be well-described by a system of differential equations that account for nutrient uptake and biomass accumulation with good corroboration to empirical data. These equations are generally constrained by culture conditions such as density or substrate dependence. The most common mathematical model for microbial growth follows Verhulst kinetics, which are density dependent kinetics that follow a standard logistic equation and rely on two microbial constants with a single initial condition:

$$\frac{dX}{dt} = rX\left(1 - \frac{N}{K}\right) \tag{1}$$

Here, K represents the carrying capacity of the microbe within the context of the system while r represents the specific growth rate. The empirical solution may be solved using an initial biomass condition.

$$X(t) = \frac{K}{(1 - (\frac{K}{X_0} - 1)e^{-rt})} \tag{2}$$

$$X_0 = X(0) \tag{3}$$

While the logistic equation is extremely versatile and can be solved empirically, it suffers from oversimplification. Carrying capacity is derived purely from lumped experimental conditions dependent on the experimental context and cannot be extrapolated between conditions. Nonetheless, Equation 1 often well describes systems with slow growing constituents like algae. More robust models tend to follow Michaelis-Menten kinetics in which a substrate

gradually forms a product as a function of the substrate concentration. Michaelis-Menten were originally used to describe conversion of sucrose to glucose and fructose by enzymatic hydrolysis using invertase [28].

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_s + [S]} \tag{4}$$

Here, the rate of product formation is a function of the substrate and dictated by a dissociation constant, K_s, and the maximum rate of production, V_{max}. The Michaelis-Menten equation has robust applications in microbiology where it can describe microbial growth under substrate limited kinetics via the Monod equation [29].

$$\frac{dX}{dt} = \frac{\mu_{max} * S}{K_s + S} \tag{5}$$

In this case, X is the cell concentration and analogous to the product of the Michaelis-Menten equation. μ_{max} is the maximum growth rate. Equation 5 can readily describe substrate, biomass, and product concentrations by using yield coefficients. Michaelis-Menten kinetics may also be rearranged to describe repressor and inducer kinetics where the concentration of bound molecules is a function of inducer concentration, S_x, and the repressor protein, X_T [30].

$$X_t = X + [XS_x] \tag{6}$$

$$\frac{d[XS_x]}{dt} = k_{on} * XS_x - k_{off}[XS_x] \tag{7}$$

$$\frac{k_{off}}{k_{on}} = K_x \tag{8}$$

Such that at chemical equilibrium, an equation analogous to Michaelis-Menten arises [30].

$$[XS_x] = \frac{X_T S_x}{K_x + S_x} \tag{9}$$

In synthetic biology, multiple inducer molecules are often required to produce expression. For example, n = 2 for systems using the lac operon as induction requires 2 molecules of IPTG [31]. "Hill equations" describe kinetics for which "n" molecules are required for induction [30].

$$\frac{[nS_x X]}{X_T} = \frac{S_x^n}{K_x^n + S_x^n} \tag{10}$$

Higher order equations are called cooperative reactions as they require multiple molecules. Furthermore, as n increases the maximum rate of change of bound inducer to substrate concentration increases. Unbound repressor may also be described [30].

$$\frac{X^*}{X_T} = \frac{1}{1 + (S_x/K_x)^n} \tag{11}$$

These systems of equations can reliably explain synthetic circuits. An understanding of reaction kinetics and broad metabolic modeling can be compiled into Boolean networks and analyzed using linear algebra for pathway optimization.

1.2 A tripartite microbial consortium as a platform for bioproduction

Synthetic microbial communities may be engineered to produce bioproducts that are more sustainable than petrochemical analogues. The challenge of sustainability is to establish a process that fulfills a valuable societal function while reducing environmental and economic stress. To date, much of the global infrastructure was established through a common and general use of fossil fuels that, while contributing to prolonged prosperity, have led to undue environmental detriment. In 2010 it was estimated that global annual plastic pro-

duction exceeded 300 million metric tons, accounting for just under 10% of total worldwide oil consumption [32]. Sustainability is science driven innovation that works to safeguard the planet's life support systems while addressing sociopolitical paradigms. Such a microbial consortia is showcased by the coculturing of *Trichoderma reeseii*, which can hydrolyze lignocellulosic corn stover (*i.e.*, agricultural waste) into sugars, and a modified strain of *Escherichia coli*, which can convert sugars into isobutanol [33]. Together, a coculture of the modified *T. reeseii* and *E. coli* strains successfully degraded corn stover to ultimately yield higher isobutanol production than previously reported.

Recently, photosynthetic microbes have gained momentum due to their autonomous carbon metabolism, which enables renewable or semisynthetic substitutes for petroleum-based products in the pursuit of carbon neutral or carbon negative technologies. Common examples include the production algal biofuels as drop-in fossil fuel alternatives as well as production of bioplastics. In Europe, bioplastics like polyhydroxyalkanoate (PHA) and polylactic acid (PLA) have gained momentum due to their unique role in decoupling economic growth from resource depletion [34]. Industrial production of such biopolymers is under 1% with respect to petrochemical analogues, though this proportion has been rapidly increasing over the last decade [35]. Although useful in many instances, these bioplastics are biodegradable and limited in their applications. PHA and PLA, for example, have serious drawbacks in terms of their thermoplasticity, longevity, and subsequent applications in industry [32].

Although bioplastics and biofuels are derived from biological organisms their direct and indirect emissions often make them unsustainable even as alternatives to petroleum-derived products. Current biofermentation strategies suffer from high substrate-derived emissions. Carbohydrate sources often account for as much as 50% of total production costs [36]. Likewise, nitrogen fertilizers stem from Haber-Bosch ammonia, which accounts for an estimated 1 to 2% of global energy expenditures alone [36, 37]. Haber-Bosch ammonia synthesis also has an estimated 3 to 5% indirect conversion to NO_x, with recent studies asserting that these indirect emissions offset whatever positive impact bioproducts could have on the environ-

ment [38].

Carbon and nitrogen demand in biochemical production may be minimized by using organisms capable of producing their own substrates through BNF and photosynthesis. One such pair of organisms is *Azotobacter vinelandii* DJ and *Synechococcus elongatus* (PCC 7942). *A. vinelandii* DJ is a free-living proteobacterium capable of BNF in aerobic environments using Mo-, V-, and Fe-based nitrogenases. *A. vinelandii* employs two mechanisms for conserving nitrogenase under aerobic conditions [39]. The first is the production of polyhydroxybutyrate (PHB) and alginate biopolymers, which alter the chemistry of the cell membrane and reduce oxygen diffusion into the cytoplasm. A second mechanism increases cellular respiration, which removes oxygen through increased conversion to CO_2 [40]. Conversion of atmospheric nitrogen into biologically available ammonia is a highly energy intensive process involving 12 essential genes and 16 adenosine triphosphates (ATP) per reaction.

$$N_2 + 8e^- + 8H^+ + 16ATP + 16H_2O \longrightarrow 2NH_3 + H_2 + 16ADP + 16P_i \tag{12}$$

A. vinelandii also maintains polycistronic expression of the regulatory operon *nifLA*, which encodes for proteins nifL and nifA. NifL mediates ammonium repression through a negative feedback loop interacting with the PII signal-transduction protein glnK, which inactivates *nifA* to stop transcription of nitrogenase *nif* genes in ammonium rich conditions [41, 42]. Alternatively, redox conditions, which indicate an oxygen rich environment, or low levels of ammonia result in a decoupling of nifL and nifA thereby derepressing ammonium production [43–45]. Much work has explored the nifL and nifA regulatory proteins in BNF microorganisms with the coexpression of these genes providing a unique advantage for *A. vinelandii* modification. Deletion or knockout of the *nifL* gene allows for ammonia overproduction and intercellular excretion at concentrations between 10 and 35 mM, levels capable of supporting industrially relevant algal species [46, 47]. Analogous studies have considered remodeling the downstream incorporation of ammonia intracellularly in the L-glutamine glutamate synthase pathway (GOGAT) or via modification of the amtB ammonia

membrane transporter [37, 41]. *AmtB* deletion and glutamine synthase have indicated increased levels of ammonia excretion and shown accumulation of up to 10 mM after depletion of the carbon source [37, 46].

Cyanobacteria are the simplest organisms capable of photosynthesis, which uses ribulose-1,6-bisphosphate carboxylase (rubisco) to fix CO_2 into bioaccessible carbon via the Calvin-Benson-Bassham cycle. Cyanobacteria maintain a remarkable diversity of metabolic functions, including nitrogen fixation, heteroautotrophic growth on varied carbon sources, and a robust metabolic toolkit. One of the most studied cyanobacteria, *Synechococcus elongatus*, has a remarkably high photosynthetic efficiency, which is appealing given the propensity to sequester atmospheric carbon into useable substrates [48]. Manipulation of the Calvin cycle and specifically the notoriously inefficient rubisco enzyme has been a holy grail of biofuel research. Despite marginal improvements in the pathway by manipulation of the *rbcS* and *rbcL* genes, which encode for the large and small subunits of rubisco, recent studies have suggested that this enzyme might be near its biological optimum due to inherent similarities in carboxylation transition states during photosynthesis [49, 50]. Focus has instead shifted downstream towards Calvin cycle metabolites like pyruvate and acetyl-CoA as well as CO_2-concentrating mechanisms [49, 51]. *S. elongatus* has been genetically modified to produce biofuel products, ranging from 1-butanol, isobutanol, and isobutyraldehyde [52]. While produced at relatively high titers, other works have focused on the exportation of hydrophilic products, namely simple sugars [36]. *S. elongatus* produces sucrose as a compatible solute and, under significant osmotic stress, this cyanobacterium can accumulate up to 300 mM sucrose intracellularly [48]. Recent efforts have capitalized on this natural mechanism by cloning the *cscB* gene, a sucrose symporter in *E. coli* W, into *S. elongatus* under an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter [48]. Results indicate extracellular accumulation of sucrose at rates up to 36 mg/L/h from photosynthesized carbon in CO_2 rich conditions, 150 mM NaCl, and 0.1 M IPTG [48]. Under semi-continuous culturing, an analogous strategy in *S. elongatus* UTEX 2973 maximized sucrose concentration at 8.6 g/L

in 21 days [53].

A. vinelandii and *S. elongatus* share abiotic factors including pH, temperature, and micronutrient concentrations for optimal growth. A coculture of *A. vinelandii* and *S. elongatus* encapsulated in hydrogels was studied in a limited capacity to demonstrate symbiotic cross-feeding for biopolymer production [47]. The engineered coculture successfully generated PHB from the *A. vinelandii* strain at a rate of 20% of total biomass without external carbon or nitrogen substrates though with a notable reduction in growth rate [47]. While promising, PHB is limiting in both scope and application. Studies conducted by Hays et al. have cocultured *S. elongatus* with various heterotrophs, including modified *S. cerevisiae*, *B. Subtilis*, and *E. coli* W and demonstrated successful and stable coculture growth in sugar free minimal media [54]. The commonality of these heterotrophs for microbial production is essential and these experiments demonstrate general design principles and an unusual robustness in pairing the *S. elongatus cscB* strain [54]. Combined, these works suggest that the chassis of nitrogen and carbon fixating organisms can sustain the growth of a third producer strain and present a novel microbial consortium for investigation.

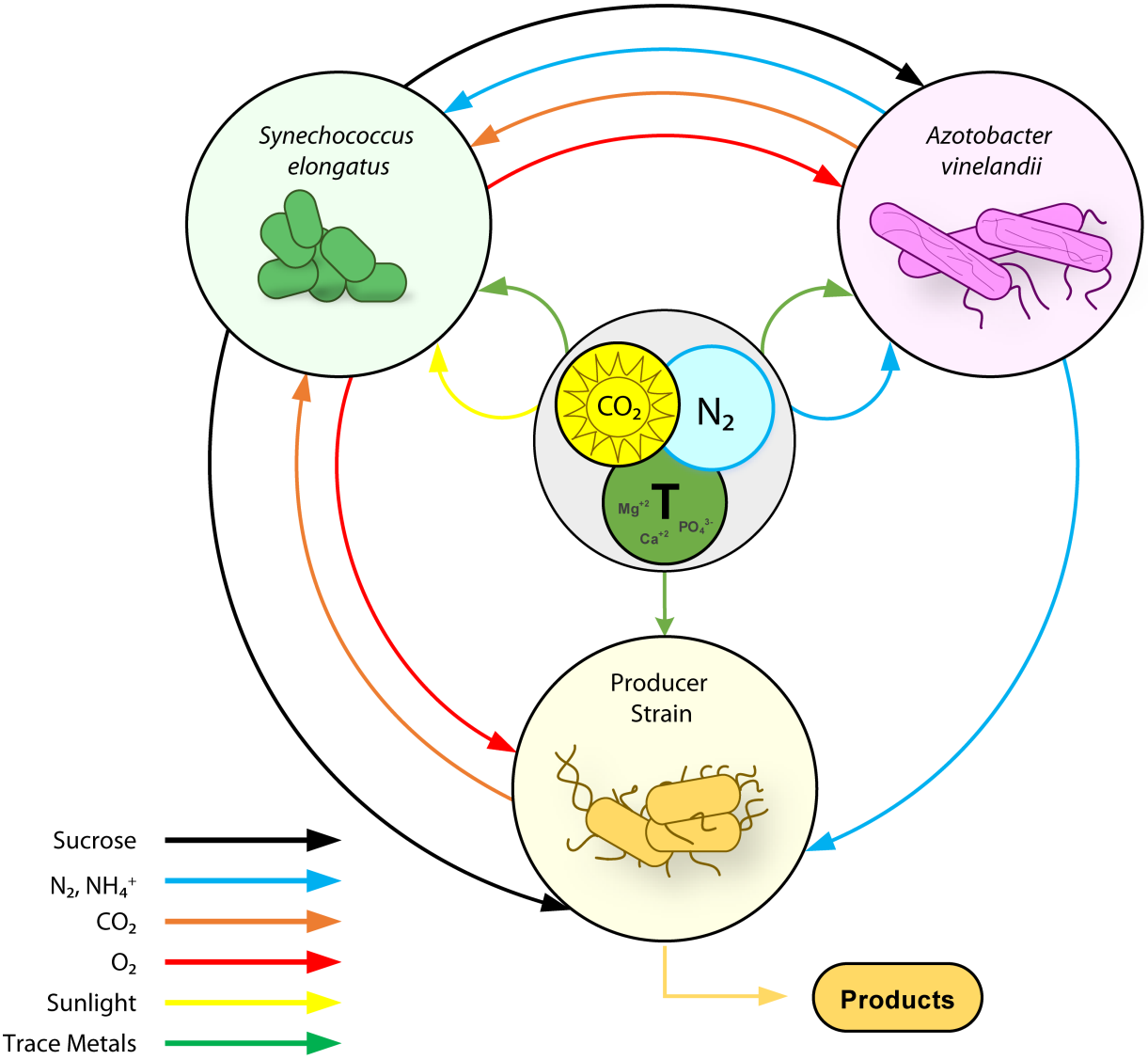


Figure 1.4: Depiction of the proposed tripartite symbiosis in which *A. vinelandii* fixes nitrogen and *S. elongatus* fixes carbon to drive growth and production of a third producer strain. The system is modular such that any strain capable of using on sucrose and ammonia can be grown.

Fig. 1.4 illustrates the overall design framework of this consortium. It depicts molecular nutrient flux between specialists. *E. coli* was selected as the baseline producer as it is exceptionally well studied and fundamental for many biological investigations. Other options for producer species are listed in Appendix A. In the context of a synthetic consortium, biologically fixed nitrogen and photosynthetic sucrose are limiting, though can be tuned by spatial and temporal control. Given that limiting dinitrogen concentrations are highly unlikely under normal atmospheric conditions and CO_2 may be fed into a growth chamber,

the rate-limiting steps are centralized around nitrogenase and rubisco [39]. Construction and tuning of this microbial consortium form the basis of this dissertation.

1.3 Chromosomal integration and heterologous gene expression

The most common strategy for heterologous gene expression is through highly transferable circular DNA sequences called plasmids. Plasmids are natural shuttles for certain genes, namely antibiotic resistance cassettes, that enable certain bacteria to thrive under unique circumstances. Unfortunately, plasmids require a method of selection (temperature, antibiotic, or intrinsic method) to maintain. As a result, plasmid expression is less desirable compared to chromosomal expression due to metabolic and monetary maintenance costs of the selection method and general instability. While the BioBrick framework discussed in section 1.1.2 facilitates construction and elucidation of genetic circuits, long term stable biofermentation platforms necessitate chromosomal integration.

Integration onto the chromosome is generally accomplished through site-specific recombineering or phage-derived transposons. λ-Red recombineering is an especially popular strategy for site-specific chromosomal integration. The λ-Red system comprises three genes: *exo*, *gam*, and *beta*. Exo degrades linear dsDNA, gam inhibits host nucleases, and beta enables binding to site-specific regions on the target chromosome simply based on flanking homology regions. While most widely used, λ-Red requires 30 to 50-bp flanking homology and suffers low efficiency for fragments over 2.5 kb such that integration of entire operons is often difficult. [55, 56] Other mechanisms such as conditional-replication, integration, and modular (CRIM) plasmids and clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 have enabled highly specific integration into the genome. However, it remains site-specific integration efficiency or optimality are often unclear due to chromosomal topography (GC content, extended protein occupancy domains, coiling, etc.). Researchers have only recently begun to explore and quantify locus dependent expression.

The tripartite microbial consortium requires organisms that can thrive on both su-

crose and ammonia as well as similar alkalinity, buffer, and micronutrients. Although *E. coli* K-12 MG1655 is by far the most common organism for genetic modification, it cannot naturally catabolize sucrose. *E. coli* W, a unique strain isolated from Rutgers University, can metabolize sucrose by a non-phosphotransferase (non-PTS) pathway and has been demonstrated to grow well in cocultures with *S. elongatus* [54, 57]. Many studies have cloned the non-PTS genes onto plasmids that can be transformed into any *E. coli* K-12 strain. By conferring sucrose catabolism, transformed K-12 strains can be grown in the tripartite system, which poses immediate interest to Section 1.2. Unfortunately, studies have reported inherent drawbacks of episomal sucrose catabolism including one instance of spontaneous chromosomal integration [58]. Furthermore, it remains unclear whether growth on sucrose is a strong enough selection to maintain the plasmid in subsequent passages as more stringent methods like antibiotics are unusable in polyculture.

Random integration of the sucrose catabolism operon into *E. coli* K-12 MG1655 using Tn5 transposase yields a library of growth rates based on integration locus. Select integrants from this library can be used to probe dynamics of the tripartite system and be transduced into production strains with local homology for carbon substrate switching from glucose to sucrose. Transduction of a fast growth rate locus into production strains enables better sustainability metrics compared to growth on glucose.

1.4 Life cycle assessments as a tool to analyze relative sustainability

Life cycle assessments (LCAs) realize the holistic energy intensity and net emissions of a selected process through a comprehensive and quantitative analysis. These assessments are fundamental to quantifying the relative sustainability a process and are the gold standard for determining the net benefits of one process (*e.g.*, burning petroleum diesel in compression-ignited direct injection engines (CIDI)) over another (*e.g.*, burning biofuel in CIDI engines). The guidelines for conducting LCAs are dictated by the International Organization for Stan-

dardization (ISO) 14040 series and define scope, inventory analysis, impact assessment, and interpretation as illustrated in Fig. 1.5 [59,60]. Each of these steps are inherently intertwined and defined, evaluated, and adjusted accordingly. Most assessments are comparative LCAs and describe two parallel pathways that generate a similar if not identical product.

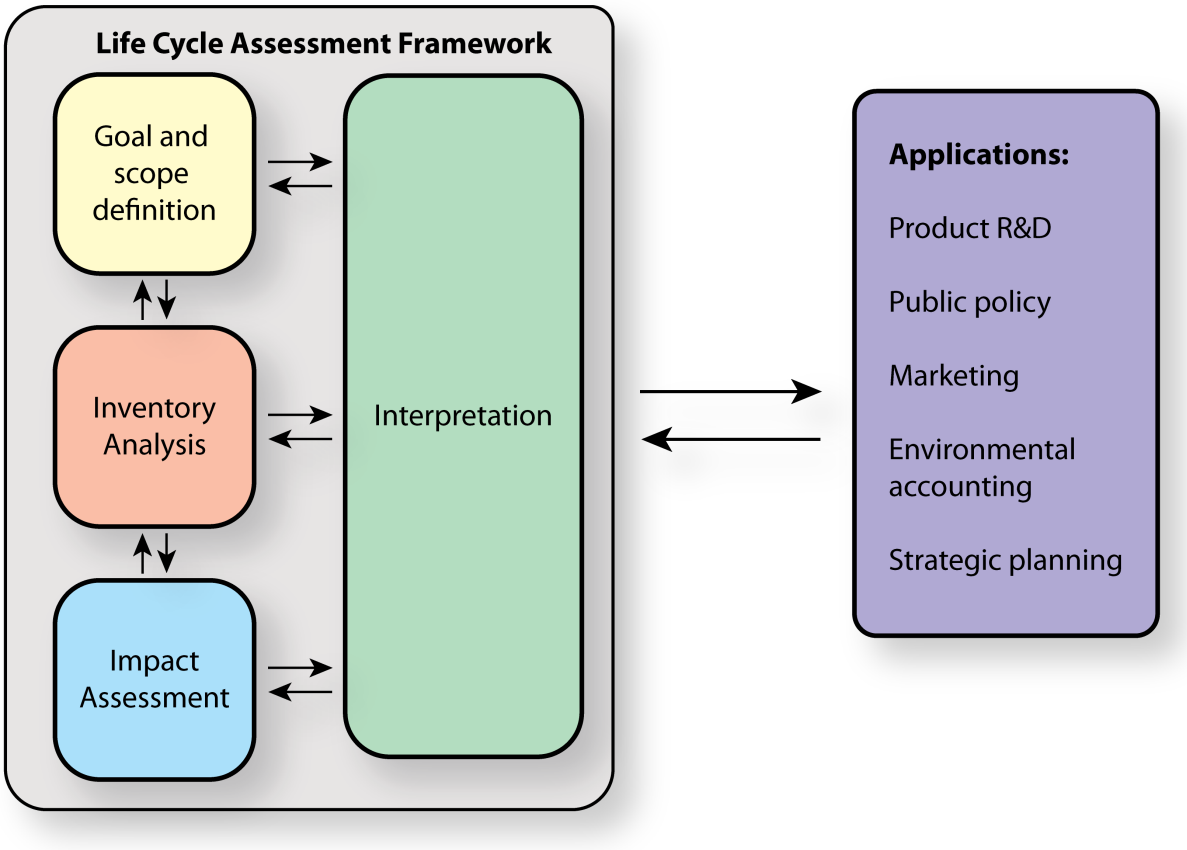


Figure 1.5: Life cycle assessment framework as described by the ISO 14040. ISO 14040 states guidelines on environmental management and the principles/framework involved in LCAs.

All inputs and outputs are enumerated within the goal and scope of a production scheme. First and foremost, this includes the system boundaries, the system functional unit, and objectives of the LCA. For biofuels, the functional unit is the volume of fuel while the boundaries dictate where accounting begins and ends. Most bioproducts are referred to as cradle-to-grave or well-to-wheel, where "well" refers to the oil well or analogous pond. The selection of an appropriate boundary condition directly affects the overall conclusions. For

example, ignoring upstream cultivation of sugarcane or corn for fermentable carbon substrate production could suggest that a bioproduct is sustainable. However, the reality of agricultural practices, land-use-change, fertilizers, etc. associated with the carbon substrates each have dramatic effects in the overall production scheme. The scope and the sheer number of variables in a given LCA necessitates many assumptions. All relevant assumptions pertaining to electricity composition, data sourcing, homogeneity of conditions, and so forth must be explicitly noted. Many resource inputs change seasonally and, especially in the case of algal biofuel production, estimating temporal and environmental fluctuations is critical for process viability.

The life cycle inventory (LCI) quantifies the inputs and outputs of a given process pathway in the context of the prescribed functional unit. Often the inventory is a model that incorporates variation within the process for sensitivity analysis using a Monte Carlo simulation. The LCI is translated from inputs and outputs to impacts by an impact assessment. Many impact assessment frameworks have been developed, including the Greenhouse gases, Regulated Emissions, and Energy in Transportation (GREET) utility, EcoInvent®, and SimaPro®, which each may assign different environmental weights within the context of the scope and goal of the project. Common impact categories for bioproduction usually include greenhouse gas emissions (GHGs), nonrenewable energy usage (NREU), and water intensity (WI). Others may also explore human health effects, acidification, ozone, eutrophication, etc., which may be weighted again based on the goal of the LCA. Finally, interpretation of results and usability draw a conclusion about the relative sustainability of the product, process, or pathway with respect to another. Often results can direct funding, research, or policy changes surrounding a specific process, pathway, or product.

The LCA framework outlined here is continually referenced or alluded to throughout this dissertation, though more explicitly noted within Chapter 4. In Chapter 4, an LCA is conducted that closes a knowledge gap between algal cultivation, extraction, and life cycle assessment through a cohesive well-to-wheel examination of various algal cultures and

culture conditions. Experimental data for the model is generated through multidisciplinary work published at the University of Michigan, comprising 37 combinations of highly productive algal species extrapolated to an industrial process [25, 37, 61, 62]. The scope includes algal pond cultivation, conversion using hydrothermal liquefaction, and combustion in CIDI engines using a functional unit of 1 MBTU transportation energy. Key empirical nutrient inputs for growth experiments include nitrogen substrates, trace metals, and phosphates that each carry different energy and emissions burdens. Interpretation of this algal LCA is framed by energy and emissions sustainability impact metrics through NREU, GHGs, and WI in pseudo-steady state. Each pathway is quantitatively analyzed using the GREET utility. The utility amalgamates transportation data with life cycle inventory coefficients so as to assess the sustainability and efficiency of a fuel pathway [63]. The work also integrates publications by the Pacific Northwest National Laboratory and Argonne National Laboratory into a modular framework within an ecological context [64, 65].

1.5 Abiotic production from CO_2

Carbon dioxide is a thermodynamically stable compound with low reactivity and requires high energy addition or specialized catalysts for chemical conversion. Indeed, efficient catalysts have been noted as a major barrier for optimization of abiotic CO_2 derived products [66].

While this work focuses entirely on photosynthetic fixation of carbon dioxide, abiotic fixation has been explored for decades as a means of converting petrochemical byproducts into useful chemicals. Many catalytic pathways have been developed for the conversion of CO_2 to usable chemicals including carbonates, carbamates, lactones, and formic acid [67]. The most common use of CO_2 feedstocks are in reactions with carbon monoxide and hydrogen gas in the mixture called syngas, which can be used to drive production of methanol and its derivatives [68]. The set of reactions, namely the water gas shift and methanol synthesis, are well studied and commercially viable.

In theory, supply of CO_2 to these reactions can be provided by atmospheric carbon capture and sequestration (CCS), however current practices rely on flue gas for commercial viability. This is largely due to the relatively low concentration of CO_2 in the atmosphere at approximately 400 ppm compared to the concentration of CO_2 in flue gas. CCS may become increasingly viable as energy paradigms shift towards renewable sources (wind, solar, etc) that can fix carbon without significant nonrenewable energy usage. The current reliance upon flue gas dramatically reduces the sustainability of abiotic CO_2 derived products and complicates assessment of life cycle metrics, which must account for nonrenewable energy usage within the boundary conditions.

To date, few life cycle assessments have compared the relative sustainability of abiotic to biotic carbon fixation, though these alternative processes can be potentially integrated in a complimentary fashion. For example, a hybrid process could combine growth of an algal or cyanobacterial production culture for bioproduct generation using photosynthetically fixed CO_2 and subsequent combustion of spent biomass to produce flue gas for abiotic conversion. The most efficient strategy for CO_2 derived chemicals may turn out to be one with similar integrative approaches. This is an exciting opportunity and of particular interest for the future of sustainability.

1.6 Dissertation overview

Broadly, this dissertation describes how microbial consortia can be engineered or manipulated to enhance sustainability. This includes genetic modification of specific organisms to enable nutrient cross-feeding within photosynthetic consortia (Chapter 2 and Chapter 3) as well as capitalizing upon natural microbial predispositions for bioproduction (Chapter 4). The specific chapter summaries are as follows:

Chapter 2: Engineering modular carbon and nitrogen fixing microbial consortia for sustainable biochemical production: This chapter explores tripartite synthetic consortia as bioproduction platforms in which carbon dioxide and atmospheric nitro-

gen are fixed by constituent bacteria for feeding to a producer strain. Growth of *E. coli* K-12 *cscBKA*::*nlpA*, *B. subtilis*, and *C. glutamicum* is demonstrated over three dilutions of iterative continuous culturing experiments. A comprehensive mathematical model elucidates metabolic bottlenecks within the consortium.

Chapter 3: Random chromosomal integration of sucrose catabolism in *E. coli* K-12 yields readily transducible high growth rate loci: Chromosomal expression is advantageous compared to episomal expression, though is affected by a myriad of locus-specific factors. This chapter describes random integration of a non-PTS sucrose catabolism operon onto the *E. coli* K-12 chromosome. Random integration generates of a library of integrants with variable growth rates on sucrose. A high growth rate locus is transduced into a chromosomally integrated isobutanol producing strain of *E. coli*, yielding better sustainability metrics.

Chapter 4: Microalgal communities enhance multifunctionality in a life cycle assessment of biofuel production: A module for assessing the relative sustainability of algal polycultures compared to monocultures is created and applied to empirical pond growth and biocrude conversion data. Certain polycultures were able to outperform monocultures due to enhanced multifunctionality across multiple factors including enhanced biomass production, relative nutrient use, and resistance to pond crashes.

Chapter 5: Conclusion and future directions: Here, I contextualize this dissertation within the field of synthetic biology and briefly explore avenues for the next stages of research and development for each chapter.

Chapter 2: Engineering Modular Carbon and Nitrogen Fixing Microbial Consortia for Sustainable Biochemical Production

Engineering of synthetic microbial consortia has emerged as a new and powerful biotechnology platform. To date, most microbial consortia have focused on biofuel development, though they also have enormous potential in the production of biobased commodity chemicals. We describe a tripartite system in which three microbes of differentiated specializations can convert sunlight, carbon dioxide, and atmospheric nitrogen into chemical precursors for bulk polymer production. Specifically, *Azotobacter vinelandii*, a nitrogen-fixing bacterium that secretes ammonia, and *Synechococcus elongatus*, a photosynthetic cyanobacterium that exports sucrose, form a symbiotic chassis hypothesized to support a third producer strain. We demonstrate supported growth for a wide range of producer strain candidates, namely a transducible fast-growth K-12 derivative (*E. coli* K-12 *cscBKA*::*nlpA*), an industrial production strain (*Corynebacterium glutamicum*) and a robust chassis strain (*Bacillus subtilis*). Tripartite consortia were grown under continuous culture conditions and complemented by an empirical kinetic model for platform optimization. This framework bridges a fundamental gap between commercial biofermentation and community engineering, providing a potential strategy to alleviate the energetic and economic constraints barring market entry of specific bioproducts.

2.1 Introduction

Advances in metabolic engineering have contributed to the development of microbial cell factories for bioproduction. Platforms have demonstrated production of a vast library of chemicals from simple alcohol biofuels to beneficial pharmaceutical compounds. However, the scalability and success of biofermentation is hampered by costly fertilizers and fermentable carbohydrates. Substrates can account for over 50% of total monetary production costs while adding significant energetic and environmental burdens depending on the life cycle of upstream processes [69, 70].

Recent studies have investigated the design and application of microbial consortia as a strategy to reduce the environmental and energetic burden associated with bioproduction substrates. Natural microbial consortia are exemplified by the plant rhizosphere in which mycorrhizal fungi, plant-growth promoting rhizobacteria, and plants have coevolved syntrophic interactions for nutrient fixation and cross-feeding that reduce the metabolic burden of complex processes (*e.g.*, biological nitrogen fixation (BNF) and photosynthesis) on any single organism within the community. This division of labor contributes to increased resistance to invasion and abiotic stress, increased biogeochemical recycling, and ultimately enhanced productivity of the community as a whole [4]. Synthetic microbial consortia can capitalize upon natural predispositions by selecting for interspecies relationships or specific community functions. Pairing photosynthetic and diazotrophic organisms with metabolic engineering, for example, is a powerful strategy for simultaneously driving bioproduction while reducing the environmental, energetic, and monetary burden of carbon and nitrogen substrates.

Nitrogen is scarce in terrestrial and aquatic ecosystems and, as a result, diazotrophic bacteria have complex regulatory pathways to facilitate BNF and maintain fixed nitrogen. *Azotobacter vinelandii* DJ is a canonical free-living diazotoph that produces three heterometal V, Mo, and Fe based nitrogenases. Paradoxically, *A. vinelandii* is capable of

fixing nitrogen under strict aerobic conditions by tight regulation of nitrogenase production, cellular respiration, and cell membrane permeability by production of alginate and polyhydroxybutyrate (PHB) [71]. The diazotroph is also capable of accumulating iron chelating siderophores that facilitate nitrogenase biosynthesis. Despite the relative complexity of the nitrogenase enzyme and its myriad of accessory nif genes and metal complex co-factors, its regulation is controlled by the operon *nifLA*. *nifLA* encodes for a sensor protein *nifL* and a σ^{54}-dependent transcriptional activator *nifA* [72]. Under conditions of ammonium excess, NifL binds and inactivates NifA, halting the production of nitrogenase [42]. Disruption or knockout of *nifL* has yielded viable strains that overproduce nitrogenase and secrete between 10 mM and 30 mM NH_4^+ [73–75]. Synthetic commensal relationships of *A. vinelandii* and freshwater algae have been studied to exploit *Azotobacter* as an ammonia overproducing biofertilizer, supplying either ammonia (*C. sorokiniana* RP, *S. obliquus* C1S, *Pseukirchneriella* sp. C1D) or the siderophore azotobactin (*N. Oleoabundans* and *Scenedesmus* Sp. BA032) [75–77].

Many photosynthetic organisms have also been reengineered to reroute fixed carbon towards desirable products. Cyanobacteria naturally produce carbohydrate osmolytes under salt stress, including glycogen, trehalose, and sucrose in freshwater species and glucoslyglycerol, glucosulglycerate, and glycine betaine in saltwater species [78]. Osmolytes can account for up to 50% of total fixed carbon [78]. Addition of osmolyte specific transporters has been demonstrated to enable export of carbohydrate into the extracellular media [69, 79]. In a specific study *Synechococcus elongatus* (PCC 7932) was modified for heterologous expression of a non-PTS sucrose permease (*cscB*) under an IPTG-inducible promoter [69]. Growth in mildly alkaline conditions and 150 mM NaCl enabled sucrose export of up to 85% of fixed carbon at high rates of 36 mg/L/hr into the extracellular medium [80]. Further modification of the sucrose phosphate synthase (SPS) gene enabled inducible sucrose production in the absence of osmotic pressure, thereby maintaining a comparable growth rate to WT [81]. Interestingly, strain engineering enhanced rates of photosynthesis, suggesting

that carbohydrate export is a sink that relieves photosynthetic inhibition. Sugar exporting *S. elongatus* derivatives have been demonstrated to support heterotrophic organisms (*E. coli* W, *S. cerevisiae*, and *B. subtilis*) purely on photosynthetic carbohydrates and without preexisting relationships [78, 82–84].

The relative similarity of growth conditions for *S. elongatus* and *A. vinelandii* as well as their capacity to generate syntrophic nutrients has been explored generate a self-sustaining coculture. By coculturing *S. elongatus* (*cscB*) and *A. vinelandii* (AV3), Smith et al. were able to demonstrate low-level production of PHB in a small scale, membrane separated system [85]. The study serves as a proof of concept for consortia that capitalize upon this division of labor and presents an opportunity for integration of more robust bioproduction candidates.

Here, we describe a tripartite platform in which *S. elongatus* (*cscB* + SPS) [81] and *A. vinelandii* (AZBB163) [74] drive production of a third producer species in continuous culture with limited nutrient addition. As ammonia and sucrose are widely utilizable substrates, the platform enables modular addition of an industrially relevant species including *Corynebacterium glutamicum*, a global producer of amino acids, as well as chassis strains like *S. cerevisiae* and *B. subtilis* which maintain robust metabolic toolkits. Specifically, we demonstrated continuous culturing of a fast-growth sucrose integrant *E. coli* K-12 MG1655 *cscBKA*::*nlpA*, *C. glutamicum* 13032, and *B. subtilis* (168) over three sequential dilutions (Fig. 2.1). We also developed a kinetic model for elucidation of platform bottlenecks and opportunities for further strain optimization. While further genetic modification is necessary to realize the potential of the tripartite platform, our work is a key development for sustainable bioproduction.

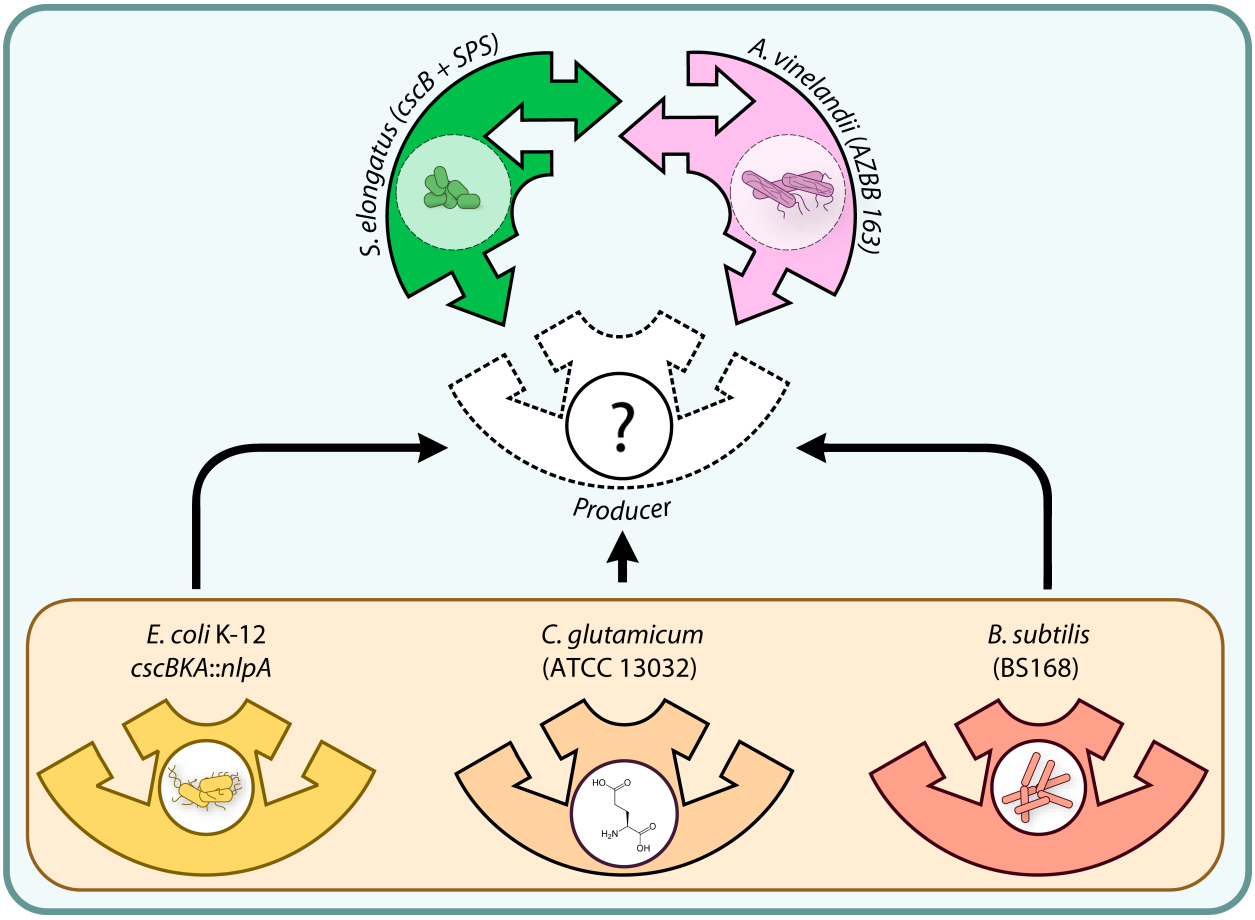


Figure 2.1: The *S. elongatus* and *A. vinelandii* coculture provides a modular framework for production strains, including readily modifiable strains like *E. coli* and *B. subtilis* as well as industrial producers like *C. glutamicum*.

2.2 Results and discussion

2.2.1 System overview

Development of the tripartite platform required investigation of potential diazotrophic and photosynthetic candidates. Ultimately, *S. elongatus* was selected due to high photosynthetic productivity and *A. vinelandii* was selected due to its unique role as a free-living aerobic diazotroph. These disparate organisms were grown on unique media with nutrient requirements that differ significantly from both each other and from common production strain media. A union minimal media was synthesized by using design of experiment and

screening for both optimal strain production and biomass accumulation (Fig. 2.2A). Ten production strains were identified as potential candidates for integration into the tripartite system. Candidates included chassis strains like *E. coli* W, *E. coli* K-12 with a transducible fast-growth *cscBKA* locus, *B. subtilis* 168, and *S. cerevisiae* as well as production strains like *B. licheniformis* (ATCC 9945a), *B. subtilis* "natto", *C. glutamicum* 13032 and *C. glutamicum* 25126. Lastly, candidates included two *E. coli* K-12 strains with plasmid mediated production of a fluorescent protein and sucrose catabolism. A physical culturing apparatus was constructed and tailored to the needs of individual consortium constituents, namely through membrane separation of each member, irradiance, and gas delivery (Fig. 2.2B). Growth and production regimes were identified to separate initial nutrient priming of the *A. vinelandii* − *S. elongatus* coculture and the subsequent continuous culturing of selected production strains (Fig. 2.2C). Finally, a tripartite model using Monod kinetics had good corroboration with experimental data and enabled a correlation analysis of parameter values to system outputs, namely nutrient and constituent concentrations for future optimization of the tripartite platform.

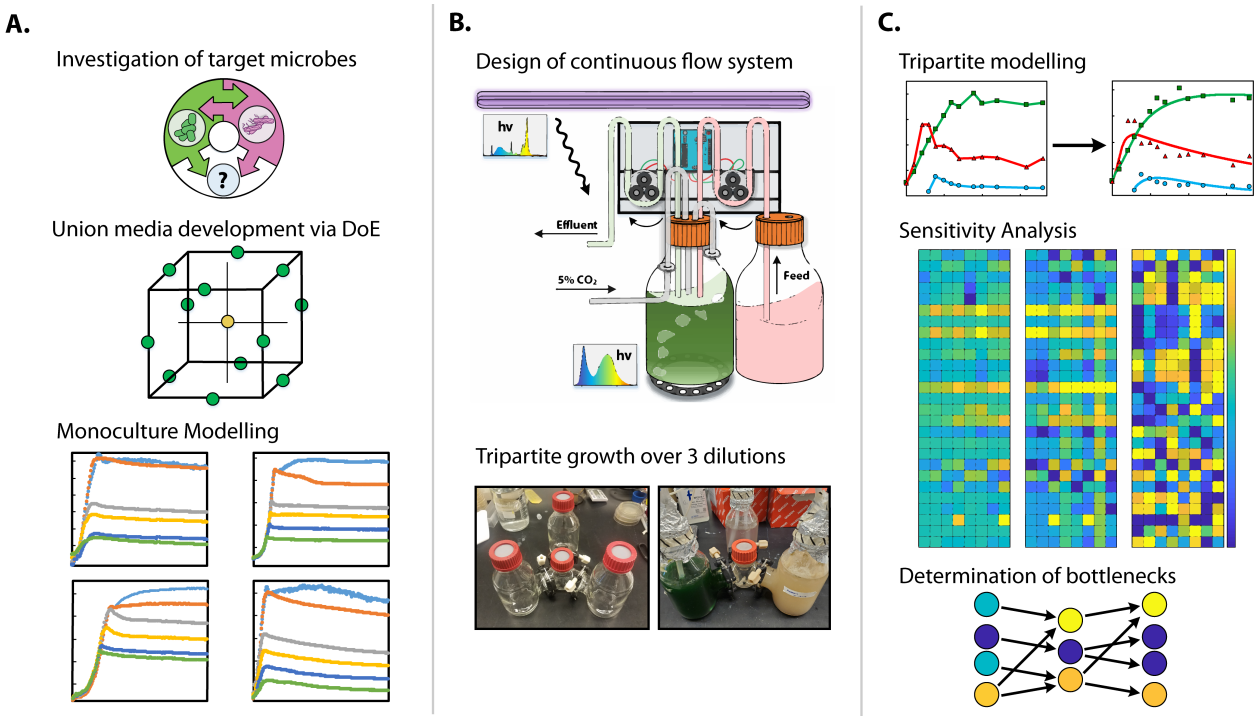


Figure 2.2: Overview of the investigation, construction, and analysis of the tripartite platform. (A) Strains were initially investigated for integration into the system and grown in a union minimal medium that was synthesized from respective strain minimal media. Monoculture growth and production were modeled using Monod kinetics. (B) A physical platform for tripartite cultivation was developed using continuous flow and membrane separation. (C) The kinetic model was compared to experimental data for each consortium. Lastly, a correlation analysis was conducted to identify key bottlenecks, which represent opportunities for further optimization and genetic modification.

2.2.2 Formulation of a union medium using design of experiment

A union medium was synthesized between Burk's Nitrogen-free medium for *A. vinelandii*, BG-11 medium for *S. elongatus*, M9 medium for *Escherichia coli* K-12 MG1655, and CG XII medium for *Corynebacterium glutamicum*. Extensive design of experiment was conducted to optimize monocultural growth of constitutive co-cultures and tri-cultures by varying buffer, salt, and metal concentrations. In total, 235 experimental conditions for optimal product yield (*i.e.*, sucrose from *S. elongatus* and ammonia from *A. vinelandii*) and biomass yield from constitutive species were screened under carbon and nitrogen limited conditions. Some findings, for example that high sucrose concentration favored diazotroph growth while high

ammonia concentration favored cyanobacterial growth, were obvious under nutrient limiting conditions. However, design of experiment also indicated that high phosphate buffer concentrations were detrimental to *S. elongatus* and low iron concentrations hindered both *A. vinelandii* growth and production of ammonia. Such micronutrients were therefore tuned specifically for those members of the consortium. Overall pH and temperature were amenable though not necessarily optimal for all strains evaluated (Fig. 2.3).

	Medium	BG11	Burk's	M9	CG XII	BB11
	Organism	*S. elongatus*	*A. vinelandii*	*E. coli*	*C. glutamicum*	Union
Carbon Source	Sucrose		20.0			V
	Glucose			4.0	20.0	
Nitrogen Source	$NaNO_3$	1.5				V
	NH_4Cl			1.0		
	NH_4SO_4				20	
	Urea				5	
Buffers	K_2HPO_4	0.04	0.2		1.0	0.2
	KH_2PO_4		0.6	3.0	1.0	0.8
	$Na_2HPO_4*7H_2O$			6.0		
	MOPS				42	
	HEPES	1.0				1
	Na_2CO_3	0.02				0.02
Salts	$MgSO_4*7H_2O$	0.075	0.2	0.25	0.25	0.2
	$CaCl_2*2H_2O$	0.036	0.09	0.015	0.013	0.09
	NaCl			0.5		1.0
Metals	Critic acid	0.006				0.018
	Ammonium ferric citrate	0.006				0.018
	$Na_2EDTA*2H_2O$	0.001				0.001
	BG11 Trace Metal Mix	1 mL				1 mL
	$FeSO_4*7H_2O$		0.005			0.015
Growth Factors	PCA				0.030	0.030
	Biotin				$2.0*10^{-4}$	$2.0*10^{-4}$
	IPTG	1 mM				1 mM
Conditions	pH	8.0	7.0	7.0	7.0	8.0
	Temperature	30	30	37	30	30

Figure 2.3: Union Medium composition consisting of a union between BG-11 [80], Burks N-free [70, 74, 86], CG XII [87], and M9. Components are listed in grams with the exception of pH and temperature and unless otherwise specified.

Solution 1 was a 10x solution of 2.0 g KH_2PO_4, 8.0 g K_2HPO_4, and 10.0 g NaCl in 1 L

of deionized H_2O. Solution 2 was composed of 0.2 g $MgSO_4$*$7H_2O$, 0.9 g $CaCl_2$*$2H_2O$, 0.1 mL of 10 mM Na_2MoO_4 * H_2O, 1.0 mg of Na_2EDTA*$2H_2O$, 20 mg Na_2CO_3, and 1 mL of 1000x solution BG-11 metal mix in 800 mL deionized H_2O. Solution 3 was a sterile filtered iron solution containing 0.25 g $FeSO_4$*$7H_2O$, 0.30 g citric acid, and 0.30 g ammonium ferric citrate in 50 mL deionized H_2O.

Skeletal Union Medium consists of 100 mL of autoclaved 10x Solution 1, 800mL of autoclaved Solution 2, and 3 mL of 1000x sterile filtered Solution 3 for an excess of iron. Solutions of protocatechuic acid (PCA, 3,4-dihydroxybenzioic acid), biotin, and HEPES were added for final concentrations of 30 mg/L, 0.2 mg/L, and 1 g/L, respectively. The solution pH was then adjusted to 8.0 with approximately 875 μL of 5 M NaOH and deionized water was used to adjust the volume to 1 L to form the skeletal Union Medium (*i.e.*, Union Medium without primary carbon or nitrogen sources). The Union Medium Primer further contained 5.0 g/L sucrose and 0.4 g/L $NaNO_3$ to facilitate initial growth of the *S. elongatus* and *A. vinelandii* strains prior to inoculation with the producer.

2.2.3 Kinetic analysis of strain growth and production

Production strain candidates were grown in 200 μL volume 96-well plates in the Union Medium, incubated at 30 °C with linear shaking, and provided with baseline initial nutrients (5 g/L sucrose and 0.4 g/L NH_4Cl). Strains were selected based on their robust genetic toolbox and prominence in the literature (*E. coli*, *B. subtilis*, and S. cerevisiae), use in industrial bioproduction (*C. glutamicum* 13032, *C. glutamicum* 25126, *B. licheniformis* 9945A, and *B. subtilis* "Natto"), or as strains to probe the system (*E. coli* K-12 pAKB, *E. coli* K-12 pBKA, and *E. coli* K-12 *csc* chromosomal integrants. Variable growth rate integrants were of particular interest as candidates in the tripartite system due to their differing growth kinetics and subsequent effects on overall nutrient concentration. The plasmid strains were also fluorescently labeled for secondary screening.

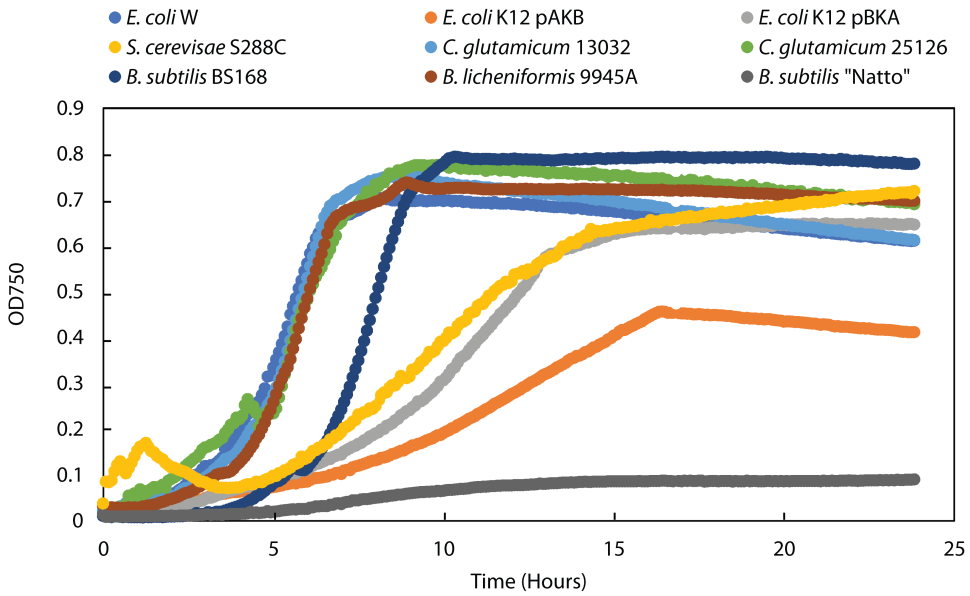


Figure 2.4: Growth curves of all production strain candidates in microplates with Union Medium with 5 g/L sucrose and 1 g/L NH_4Cl (n = 3).

All production strain candidates grew under the aforementioned conditions, but with markedly different kinetics as illustrated in (Fig. 2.4). Slow *E. coli* pAKB growth with respect to pBKA is consistent with published data. The relatively low absorbance of *B. licheniformis* stems from accumulation of cells at the sides of plate wells due to linear shaking. Although difficult to discern, *E. coli* W had the highest growth rate, again corroborating published work. *E. coli* W, *C. glutamicum* 13032, *B. subtilis* 168, and *E. coli* K-12 MG1655 *cscBKA*::*nlpA* served as initial candidates due to significantly different growth kinetics and, in the case of *C. glutamicum* 13032, bioproduct generation.

Selected strains were grown under carbon and nitrogen nutrient limited conditions to analyze growth kinetics. It was observed that most strains required an order of magnitude (10x) more sucrose than NH_4Cl by mass. High resolution growth curves were generated by kinetic reads in microplate under similar conditions to the initial strain investigation but with varied sucrose concentration. Strains were passaged once in Union Medium prior to kinetic analysis. Concentrations varied from 5.0 g/L, which is the level used for tripartite system priming, to 0.25 g/L. Four selected strains are depicted in Fig. 2.5.

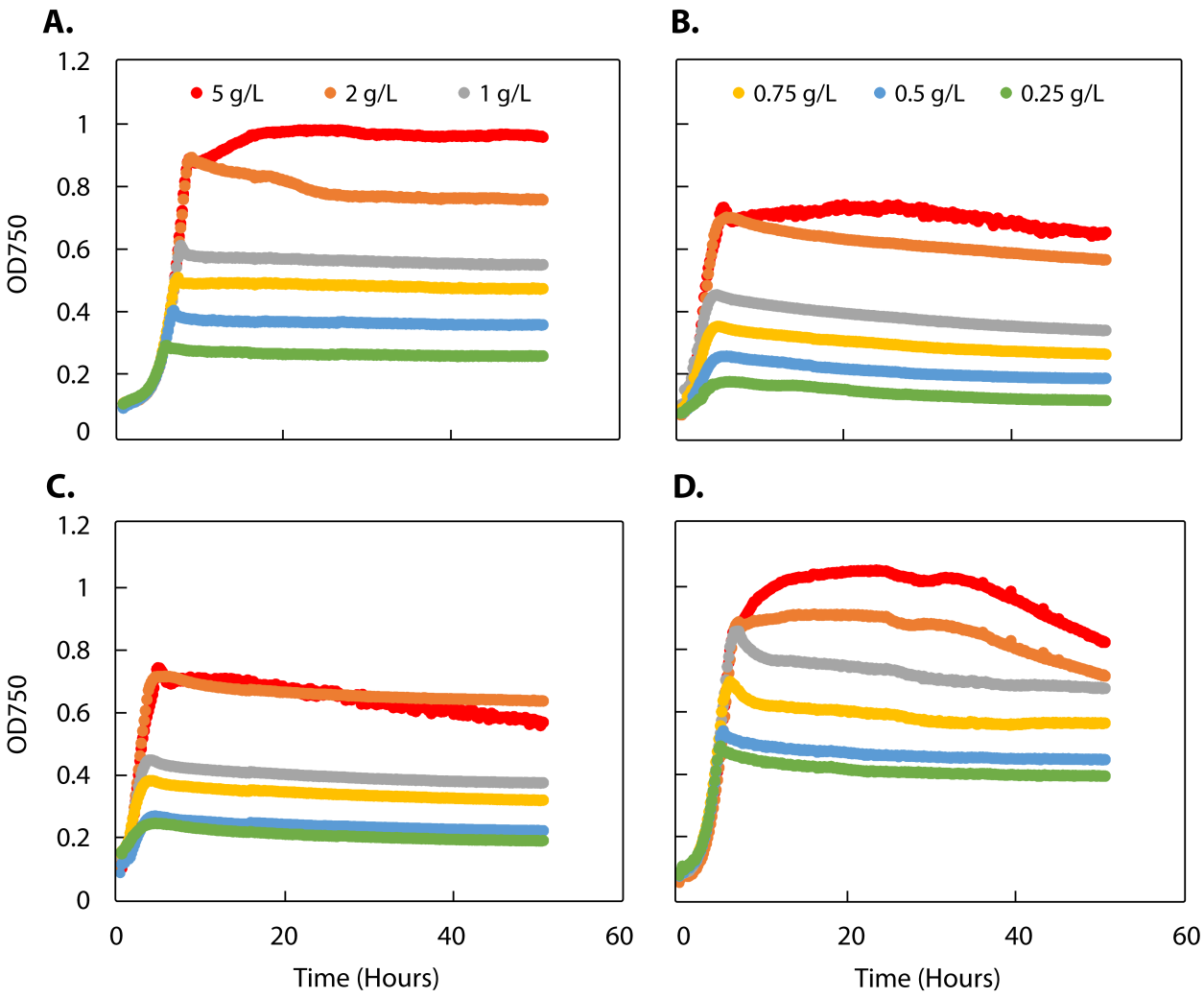


Figure 2.5: Growth curves of potential producer species under sucrose limited conditions and excess NH_4^+ (n = 3). (a) *C. glutamicum* 13032, (b) *E. coli* W, (c) *C. glutamicum* 21526, and (d) *B. subtilis* 168

Growth curves were fit to a Monod kinetic model assuming sucrose limited conditions ($OD_{750} < 0.8$). For the four selected production strains in Fig. 2.5, this included growth on sucrose between 0.25 g/L and 1.0 g/L. Curves were fit using growth data for each strain at each concentration and fit using the Levenberg-Marquardt algorithm to find the non-linear least squares using Equations 26 and 32. Best fit parameters yielded high confidence parameters that were well fit for different conditions. Between 1.0 g/L to 0.25 g/L for *C. glutamicum* 13032, for example, $\mu_{max} = 0.202 \pm 0.02$ hr^{-1}, $K_s = 0.006 \pm 0.004$, $Y_x = 1.84$

$\pm$ 0.25, and m = 0.0028 $\pm$ 0.0008. The parameter K_d was held at 0.0006 hr^{-1} across all experiments based on the literature. The half-velocity constant or affinity K_s varied the most across experimental conditions while μ_{max} values tended to be the most consistent. *E. coli* W demonstrated a significantly lower μ_{max} at low sucrose concentrations, which is again consistent with more comprehensive studies of *E. coli* W sucrose catabolism [88].

Production and growth by *A. vinelandii* (AZBB163) and *S. elongatus* (*cscB*) were investigated in Union Medium with nutrient priming (5 g/L and 0.4 g/L $NaNO_3$, respectively). Strains were cultured in 50 mL union medium and incubated at 30 °C with 200 rpm orbital shaking. Plots of absorbance and nutrient production are depicted in Fig. 2.6.

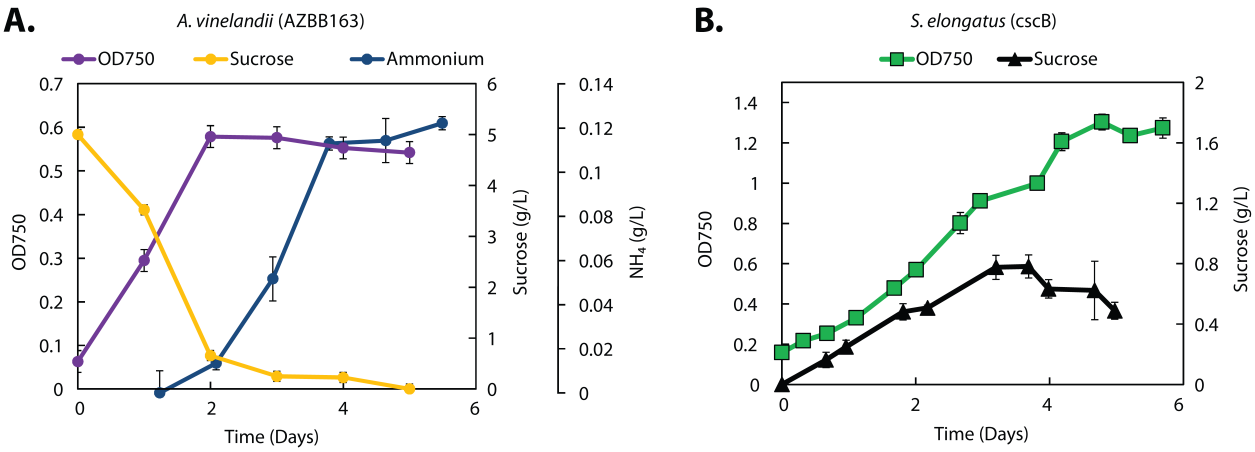


Figure 2.6: A. vinelandii and *S. elongatus* production in monoculture in union medium. (A) *A. vinelandii* production of ammonium on minimal sucrose. (B) *S. elongatus cscB* production of sucrose on ammonium.

Approximately 7 mM ammonia or 0.12 g/L NH_4 was generated by the modified *Azotobacter* strain after three days of culturing in limited media and approaches the concentration of ammonium chloride added in the production strain growth assays. Approximately 0.9 g/L sucrose was generated under a similar time frame by *S. elongatus*, which is significantly less than the concentration added to the production strain growth assays and indicates a likely bottleneck for the tripartite platform. Yet, as demonstrated in Fig. 2.5, 1.0 g/L sucrose still enabled significant bacterial growth.

2.2.4 Continuous culturing of tripartite system

A culturing strategy was developed to spatially tune strain concentrations as well as to temporally segregate the growth and production regimes of *S. elongatus* and *A. vinelandii*. Inoculation of the tripartite system without initial carbon and nitrogen priming yielded negligible growth and eventual crash of the overall system within several days as characterized by *S. elongatus* bleaching and an immediate decrease in culture absorbance (data not shown). However, excess nutrient addition undermines the goal of the tripartite consortium. Priming of the system with low level nutrient using 5 g/L sucrose and 0.4 g/L $NaNO_3$ or 25% of the limiting nutrients from Burks N-free and BG-11, respectively, served as a rational compromise. In particular, priming with $NaNO_3$ enabled selective nutrient delivery as it can only be metabolized by *S. elongatus* under aerobic conditions. Monoculture data indicated that most sucrose and $NaNO_3$ is consumed after 72 hours, while species-specific titers of NH_4^+ and sucrose approached maximum values between 48 and 72 hours. Growth and production regimes were segregated along this 72-hour mark. At 72 hours the production strain was introduced and continuous flow was initiated. While the system was designed for independent dilution rates for each strain, only the producer strain (and necessarily the entire shared culture volume) was under continuous flow.

2.2.5 Tripartite consortium with *E. coli* K-12 MG1655 *cscBKA*::*nlpA* under continuous flow conditions

A library of *E. coli* K-12 strains was generated through random integration of non-phosphotransferase sucrose catabolism by use of Tn5 transposase (Chapter 3). One strain exhibited a fast-growth phenotype with intragenic integration of the *cscBKA* genes into *nlpA*. This locus could also be readily transduced into other *E. coli* production strains (Chapter 3) and epitomizes the plug-and-play nature of the system. The system was primed with 5 g/L sucrose and 0.4 g/L $NaNO_3$ to facilitate initial growth of *A. vinelandii* and *S. elongatus* (both 0.25 OD_{750}), respectively. After this 72 hour growth regime, the production strain *E. coli*

K-12 MG1655 *cscBKA*::*nlpA* was inoculated at approximately 0.10 OD_{750} and continuous flow began at the relatively low dilution rate of 0.20 (100 mL of 500 mL chamber volume per day) such that every 5 days a chamber volume would be entirely replaced.

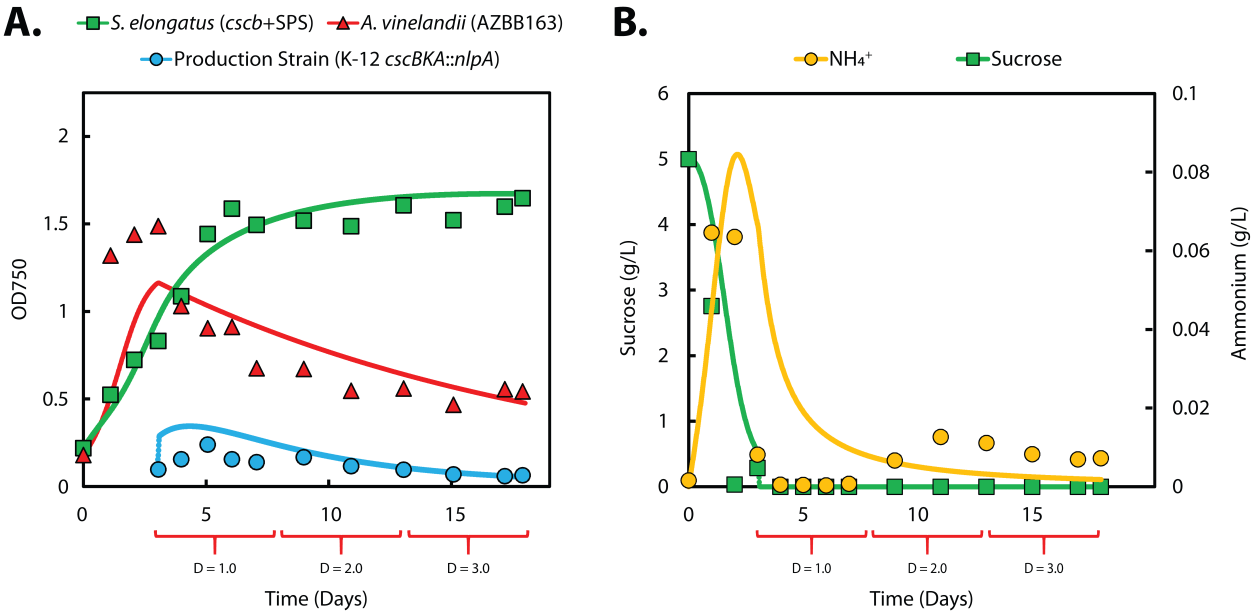


Figure 2.7: Absorbance and nutrient data for a tripartite consortium with *E. coli* K-12 MG1655 *cscBKA*::*nlpA* under continuous flow conditions. Aggregate dilutions are also listed beneath the time axis. (A) Empirical absorbance data (markers) with model prediction from monoculture data (line) (B) Empirical nutrient data (markers) with model prediction from monoculture data (line).

Fig. 2.7A depicts absorbance data for the tripartite system with growth of the *E. coli* K-12 integrant as the production strain. Experimental data shows a remarkable overshoot and decline of *A. vinelandii* during the growth regime. *S. elongatus* grew for ten days then remained at a constant absorbance for the remainder of the experiment. Producer growth appears to reach a relative maximum after 48 hours followed by a generalized decline and stabilization. This generalized decline may indicate washout of the producer strain, which suggests that the dilution rate is outpacing the producer growth rate current conditions. Nonetheless, the rate of decline of the production strain is significantly less steep than if nutrient addition were entirely absent (*i.e.*, 50% reduction in absorbance after every dilution). The corroboration between the model and experimental data as depicted in Fig. 2.7A is good

with the exception of the *A. vinelandii* overshoot between days 2 and 4. The overshoot is concerning due potential causal factors like negative interspecies interactions from resource competition or allelochemical responses. Good corroboration between empirical data and the kinetic model indicates that the monocultural growth parameters adequately describe the tripartite system.

Nutrient production and consumption within the tripartite system were also quantified and compared to model predicted values as illustrated in Fig. 2.7B. Nutrient data and model predictions follow similar trends with a steep decrease in sucrose concentration due to *A. vinelandii* uptake. Ammonium concentration immediately increases after inoculation and dramatically decreases around day 3 due to *S. elongatus* and producer strain consumption. While the model and data are similar for sucrose, they differ considerably for ammonium especially during the second and third production strain dilution. This observation supports a hypothesis that the system is extremely sucrose limited, so limited in fact that a low concentration of excess ammonium persists through subsequent dilutions.

2.2.6 Investigation of biofilm formation and possible interspecies interactions

Cocultures of *A. vinelandii* AZBB163 and a fluorescently labeled *E. coli* K-12 MG1655 *cscBKA*::*nlpA* were studied to determine whether an interspecies interaction between *E. coli* and *A. vinelandii* could be adversely affecting *A. vinelandii* and responsible for the sudden loss of absorbance. Cultures were grown in 25 mL union medium with 5 g/L sucrose in 125 mL flasks over seven days. Where applicable, *E. coli* ($\sim 10^8$ cells/mL) was added after 72 hours. Absorbance and YFP fluorescence (λ_{ex} = 475 nm, λ_{em} = 517 nm) were used as proxies for biomass accumulation Fig. 2.8.

As depicted in Fig. 2.8, the *A. vinelandii* control culture produced the characteristic overshoot and gradual decline in absorbance that was observed in continuous culturing experiments. Furthermore, coculturing with *E. coli* after 72 hours had a negligible effect on *A. vinelandii* absorbance compared to the *A. vinelandii* monoculture control. *E. coli*

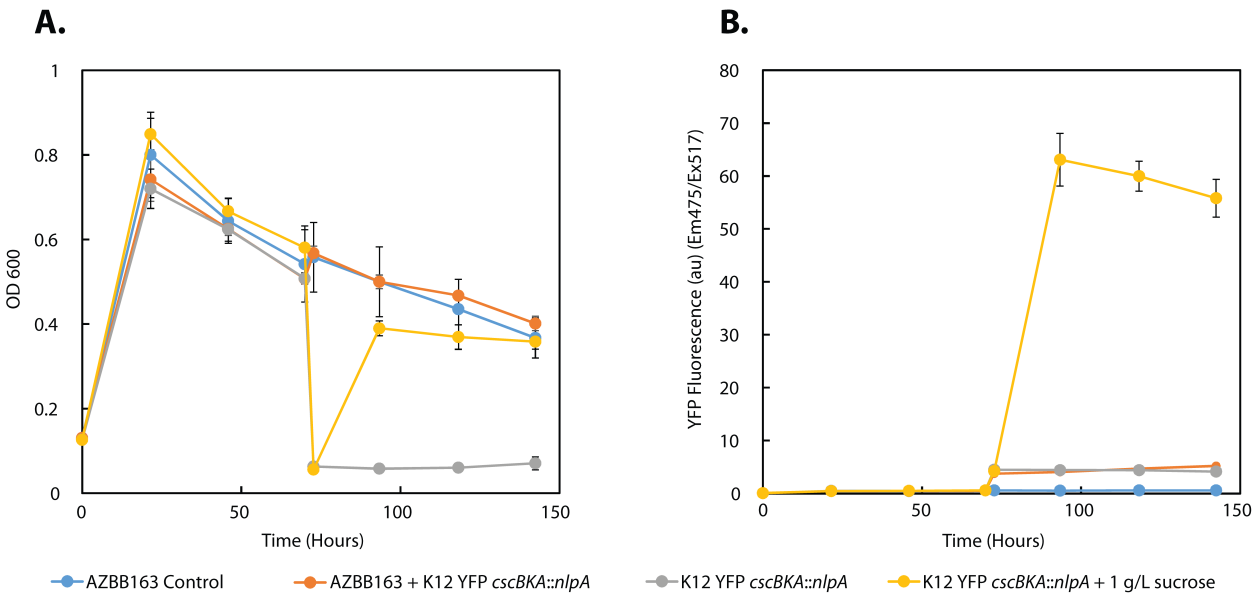


Figure 2.8: (A) Growth of *A. vinelandii* over 7 days with a fluorescently labeled *E. coli* strain capable of sucrose catabolism. Experimental conditions included a control, addition of *E. coli* to an *A. vinelandii* culture, addition of *E. coli* to *A. vinelandii* conditioned media, and addition of *E. coli* to *A. vinelandii* conditioned media with 1 g/L sucrose (n = 4). (B) Fluorescence measurements of cultures.

growth on conditioned media was only observed with the addition of 1 g/L sucrose. This also indicated that *A. vinelandii* could produce enough NH_4^+ to sustain *E. coli* growth. *E. coli* growth on conditioned media or in *A. vinelandii* cocultures was negligible. This simple experiment indicates that the overshoot and gradual decline in absorbance is intrinsic to *A. vinelandii* cultures under these conditions. The formation of PHA rich biofilms due to nutrient limiting conditions may also have contributed to the anomalous initial increase in absorbance (Appendix C).

2.2.7 Tripartite consortium with *C. glutamicum* 13032 under continuous flow

C. glutamicum was grown in the tripartite system to demonstrate bioproduction of amino acids following the same framework as *E. coli* K-12 MG1655 *cscBKA*::*nlpA*. *Corynebacterium* is cultured as a global producer of amino acids, namely L-lysine and L-glutamic acid.

As expected, *A. vinelandii* and *S. elongatus* mirrored the growth trends observed with

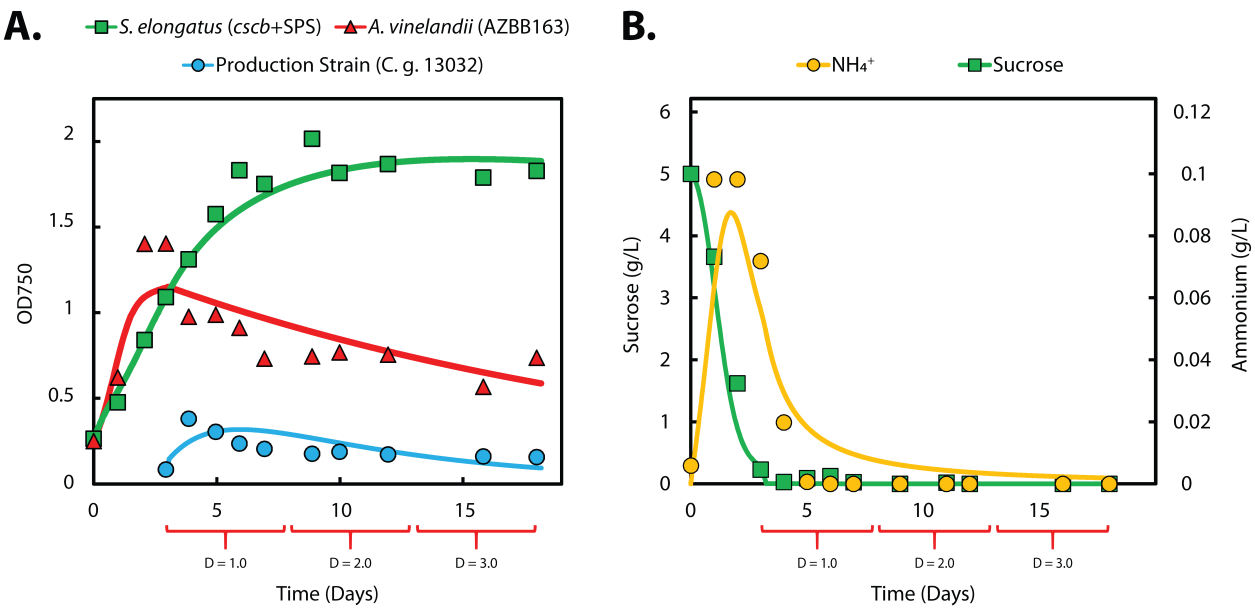


Figure 2.9: Absorbance and nutrient data for a tripartite consortium with *Corynebacterium glutamicum* 13032 under continuous flow conditions. (A) Empirical absorbance data (markers) with model prediction from monoculture data (line) (B) Empirical nutrient data (markers) with model prediction from monoculture data (line).

the K-12 integrant (Fig. 2.9A). Specific characteristics include the initial *Azotobacter* overshoot between days 2 and 4 as well a dramatic loss of absorbance before stabilizing around 0.6 OD_{750}. The production strain reached a maximum density after 24 hours, then gradually stabilized over the remaining several dilutions. The model again well describes this system with the exception of the *Azotobacter* overshoot. The stabilization of *C. glutamicum* over several dilutions is encouraging for platform viability.

In contrast to Fig. 2.7B, 2.9B shows good corroboration of nutrient quantification between empirical and that model predictions with the exception of a more severe decline in ammonium concentration after introduction of the production strain. Better corroboration may stem from a different carbon to nitrogen ratio necessary for *C. glutamicum* propagation compared to the *E. coli* K-12 derivative. It should be noted that sucrose and ammonium nutrient concentrations are largely independent of production strain additions because the system is under nutrient limited conditions and extracellular sucrose is expected to only be detectable at trace levels. Plots of ammonium and sucrose are nearly identical to con-

tinuous culturing with the K-12 derivative and defined by an immediate decline of sucrose concentration with ammonium reaching a peak concentration around day 2.

C. glutamicum naturally overproduces and secretes L-lysine due to an absence of L-lysine degrading enzymes [89]. Conversely, biotin is a major factor in regulating production of L-glutamic acid [90]. HPLC was used to quantify amino acid production and indicated low level concentrations throughout the experiment as depicted in Fig. 2.10.

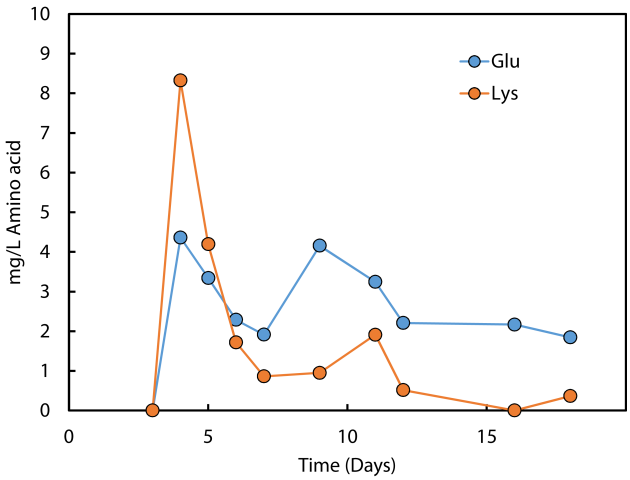


Figure 2.10: Production of L-glutamic acid (Glu) and L-lysine (Lys) by *C. glutamicum* over three dilutions.

Amino acid production in the tripartite system was at trace levels and significantly lower than the expected yield from monocultures in grown on Union Medium with nutrient additions. Low titer of either amino acid was likely due to severe nutrient limitation. As a result, excess nutrient accumulation and secretion is extremely unfavorable.

2.2.8 Tripartite consortium with *B. subtilis* (BS168) under continuous flow

Bacillus subtilis 168 was also grown in the tripartite platform to demonstrate the versatility of the system and to investigate how the growth of a strain with distinct parameters would affect culture kinetics. Specifically, *B. subtilis* had the highest absorbance yield per gram sucrose in monoculture.

S. *elongatus* and *A. vinelandii* growth was nearly identical to the previous two tripartite

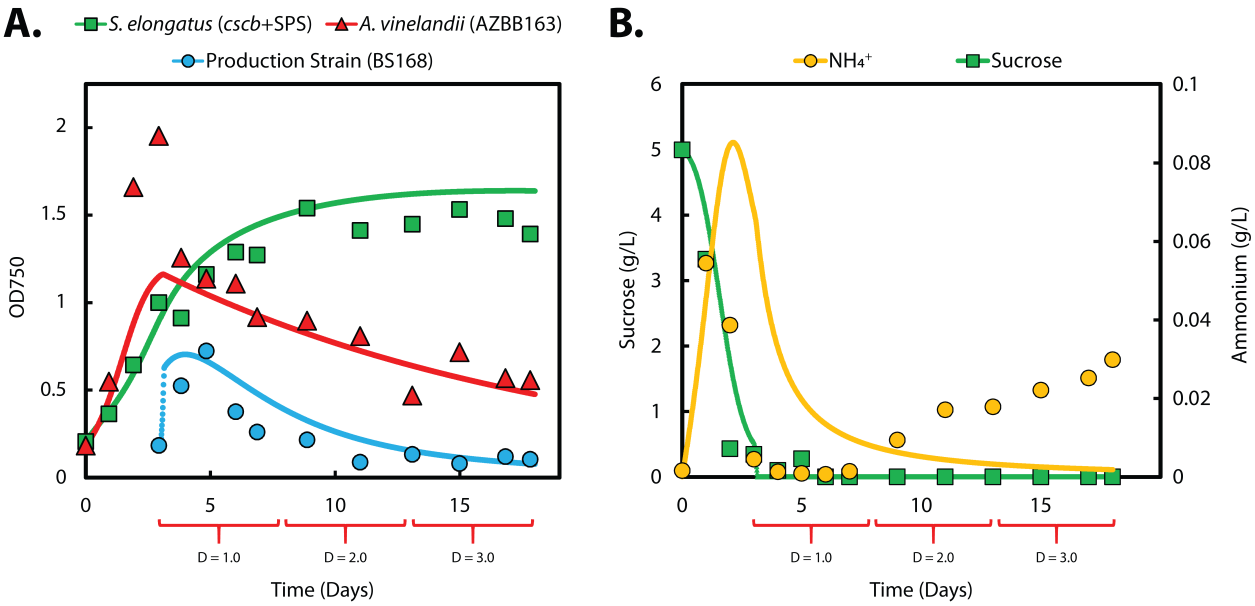


Figure 2.11: Absorbance and nutrient data for a tripartite consortium with *Bacillus subtilis* 168 under continuous flow conditions. (A) Empirical absorbance data (markers) with model prediction from monoculture data (line) (B) Empirical nutrient data (markers) with model prediction from monoculture data (line).

experimental iterations, again demonstrating significant *A. vinelandii* overshoot and a more gradual stabilization of *S. elongatus* (Fig. 2.11A). As predicted from the monoculture growth experiments, production strain growth rate is significantly faster and absorbance yields are significantly higher than either *C. glutamicum* 13032 or the *E. coli* K-12 integrant. The model data corroborates well with the experimental data for all three strains. Ammonium concentrations differ significantly between experimental values and model predicted values by both overestimating initial production and by underestimating production after several dilutions (Fig. 2.11B). The gradual increase of ammonium concentration over subsequent dilutions was similar to the profile of the *E. coli* integrant in continuous culture, which is indicative of a severely sucrose limited system. As with the previous two iterations, sucrose concentrations were near trace levels following the growth regime.

2.2.9 Computational analysis of bottlenecks in tripartite system

A correlation analysis was conducted to determine how parameter values correlated with desirable outputs. Model parameters were varied using Latin hypercube sampling (LHS) to form a uniform distribution of 10,000 parameter values between designated upper and lower boundaries. Boundaries were generally set between 5-fold increases and decreases in parameter value with respect to empirical data. Certain parameters that yielded exponential growth or production (*e.g.*, extraordinarily high sucrose yield by *S. elongatus*) had more limited boundaries. Limiting boundaries in these scenarios had no effect on the relative strength of correlation between input parameters and outputs.

The outputs were also divided into two distinct regimes. The growth regime is defined as the first 72 hours of culturing and includes the initial priming of the system prior to the introduction of the producer. The production regime is defined as the remaining time after the production strain is inoculated and continuous flow begins. Outputs were calculated by integrating interpolated functions for each output variable (biomass and nutrients) for each regime. The "overall" category represents integration of an interpolated function for the entire system (both regimes). The results of this correlation analysis were interpreted using Spearman rank correlation coefficients and are depicted as heatmaps in Fig. 2.11.

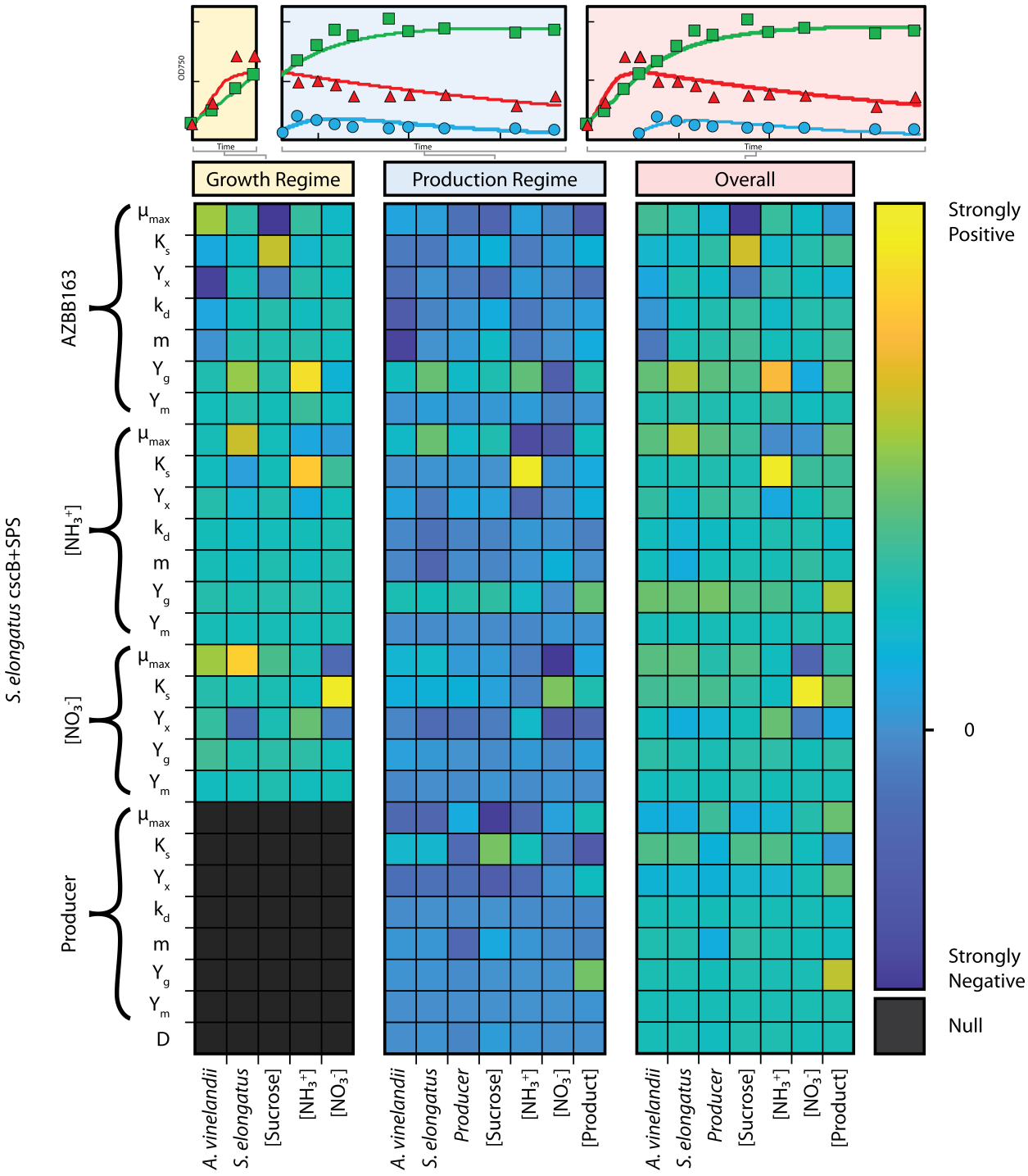


Figure 2.12: A heatmap of a tripartite model correlation analysis depicting growth and production regimes as well as the combined overall system (Spearman's correlation coefficient, $p < 0.05$ '*', $p < 0.005$ '**', $p < 0.0005$).

Many of the input parameters are strongly correlated to outputs with statistical sig-

nificance. Some correlations are self-explanatory, like correlations between respective strain μ_{max} values to concentration of strain. Likewise, the yield of a specific nutrient or product is strongly correlated with that nutrient or product concentration. Other correlations, however, are more nuanced. Fig. 2.12 indicates that nutrient concentration in the growth regime is dictated strongly by the nutrient specific growth rates, affinities, and yields. Monod affinity constants K_s correlate strongly with nutrient concentration, which is reasonable given that a higher affinity value (hr^{-1}) equates to slower consumption of a given nutrient. Likewise, the maximum growth rates and yields of biomass per unit substrate are strongly negative with respect to the given nutrient concentration due to corresponding increases amount or rate at which the nutrient is consumed. Uptake of ammonia by *S. elongatus* is less strongly correlated with *S. elongatus* concentration ($\rho = 0.04$) compared to uptake of nitrate ($\rho = 0.08$) due to the relative abundance of ammonia and nitrate during the growth regime. All production strain values are null for the growth regime as the production strain has not yet been added.

The production regime is largely dictated by the consumption of ammonia by *S. elongatus* and subsequent yield of sucrose Y_g on ammonia, which strongly correlates to increased growth of a producer strain ($\rho = 0.20$), *A. vinelandii* ($\rho = 0.17$), and overall product concentration ($\rho = 0.29$). The yield of ammonia by *A. vinelandii* also has a strong correlation on growth of all constituents in the system. Product concentration, which is arguably the most valuable system output, is most significantly correlated with increased product yield ($\rho = 0.32$) but is also correlated with the respective yields of *S. elongatus* and *A. vinelandii* ($\rho = 0.18$). Remarkably, dilution rate has only a weakly negative correlation on producer strain concentration (ρ = -0.02) and a weakly positive correlation with sucrose concentration ($\rho = 0.03$) suggesting that the system is so sucrose limited that active removal of cell concentration is less correlated with cell concentration than sucrose production.

Mathematically, the relative influence of sucrose or ammonia yield on product and producer strain concentration is identical if parameters 1) are varied over identical ranges

and 2) if producer consumption is a function of whatever nutrient is limiting for that subset of parameters. Sucrose is parametrically defined over a narrow range of values based on empirical data and, as a result, sucrose production implicitly has more influence on the overall system compared to ammonia. This assumption is appropriate based on experimental sucrose growth limitation. It is important to note that the ranges of ammonia and sucrose yield in this correlation analysis were quite different (5x range for ammonia and 1.5x for sucrose) due to the tendency of high yield sucrose simulations to singly produce exponential system growth and production. While continuous exponential growth would be ideal for this platform, it is physically unrealizable and not described by Monod kinetics.

In any case, the correlation analysis confirms that the greatest gains stem from increasing growth and yield of *S. elongatus* followed by increasing the growth and yield of *A. vinelandii*. These findings confirm that optimization of production throughput necessitates enhanced growth of *S. elongatus*. The relative magnitude of the correlation coefficients between sucrose and ammonia production also confirm that sucrose is indeed the limiting nutrient in the system.

To further visualize this correlation analysis, Spearman's rank correlation coefficients are scaled as a heatmap by inputs (rows), scaled as a heatmap by outputs (columns), as well as listed by values in Appendix F. While not necessarily providing new information, these supplementary heatmaps qualitatively show differences in how outputs are selectively affected by inputs.

2.2.10 Generation of a stable photosynthetic consortium

Single vessel experiments were conducted prior to use of the spatial separation. Although these experiments tended to washout after several weeks due to substantial loss of *A. vinelandii* and *S. elongatus* cells, one experiment generated a particularly interesting consortium. In this initial experiment, a single vessel triculture of *A. vinelandii* AZBB163, *S. elongatus cscB*, and *E. coli* W was ultimately contaminated due to backflow of the peri-

staltic pumps. Rather than terminating the experiment, the vessel was sealed and culturing was continued with shaking at 200 rpm and incubation at 30 °C in a closed 500 mL Pyrex bottle and 100 mL headspace. After 100 days of culturing, culture aliquots were spotted on respective minimal media (*i.e.*, BG-11 with 12.5 μg/mL chloramphenicol for *S. elongatus* SPS *cscB*, Burks N-free with 3 μg/mL kanamycin for *A. vinelandii* AZBB163, and LB without antibiotic to screen for heterotrophs). Remarkably, viable colonies of *S. elongatus* grew and were isolated from BG-11 plates. Similarly, a variety of heterotrophic strains were morphologically identified and restruck from initial growth on LB plates.

The 100-day culture is a fascinating albeit tangential discovery from the tripartite system in that its constituents could survive without additional nutrients for an extended period. Elucidation of culture composition could point towards potential synergistic interspecial relationships, thus the community was analyzed by sequencing of 16S fourth hypervariable (V4) region shown in Fig. 2.13.

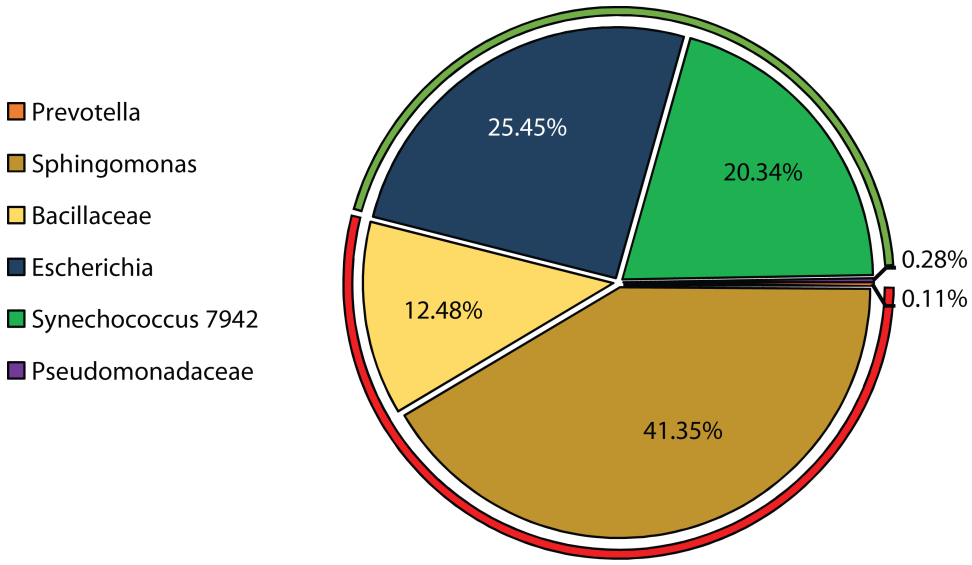


Figure 2.13: Community composition of the 100-day culture. Sequencing data shows a range of taxonomical accuracy based on the amplified region.

Bacteria of the genus *Escherichia* or *Synechococcus* composed 49% of the culture and most likely originated from the initial strains of the tripartite system, confirmed by growth

of colonies from culture samples on selective BG-11 plates. Another major constituent was of the family *Bacillaceae* (12.48%), which consists of gram-positive, heterotrophic endospore forming bacteria [91]. Unfortunately, the resolution of the 16S region makes further elucidation difficult, though its presence is almost certainly due to the ubiquity of *Bacillaceae* in the environment. *Azotobacter* is the likely candidate within the family Pseudomonadaceae with very low overall composition (0.28%). Interestingly, most of the culture was composed of *Sphingomonas*, a gram-negative aerobic bacterium characterized by their glycosphingolipid membrane that produces in bright yellow colony formation [92]. Sphingomonads are widely distributed among soil, freshwater, and marine environments and can thrive under low nutrient conditions and substantial investigation into their human health effects has been conducted due to commonality in clinical settings [93]. Uniquely, sphingomonads also possess pathways for organic pollutant degradation including dioxins, chlorinated phenols, aromatic hydrocarbons, and even microcystin-RR [94,95]. They have been classified as candidates for bioremediation and extensively studied in PGPR studies due to their specialized pathways relative to other soil bacteria [95]. The presence of *Bacillaceae* and *Sphingomonas* may be rationalized by their ubiquity in the environment and aptitude for survival under low nutrient conditions. It is unclear specifically how or why a species of *Sphingomonas* dominated this stable consortium, but the fact that it propagated from a background contaminant to a dominant member of the culture without killing the other constituents suggests it could have some beneficial impact (nutrient cross-feeding, biodegradation, etc.) that facilitates culture survival. A simple set of coculture experiments with isolated *S. elongatus* and *Sphingomonas* strains was attempted, though each crashed after a week. Crashes were perhaps due to poor initial culture conditions or influence by other consortia constituents (*i.e.*, *E. coli* and *Bacillaceae*). This stable photosynthetic consortium is fascinating and could point toward *Sphingomonas* as a candidate for future consortia.

2.3 Conclusion

The tripartite platform demonstrates the propensity of *S. elongatus* and *A. vinelandii* to support production strain candidates and, in one case, yield low-level bioproduct titers. While the tripartite platform is not yet competitive with direct microbial cultivation on added nitrogen and fermentable carbon substrates at its current stage, this prototype has established a baseline for platform viability. The accompanying kinetic model has specifically elucidated major systematic bottlenecks and opportunities for strain optimization. Indeed, integration of recent advances in pathway optimization or yield of either *A. vinelandii* or *S. elongatus* could have important implications for further optimization of the tripartite system.

Improvements to the physical cultivation apparatus, building resilience to heterotrophic contamination, and integration of optimized strains would facilitate higher overall production, thereby increasing the relative competitiveness of the platform with petrochemical derived products.

2.4 Materials and methods

2.4.1 Strain background

S. elongatus

The *cscB* + SPS strain of *S. elongatus* was used in this work over the *cscB* variant to simultaneously maintain high growth rate and high sucrose export. See Abramson et al. for more details. The strain was kept in replicate cryostocks provided by D. Ducat. Cryostocks were revived on BG-11 agar plates with 12.5 μg/mL chloramphenicol for selection. Prior to inoculation into union media, precultures were prepared in liquid BG-11 media with 1 g/L HEPES and without antibiotic. Precultures were meticulously screened by spotting on LB agar plates without antibiotic to screen for contamination.

A. vinelandii

The *A. vinelandii* DJ derivative AZBB163 was used in this study. The strain maintains an in-frame *kanR* insertion intragenically in the *nifL* gene as well as a point-mutation that enabled continued growth and nitrogenase production as depicted in Fig. 2.13 [74]. More details regarding strain modification are available in Barney et al.

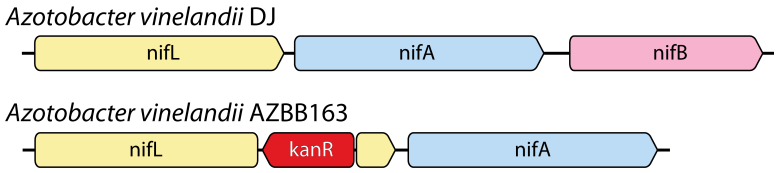


Figure 2.14: Illustration of *kanR* insertion into the *nifL* gene of *Azotobacter vinelandii* DJ to construct the AZBB163 strain. Adapted from Ref [74].

In a similar fashion to *S. elongatus*, *A. vinelandii* was stored in cryostock replicates. Agar plates were obtained from Brett Barney. Individual colonies were struck on Burks N-free agar plates with 3 μg/mL kanamycin sulfate. Colonies were also screened for the target phenotype by ammonium ion selective electrode. Cryostocks were prepared from single colonies and stored in 900 μL 10x Burk's N-free phosphate solution with 100 μL DMSO. AZBB163 was revived on Burk's N-free agar plates with 3 μg/mL kanamycin sulfate for selection and 20 g/L sucrose as carbon substrate. All precultures were screened for ammonium production and aliquoted onto LB plates without antibiotic prior to inoculation into the tripartite system. *A. vinelandii* grows poorly on LB such that heterotroph contaminants are readily discernible. In rare instances precultures produced low or trace ammonium levels which generally indicated heterotroph contamination.

Production Strains

Candidates for production in the tripartite consortium with *A. vinelandii* and *S. elongatus* were selected based on native capacity for growth on sucrose and ammonia and absence of allelochemical production.

2.4.2 Monoculture model of growth and production

Algorithms for microbial growth generally follow density-limited Verhulst kinetics or substrate-limited Monod kinetics depending on physical culture constraints [96, 97]. Both kinetic models may be readily solved using empirical data and initial conditions. Verhulst kinetics follow a simple logistic function:

$$\frac{dX}{dt} = rX\left(1 - \frac{N}{K}\right) \tag{13}$$

Where X is the number of cells or an approximation of cell density, r is the specific growth rate of the culture, and K is the carrying capacity or upper limit of cell density as defined by the culture system (*e.g.*, absorbance). The logistic equation accurately describes many microbial growth systems that are density dependent, however, the carrying capacity changes dramatically based on culture conditions and geometry. Conversely, the Monod equation describes a system that is limited by a single nutrient and follow Michaelis-Menten kinetics:

$$\frac{dX}{dt} = \mu_{max} * X * \left(\frac{S}{K_s + S}\right) \tag{14}$$

In the Monod equation, S represents the limiting substrate (g/L), K_s represents the half-velocity constant or substrate affinity (g/L), and $\mu_{\max}$ is the maximum specific growth rate (hr^{-1}). Unlike carrying capacity, affinity is a biological component related to substrate uptake and transport kinetics. Likewise, Monod kinetics describe nutrient depletion as the mechanism for decreased growth rate as opposed to the arbitrary carrying capacity in Verhulst kinetics. The growth rate or the change in cell count over time is also commonly referred to as the specific growth rate and denoted μ (hr^{-1}):

$$\mu = \mu_{max} * \left(\frac{S}{K_s + S}\right) \tag{15}$$

In the tripartite system, nutrient concentrations are necessarily low due to the energetic burdens of photosynthesis and nitrogen fixation, which result in relatively low nutrient fluxes of sucrose and ammonia. As a result, Monod kinetics are ideal for describing the system. Bacterial substrate consumption may also be allocated into two more distinct categories: growth associated consumption and non-growth or maintenance associated consumption. As with growth associated consumption (Equation 15) non-growth associated consumption can also be derived [33]:

$$K_{max} * \left(\frac{S}{K_s + S} \right) = Y_1 * \mu + m \tag{16}$$

Here, Y_1 is the yield of cells on the limiting substrate ($g_{substrate}/g_{cells}$). The term K_{max} is the maximum specific uptake rate and defined as a sum of the non-growth associated substrate consumption "m" and the substrate-biomass yield coefficient ($g_{substrate}/g_{cells}/hr$), which is assumed constant:

$$K_{max} = Y_1 * \mu_{max} + m \tag{17}$$

Equation 16 may be rearranged in terms of growth rate μ where maintenance has a negative effect on overall growth rate due to split substrate allocation:

$$\mu = \frac{1}{Y_1} * \left(\frac{K_{max} * S}{K_s + S} - m \right) \tag{18}$$

Combining these equations yields an expression of growth rate that includes contributions to both growth and non-growth associated substrate consumption:

$$\mu = \left(\left(\mu_{max} + \frac{m}{Y_1} \right) * \left(\frac{S}{K_s + S} \right) - \frac{m}{Y_1} \right) \tag{19}$$

Equation 19 may also be adjusted to better describe bacterial growth by incorporating death rate k_d (hr^{-1}) and non-growth associated substrate consumption:

$$\frac{dX}{dt} = (\mu - k_d) * X \tag{20}$$

Substrate consumption can be derived in an analogous fashion:

$$\frac{dS}{dt} = -\left(\frac{K_{max} * S}{K_s + S}\right) * X = -(Y_1 * \mu_{max} + m) * \left(\frac{S}{K_s + S}\right) * X \tag{21}$$

Rearranged for simplicity:

$$\frac{dS}{dt} = -\left(Y_1 * \mu_{max} * \left(\frac{S}{K_s + S}\right) + m * \left(\frac{S}{K_s + S}\right)\right) * X \tag{22}$$

Finally, product yield is derived in an analogous manner such with growth-associated production and non-growth associated production using Y_2, which represents yield of product ($g_{product}/g_{cells}$). As written, it assumes that the product is not inhibitory:

$$\frac{dP}{dt} = Y_1 * \mu * X \tag{23}$$

$$\frac{dP}{dt} = \left(Y^{growth} * Y_1 * \mu_{max} * \left(\frac{S}{K_s + S}\right) + Y^{maint} * m * \left(\frac{S}{K_s + S}\right)\right) * X \tag{24}$$

Equation 24 is a modification of Equation 22 with the addition of two yield terms, Y^{growth} and Y^{maint}. The terms are both in units $g_{product}/g_{cells}$ but account for differences between growth and maintenance (*i.e.*, Y_1 and m, respectively). Collectively, these equations for growth and non-growth associated production and consumption well describe all nutrient fluxes within a nutrient confined microbial system.

2.4.3 Monoculture model under continuous flow

The kinetics of chemostat or continuous flow cultures are dictated by the feed substrate concentration and flow rate. Flowrate is tuned to the growth rate of the microbial species

such that continuous flow culturing tends toward substrate-limited kinetics. The dilution rate in continuous flow cultures may be easily calculated:

$$D = F/V \tag{25}$$

Thus, growth in a continuous flow system is described collectively by building upon Equation 20:

$$\frac{dX}{dt} = (\mu - k_d - D) * X \tag{26}$$

Where F is the flow rate in and out of a vessel with volume V. The dilution rate of a chemostat must not be higher than the growth rate of the bacteria or else washout (gradual loss of culture biomass) will occur. The ratio of flowrate in and out of the culture vessel may also be tuned to maintain a threshold of biomass in the culture. The limiting substrate balance for a continuous culture is therefore defined as follows:

$$\frac{dS}{dt} = D * (C - S) - \mu * X \tag{27}$$

Where C is the concentration of substrate in the feed and all other variables have been previously defined. Lastly product concentration is the sum of growth and non-growth associated production as functions of cell concentration against the dilution or removal of product from the system:

$$\frac{dP}{dt} = -D * P + \left(Y^{growth} * Y_1 * \mu_{max} \left(\frac{S}{K_s + S} \right) + Y^{maint} * m * \left(\frac{S}{K_s + S} \right) \right) * X \tag{28}$$

A consolidated description of the terms used in modeling is available in Table 2.1.

Table 2.1: List of terms, definitions, and units of constants within the series of ODEs for tripartite culture modeling.

Term	Definition	Units
μ	Growth rate	hr^{-1}
μ_{max}	Maximum specific growth rate	hr^{-1}
m	Maintenance coefficient	$g_{substrate}/g_{biomass}/hr$
K_s	Affinity coefficient	$g_{biomass}/L$
X	Cells/biomass	$g_{biomass}$
P	Product	$g_{product}$
S	Substrate	$g_{substrate}$
D	Dilution rate	hr^{-1}
$Y_{s/x}$	Yield of cells per substrate	$g_{substrate}/g_{biomass}$
K_d	Cell death rate	hr^{-1}

2.4.4 Model of tripartite culture under continuous flow

Monocultures were parametrized in MATLAB using the Levenberg-Marquardt algorithm for the least squares curve fit of the modified kinetic equations [29, 33, 98]. Conditions were set such that growth was substrate rather than density dependent. Extrapolation of monoculture growth parameters to polyculture growth was conducted with seven assumptions.

1. All consortium members are substrate limited such that they can be well explained using Monod kinetics.

2. Producer strain is not dual substrate limited and primarily limited by sucrose.

3. Utilization of $NaNO_3$ and NH_4Cl by *S. elongatus* is homologous under nitrogen limited

conditions [99].

4. Producer consumption rate of NO_3 is negligible or nonexistent due to highly aerobic conditions.

5. No inhibition of product concentration or interspecies interactions (positive/negative) between constitutive species.

6. The culture is uniformly mixed and pulse addition of feed and removal of culture volume is a good approximation for continuous flow.

7. Diffusion through cellulose acetate membranes is not limiting.

Multispecies culturing was modeled using monoculture data for baseline parameters and then analyzed accounting for multisubstrate consumption for *S. elongatus cscB* + SPS. The equations account for changes in dilution rate such that the dilution of individual species can be controlled. Dilution of nutrient is then the sum of the combined dilution rates assuming the bottles are uniformly mixed and diffusion through cellulose acetate membranes is not limiting. Modeling of *A. vinelandii* growth is assumed sucrose limited as sufficient N_2 was bubbled into the system and the media (Fig. 2.3) contains an excess of phosphorus:

$$\mu_{Av} = \left(\left(\mu_{max,Av} + \frac{m_{Av}}{Y_{S_{suc}/X_{Av}}}\right) * \left(\frac{S_{suc}}{K_{s,Av} + S_{suc}}\right)\right) - \frac{m_{Av}}{Y_{S_{suc}/X_{Av}}} \tag{29}$$

$$\frac{dX_{Av}}{dt} = (\mu_{Av} - K_{d,Av} - D_1) * X_{Av} \tag{30}$$

Similarly, growth limitations of producer strains were found to be primarily sucrose limited with a C/N ratio of 10 (*i.e.*, 10x more carbon substrate than nitrogen substrate) before ammonia became growth limiting on a per gram basis:

$$\mu_{P,suc} = \left(\left(\mu_{max,P} + \frac{m_P}{Y_{S_{suc}/X_P}}\right) * \left(\frac{S_{suc}}{K_{s,P} + S_{suc}}\right)\right) - \frac{m_P}{Y_{S_{suc}/X_P}} \tag{31}$$

$$\frac{dX_p}{dt} = (\mu_P - K_{d,P} - D_2) * X_p \tag{32}$$

Unlike *A. vinelandii* and the selected producer strains, *S. elongatus* growth is more nuanced due to initial priming with sodium nitrate. *S. elongatus* is capable of dual-substrate uptake. Control experiments in monoculture indicated slightly higher growth rate on ammonia than nitrate such that the overall growth rate of *S. elongatus* is defined as:

$$\begin{aligned}\mu_{Se} = &\left(\mu_{max,Se,NH_4} + \left(\frac{m_{Se,NH_4}}{Y_{S_{NH_4}/X_{Se}}}\right) + \left(\frac{m_{Se,NO_3}}{Y_{S_{NO_3}/X_{Se}}}\right)\right) \\ &* \left(\frac{\left(\frac{\mu_{max,Se,NH_4}}{K_{s,Se,NH_4}}\right) * S_{NH_4} + \left(\frac{\mu_{max,Se,NO_3}}{K_{s,Se,NO_3}}\right) * S_{NO_3}}{\mu_{max,Se,NH_4} + \left(\left(\frac{\mu_{max,Se,NH_4}}{K_{s,Se,NH_4}}\right) * S_{NH_4} + \left(\frac{\mu_{max,Se,NO_3}}{K_{s,Se,NO_3}}\right) * S_{NO_3}\right)}\right) \\ &- \frac{m_{Se,NH_4}}{Y_{S_{NH_4}}/X_{Se}} - \frac{m_{Se,NO_3}}{Y_{S_{NO_3}}/X_{Se}}\end{aligned} \tag{33}$$

$$\frac{dX_{se}}{dt} = (\mu_{Se} - K_{d,Se} - D_3) * X_{Se} \tag{34}$$

These three ODEs dictate the overall cell balances for the tripartite continuous culturing platform. Single vessel culturing necessitates a single dilution rate such that all dilution rates are equal:

$$D_1 = D_2 \ldots = D_i \tag{35}$$

Partitioning or spatial isolation of constituent species enables dilution tuning such that the collective dilution rate of the media is as follows:

$$D_4 = \epsilon * \sum_{i=1}^{3} D_i \tag{36}$$

Here, D_4 is the dilution rate of the substrates and product within the culturing system. ϵ is the volume of the diluted partitions over the total system volume. This is critical as nutrient and product concentration are affected by each partitioned dilution whereas

partition cell concentration is affected only by the partition dilution rate. The most complex substrate balance in the tripartite system is for sucrose, which is a product of *S. elongatus* as well as a substrate for the other two constituent species. Furthermore, *S. elongatus* has growth and maintenance associated production on both nitrate and ammonia, lending to four production expressions. *A. vinelandii* and the producer strains have two terms each for growth and non-growth associated consumption. The balance also includes removal due to continuous flow:

$$\begin{aligned}
&\frac{dS_{suc}}{dt} \\
&= D_4 * (C_{suc} - S_{suc}) \\
&\quad + \left(Y^{growth}_{S_{suc}/S_{NH_4}} * Y_{S_{NH_4}/X_{Se}} * \left(\mu_{max,Se,NH_4} * \left(\frac{S_{NH_4}}{K_{s,Se,NH_4} + S_{NH_4}} \right) \right) * X_{Se} \right) \\
&\quad + \left(\left(Y^{maint}_{S_{suc}/S_{NH_4}} * m_{Se,NH_4} * \left(\frac{S_{NH_4}}{K_{s,Se,NH_4} + S_{NH_4}} \right) \right) * X_{Se} \right) \\
&\quad - \left(\left(Y_{S_{suc}/X_{Av}} * \mu_{max,Av} * \left(\frac{S_{suc}}{K_{s,Av} + S_{suc}} \right) \right) * X_{Av} \right) - \left(m_{Av,suc} * \left(\frac{S_{suc}}{K_{s,Av} + S_{suc}} \right) \right) * X_{Av} \\
&\quad - \left(\left(Y_{S_{suc}/X_P} * \mu_{max,P} * \left(\frac{S_{suc}}{K_{s,P} + S_{suc}} \right) \right) * X_{Se} \right) - \left(m_{P,suc} * \left(\frac{S_{suc}}{P_{Ks} + S_{suc}} \right) \right) * X_P \\
&\quad + \left(Y^{growth}_{S_{suc}/S_{NO_3}} * Y_{S_{NO_3}/X_{Se}} * \left(\mu_{max,Se,NO_3} * \left(\frac{X_{NO_3}}{K_{s,Se,NO_3} + X_{NO_3}} \right) \right) * X_{Se} \right) \\
&\quad + \left(Y^{maint}_{S_{suc}/S_{NO_3}} * m_{Se,NO_3} * \left(\frac{X_{NO_3}}{K_{s,Se,NO_3} + X_{NO_3}} \right) \right) * X_{Se}
\end{aligned} \tag{37}$$

The ammonia balance follows only the consumption by *S. elongatus* and the producer strain:

$$\begin{aligned}
\frac{dS_{NH_4}}{dt} &= (D_4) * (C_{NH_4} - S_{NH_4}) \\
&\quad + \left(Y^{growth}_{S_{NH_4}/S_{suc}} * Y_{S_{NH_4}/X_{Av}} * \left(\mu_{max,Av} \left(\frac{S_{suc}}{K_{s,Av} + S_{suc}} \right) \right) * X_{Av} \right) \\
&\quad + \left(Y^{maint}_{S_{NH_4}/S_{suc}} * m_{Av,suc} * \left(\frac{S_{suc}}{K_{s,Av} + S_{suc}} \right) * X_{Av} \right) \\
&\quad - \left(Y_{S_{NH_4}/X_{Se}} * \mu_{max,Se,NH_4} * \left(\left(\frac{S_{NH_4}}{K_{s,Se,NH_4} + S_{NH_4}} \right) \right) * X_{Se} \right) \\
&\quad - \left(m_{Se,NH_4} * \left(\frac{S_{NH_4}}{(K_{s,Se,NH_4} + S_{NH_4}} \right) * X_{Se} \right)
\end{aligned} \tag{38}$$

The nitrate balance is simply a balance of *S. elongatus* nitrate uptake and removal due to dilution.

$$\begin{aligned}\frac{dS_{NO_3}}{dt} &= (D_4) * (C_{NO_3} - S_{NO_3}) \\ &- \left(\left(Y_{S_{NO_3}/X_{Se}} * \mu_{max,Se,NO_3} * \left(\frac{S_{NO_3}}{K_{s,Se,NO_3} + S_{NO_3}}\right)\right) * X_{Se}\right) \\ &- \left(m_{Se,NO_3} * \left(\frac{S_{NO_3}}{K_{s,Se,NO_3} + S_{NO_3}}\right)\right) * X_{Se}\end{aligned} \tag{39}$$

Finally, the product balance is the growth and maintenance associated production by the producer against the product being removed by continuous flow.

$$\begin{aligned}\frac{dP}{dt} &= -(D_4 * P) + \left(Y^{growth}_{S_{suc/P}} * Y_{S_{suc}/P} * \left(\mu_{max,P} * \left(\frac{S_{suc}}{K_{s,P} + S_{suc}}\right)\right) * X_P\right) \\ &+ \left(Y_{S_{suc/P}} * m_P * \left(\frac{S_{suc}}{K_{s,P} + S_{suc}}\right) * X_P\right)\end{aligned} \tag{40}$$

Collectively, this system of seven ODEs describes the entirety of the tripartite system and may be adjusted for either single vessel culturing or partitioned culturing.

2.4.5 Monoculture growth parametrization

Kinetic growth data for producer species were collected by inoculating 200 μL of Union Medium in 96 well plates at varied nutrient concentrations using a spectrophotometer (Molecular Devices, Versamax) at 750 nm over 50 hours (Fig. 2.6). Data from growth below OD_{750} of 0.75 were used to ensure accurate fits. Flask experiments were also conducted using 50 mL Union Medium in 125 mL flasks over a similar period. Production was measured by HPLC (Agilent) for amino acid concentration where applicable. It was observed that all producer species required approximately 10:1 carbon to nitrogen such that single-nutrient limiting growth kinetics were assumed [99, 100]. Likewise, it was assumed there were no inhibitory effects of substrate or production target on the growth of each species.

2.4.6 Culture apparatus

Bacterial cultures were housed in a custom three membrane system, which includes three 500 mL bottles modified with 45 mm NW flanges and a middle chamber with three 45 mm NW flanges (fabricated by Adam & Chittenden Scientific Glass). Bottles were equipped with knuckle clamps to interface each chamber. The knuckle clamps compress a rubber O-ring between two concentric plastic rings and the glass vessel interfaces. The membrane was then held between one glass interface and the rubber O-ring. A cellulose acetate membrane with 47 mm diameter and 0.22 μm pore size (Millipore Sigma) was used as it demonstrated the best nutrient diffusion. Stirring was necessary because orbital shaking led to membrane rupture. Initially, the bottles and middle chamber of the tripartite culture apparatus (Fig. B.3) were filled with 500 mL and approximately 250 mL of Union Medium, respectively, for a total system volume of 1750 mL.

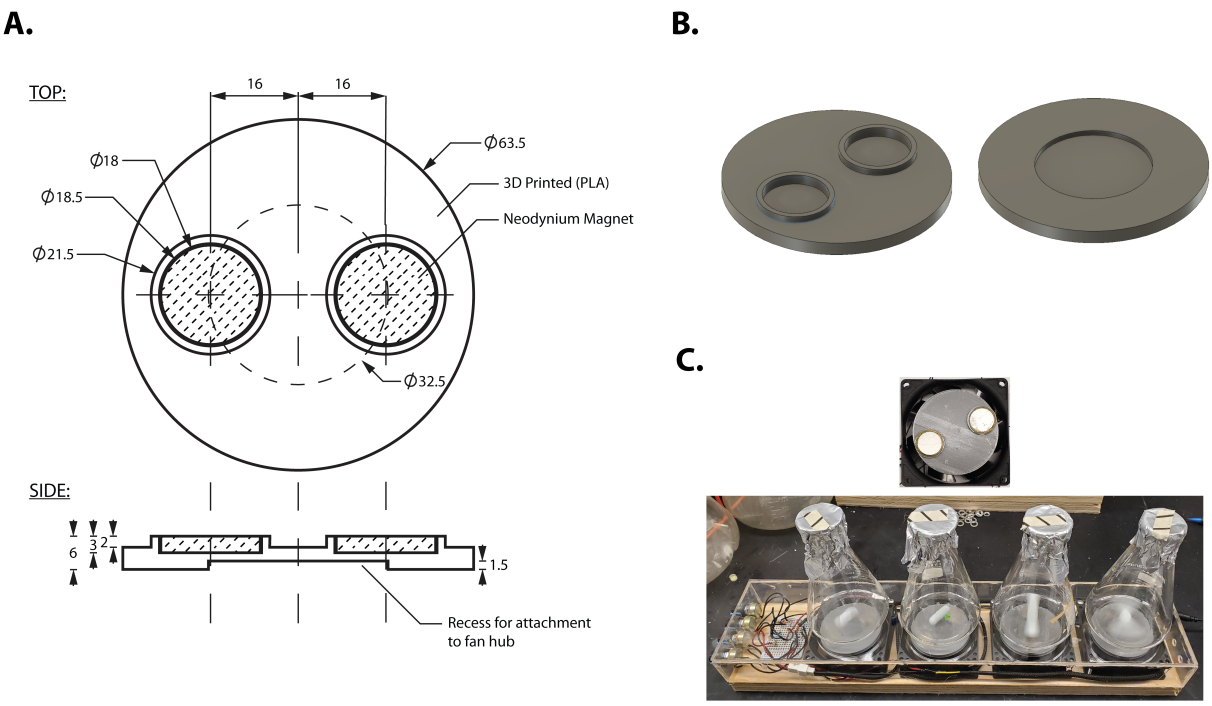


Figure 2.15: 3D printed PLA disks for custom stir plates. (A) Computer-aided design (CAD) of PLA discs for affixing neodymium magnets. The design was made by Dr. Martin P. de Beer. (B) Rendered image of discs from CAD software. (C) Image of PLA disc with epoxied magnets and a custom stir plate in action.

Custom stir plates were constructed due to the unique geometry of the glassware using neodymium magnets, a custom 3-D printed housing, and recycled computer case fans. The stir platform was engineered using four 80 mm computer case cooling fans (2500 RPM) with linear potentiometers (10 kΩ) and a power supply (5 V, 2 A). PLA discs were 3D printed using a PRUSA i3 MKA with inserts for two 0.7" neodymium disc magnets. Two neodymium disc magnets were affixed to each PLA disc with opposite polarity using epoxy. Four custom stir plates were configured and arranged beneath each glass chamber, enabling approximately 1 cm gap between the magnets and the stir bar (Fig. 2.15). The apparatus was maintained in a 5% CO_2 chamber with 60 $\mu mol_{photons}/m^2/s$ PAR (15W, Sylvania Grolux) and held at a culture temperature of 30 °C.

A. vinelandii was unable to grow without significant aeration (data not shown). Magnetic stirring is highly laminar and maintains low gas transfer with the headspace. Gas was bubbled into the system at a rate of 1.25 L/min/chamber to supply *S. elongatus* with amble CO_2 and for nonlaminar mixing. The bubbled gas consisted of 5% bone dry CO_2 in air and was sequentially mixed and hydrated in a 4 L beaker of DI H_2O, a 1 L bottle of 500 mL DI H_2O, and an empty 1 L bottle prior to bubbling into chambers. Air was filtered using HEPA-Vent Filters (GE Healthcare Whatman) ideal for gas filtration. Septa lids were pierced with modified glass pipets and fit to silicone tubing for either gas or peristaltic flow. Lids were covered with cotton swabs and foil prior to sterilization. System outflow was collected and measured to compare experimental and expected dilution rate.

2.4.7 Microcontroller set-up

Continuous flow was controlled by pulse modulation of peristaltic pumps using a microcontroller (Adafruit METRO 328) and a soldered motorshield (Adafruit Motorshield). Each microcontroller drove two peristaltic liquid pumps (12 V, 300 mA) in tandem. Peristaltic pumps were selected due to their simplicity and high accuracy for low volume aliquots. The pumps operate simultaneously such that feed is added to the system as culture is removed.

The pumps drive flow through 2 mm inner diameter silicone tubing fit with polypropylene tubing clamps. Both the tubing and clamps could be autoclaved, which facilitated removal of the culture apparatus in and out of the culture chamber without contamination. Pump pairs were tuned to a flowrate 0.5 mL/s over a 5 second pulse. Pulse width modulation within the Arduino Integrated Development Environment facilitated tuning the dilution rates of the system and worked well with the volume of the apparatus.

Photosynthetic organisms become increasingly light-limited at high optical densities. Irradiance is also impacted by vessel geometry. Light availability in a cylindrical glass chamber, for example, is maximized by radial irradiation as opposed to axial irradiation due to the increased surface area available. A custom lighting system was designed to better compensate for changing form factors and vessel geometry using variable LEDs from Adafruit Industries (NeoPixel 5050 RGBW LEDs w/ Integrated Drivers, Cool White). Flasks were cultured using NeoPixel rings (16 or 24 LEDs depending on flask volumes) for radial irradiance. Each bottle including the 500 mL modified glass bottle partition in the membrane system was outfitted with a 24 LED NeoPixel ring as well as a 30 LED NeoPixel strip. Light intensity was tuned using an Adafruit Metro microcontroller. The spectrum of LED NeoPixel white light is not specific for photosynthetic organisms, but still provides moderate gains in overall growth rate. As a result, bottle cultures were exposed to axial light sources beneath and above the vessel as well as radial light sources.

LED light intensity was tuned to maximize *S. elongatus* growth rate (approximately 60 μmol photons/m^2/s from above and 150 μmol photons/m^2/s from below). It is feasible that growth could be improved further by coding light intensity to mirror Beer-Lambert optical density during culture growth. It is also hypothesized that the growth rate could be optimized by using LEDs tuned specifically to the absorbance of photosynthetic pigments of *S. elongatus*, thereby increasing the amount of useful light (PAR) available to the photosynthetic species and reducing the amount of overall energy demand of the LEDs. While PAR is a standard for quantifying the amount of light available to photosynthetic

organisms, it should be noted that many PAR sensors do not discriminate between different wavelengths in the PAR spectrum or the quality of the radiation received. Optimization of these LED systems is an opportunity for increased growth of photosynthetic species and increased overall efficiency both of this tripartite system and feasibly for any other system driving photosynthetic production.

2.4.8 Initial conditions for tripartite culture

Spontaneous growth in Skeletal Union Medium (*i.e.*, medium without added nitrogen and carbon nutrients) from interspecies production was not observed and, without added priming, cultures crashed within days. Precultures for tripartite experiments were passaged from respective minimal media into the Union Medium with specific nutrient requirements and without inducer (*e.g.*, Union Medium with 0.4 $NaNO_3$ for *S. elongatus*, 5 g/L sucrose for *A. vinelandii* , and 0.4 g/L NH_4Cl and 5 g/L sucrose for the producer strain). The nitrogen source, $NaNO_3$, is not utilizable by *A. vinelandii*. Likewise, sucrose is not consumed by *S. elongatus*. By initially passaging cultures in these media we aimed to maintain physiological consistency between precultures and polycultures. For example, *A. vinelandii* overproduces PHB to store carbon under conditions with an excess of carbon and low nitrogen or phosphorus availability. Passaging precultures avoids intracellular carbon storage as PHB by *A. vinelandii* that could distort growth in polycultures. This initial priming has been reported in many similar co-cultures [101]. *A. vinelandii* and *S. elongatus* were inoculated at approximately 0.25 OD_{750}. The producer was inoculated after 72 hours after both nutrient sources have stabilized at much lower levels and the cultures are in a homeostatic production mode. Continuous flow of producer began immediately following inoculation.

2.4.9 Characterization of growth and production of the tripartite system

Flow cytometry (FCM) was initially chosen for community composition analysis during single vessel cultivation. While *S. elongatus* is easily detectable using chlorophyll-*a* (chl-*a*)

fluorescence, *A. vinelandii* overproduces siderophores under low iron availability that pose a challenge when using fluorescently labeled strains that have approximate emission or excitation wavelengths to certain siderophores. Siderophores are iron(III)-chelating molecules that facilitate iron acquisition in environments with low iron bioavailability. Iron is involved in many essential metabolic processes including, among others, DNA synthesis, electron transport, enzymatic cofactors, oxygen transport in respiration, and the nitrogen cycle [102]. Siderophore production is induced by intracellular iron deficiency and enables solubilization of iron in the environment and may be divided into three classes: catecholates (alkaline/neutral pH), hydroxymates (acidic pH), and mixed-types [102].

Nitrogenase is the principle enzyme of nitrogen fixation and requires 35 iron and 2 molybdenum atoms per enzyme, though Mo may be replaced by vanadium or iron depending on availability [103]. *Azotobacter* specifically can encode Mo-, Fe-, and V-only nitrogenases and produces a number of siderophores including aminochelin, azotochelin, protochelin, and 2,3-dihydroxybenzoic acid (DHBA), vibrioferrin, and azotobactin. In particular, azotobactin is yellow-green fluorescent peptide and a strong chelator and forms complexes with Mo, V, and Fe.

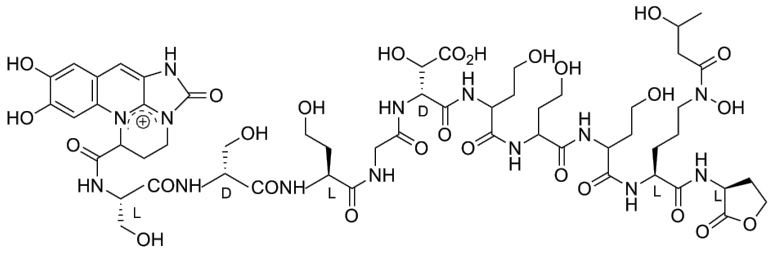


Figure 2.16: Structure of azotobactin, a yellow-green fluorescent peptide.

Fig. 2.16 illustrates the absorbance spectra of cultured *A. vinelandii* (AZBB163) supernatant with a significant peak between 375 nm and 425 nm. Siderophores are iron(III)-chelating molecules that facilitate iron acquisition in environments with low iron bioavailability.

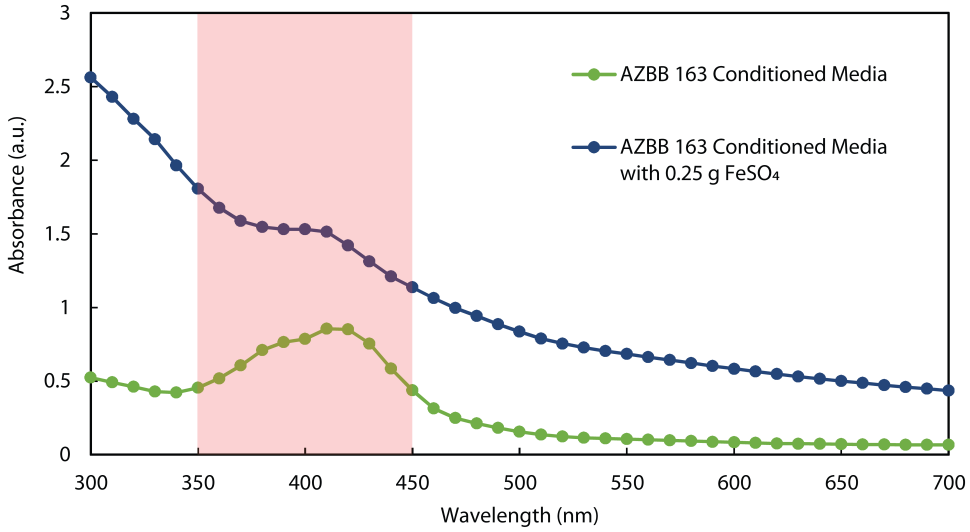


Figure 2.17: Absorbance scanning of *A. vinelandii* supernatant between 300 nm and 700 nm after three days growth in Burks N-free liquid medium (n = 3). A clear peak is visible between 350 nm and 450 nm. The relative peak magnitude is significantly dampened with the addition of $FeSO_4$, confirming the presence of siderophores.

Fig. 2.17 illustrates the necessity of high iron concentration in the union minimal medium as well as the necessity to use longer absorbance wavelengths. From a consortia perspective, strains labeled with GFP or cyan fluorescent protein (CFP) are fundamentally incompatible for co-cultivation with *A. vinelandii* and detection using spectrophotometry or FCM. FCM of *A. vinelandii* monoculture also indicated significant deviation in cell size and complexity using front and side scatter, respectively. While likely due to PHB accumulation as intracellular granules, it further complicates the use of FCM in community composition analyses for even simple studies of co-cultures differentiation using cell size (Appendix E). Spot plating of cultures, while a more accurate means of quantifying viability compared to absorbance, proved difficult due to the slow growth of *S. elongatus* cells on BG-11 agar plates.

Absorbance spectroscopy was used in lieu of FCM and plate spotting as a proxy for cell count and biomass accumulation for constituents of the tripartite system. A near IR wavelength of 750 nm was selected to avoid fluorescent molecules of *S. elongatus*, namely chlorophyll-*a* and carotenoids, as well as the siderosome of *A. vinelandii*. Chl-*a* fluorescence

data were nonetheless collected to gauge cyanobacterial health as loss of chl-*a* is a strong indicator of cell death and culture crash.

Nutrient and product concentrations were evaluated from the supernatant of samples after centrifugation at 5,000 g for 10 minutes to pellet cells. Sucrose and IPTG concentrations were quantified using high performance liquid chromatography (HPLC) (Shimadzu) fit with a Rezex ROA-Organic Acid column (Phenomenex) and 0.005 M H_2SO_4 mobile phase. Amino acid production was quantified using HPLC (Agilent) fit with a Luna C18(2) column for amino acid detection (Phenomenex). A dual mobile phase elution gradient was used according to manufacturers' specifications. The first phase was composed of 40 mM potassium phosphate dibasic (pH of 7.2) with 30 mg sodium azide as an antimicrobial reagent. The second phase was equal volumes acetonitrile and methanol. Samples were derivatized via autosampler (Agilent) using an orthophaldialdehyde reagent (orthophaldialdehyde dissolved in with methanol and added to a 0.1 M borate buffer with 3-mercaptopropionic acid) and 40 mM monopotassium phosphate (pH 11.2) then eluted with dual mobile phase gradient with diode array detection (deuterium lamp; 338 nm for L-lysine and L-glutamic acid). Ammonium concentrations were determined by treating samples with an ionic strength adjuster (Thermofisher) to convert all ammonia ions to ammonium, and detection with an ammonium ion selective electrode (Midland Scientific). Nitrate consumption by *S. elongatus* monocultures was detected by spectrophotometry using a ratio of absorbances (λ_1 = 220 nm, λ_2 = 275 nm). Nitrate absorbance in the union medium was unreliable when PCA was added.

2.4.10 Accumulation of PHB from *Azotobacter vinelandii*

A. vinelandii accumulates PHB and alginate as a mechanism for carbon storage and regulation of intracellular oxygen. As a diazotrophic aerobic organism, *A. vinelandii* faces a paradox of maintaining sufficient intracellular oxygen for respiration without poisoning nitrogenase. It is well known that many bacterial species produce PHB when under oxygen, phosphate, or nitrogen nutrient limitation in the presence of excess carbon. The tripartite

culture priming conditions contributed to significant biofilm formation in the *A. vinelandii* chamber, peaking between days 4 and 7.

To verify PHB production, cell or biofilm samples were taken, centrifuged and 10,000 g, and then dried for 72 hours at 60 °C. Dried samples were weighed and then treated with 1 mL 12 M H_2SO_4 for 60 minutes at 90 °C in open 1.5 mL Eppendorf tubes with intermittent vortexing. After 100x dilution in DI H_2O, samples were sterile filtered and analyzed using HPLC (organic acid column, UV-absorption at 210 nm) [83]. Standards of derivatized crotonic acid (Sigma-aldritch) and PHB (Sigma-aldritch) were both included (Appendix C).

2.4.11 Transduction of fast growth rate locus into strains of interest

Transductions were completed using the P1 bacteriophage. Overnight K-12 *cscBKA*::*nlpA* cultures were infected with P1 and added to Top agar. After overnight growth, Top agar was removed using a cell spreader, centrifuged, and treated with chloroform. The supernatant was then extracted for infection of target strains. A construct with YFP and flanking chloramphenicol resistance gene were integrated into the *intS* locus was used [24,104]. Following standard P1 transduction protocol, the YFP::*intS* strains was infected with P1 for 30 minutes at 37 ℃ without shaking. Infection was terminated with the addition of 1 M Na_2Citrate and recovered for 3 hours at 37 ℃. The culture volume was then spread on MacConkey agar plates with 50 μg/mL kanamycin sulfate, 34 μg/mL chloramphenicol (in ethanol), 20 g/L sucrose, and 1 mL of 1 M Na_2Citrate solution (P1 Plates). Plates were dried were incubated overnight at 37 ℃. Transductants were screened by growth on sucrose and fluorescence in minimal media (λ_{ex} = 475 nm, λ_{em} = 517 nm) indicated successful transduction.

Chapter 3: Random Integration and Screening Yields *E. coli* K-12 Derivatives Capable of Efficient Sucrose Utilization

Chromosomal expression of heterologous genes offers significant advantages over episomal expression yet remains difficult to optimize through site-specific integration. The challenge is in large part due to the unpredictability of chromosomal gene expression, which is affected by multiple factors including the recently elucidated local genomic context. In this work we exploit random integration of a three-gene *csc* operon encoding non-phosphotransferase sucrose catabolism into *E. coli* K-12 using Tn5 transposase. Isolates from the transposon library yielded a range of growth rates on sucrose, including some that were comparable to that of *E. coli* K-12 on glucose (μ_{max} = 0.70 ± 0.03 hr^{-1}). Narrowness of the growth rate distributions, improved gene expression conferring faster growth compared to that of plasmids, and enhanced growth rate upon transduction into strains that underwent adaptive laboratory evolution indicate that efficient *csc* expression is attainable and not limiting to cellular growth. We also show that transduction of a *csc* fast-growth locus into an isobutanol production strain yields high titer (7.56 ± 0.25 g/L) with significant sustainability benefits. Our results demonstrate that random integration is a viable and effective strategy for optimizing heterologous expression within the context of cellular metabolism for certain desirable phenotypes.

The majority of the work presented in this chapter has been submitted to ACS *Synthetic Biology* as follows: David N. Carruthers, Tatyana E. Saleski, Scott A. Scholz, Xiaoxia Nina

Lin. "Random integration and screening yields *E. coli* K-12 derivatives capable of efficient sucrose utilization." (2020).

3.1 Introduction

Escherichia coli K-12 is a widely employed chassis in synthetic biology. Heterologous pathways in K-12 are often implemented with plasmid-mediated expression, which exhibits significant disadvantages. Plasmids require maintenance via antibiotic selection, often yield unstable or highly variable expression due to copy number or internal rearrangement, and often place an undue metabolic burden on the strain thereby affecting growth and production performance [105–107]. Thus, despite plasmids' modularity and relative ease of transfer between strains, they are ill-suited for industrial bioproduction of bulk chemicals.

A diverse range of molecular cloning strategies have been developed and optimized for chromosomal integration, namely through either site-specific recombineering or random transposons. Recombineering is phage-mediated directed mutagenesis and exemplified by the λ-Red system. While widely used, λ-Red requires 30 to 50-bp flanking homology and suffers low efficiency for fragments over 2.5 kb such that integration of entire operons is often difficult [55, 56]. To avoid the limitations and inefficiencies of λ-Red, conditional-replication, integration, and modular (CRIM) plasmids were developed [51]. CRIM plasmids enable site-specific integration by targeting specific phage attachment sites (*attB*) and have been further optimized through the development of one-step "clonetegration" with integration of constructs over 6 kb [51, 56, 108]. CRISPR-based approaches have also been developed more recently and use distinct guide RNA sequences to cleave the host genome at specific sites with the Cas9 endonuclease, thereby facilitating high fidelity integration [109–111]. While both CRIM and CRISPR-Cas9 systems allows more efficient site-specific integration, the expression of a gene or operon is heavily influenced by the genomic context of the integration site and the optimal expression level is often unknown.

Alternatively, phage-derived transposable elements enable random integration into the

chromosome. Transposable elements are characterized as DNA sequences flanked by specific inverted repeats that are recognized by a particular transposase [112, 113]. The transposase binds to these sequences and forms a transposon complex, then binds to the chromosome either randomly (Tn5) or favoring a specific site (Tn7, *attTn7*) [114, 115]. Most regions of the *E. coli* genome are accessible for integration using tranposons with the exception of essential genes. However, all sites are affected by a myriad of associations to nucleoid folding structure, nucleoid associated proteins, as well as gene knockout and localized genetic elements [116–118]. Recent studies have probed genomic expression by Tn5 insertional transposon mutagenesis of reporter proteins, yielding highly variable locus-dependent expression, with up to 300-fold differences in expression between the highest and lowest loci [116, 118]. Genomic topography is complicated by combinatorial effects between genomic features, which in turn convolute causality in expression mapping. Like site-specific integration, transposition efficiency decreases with construct length, but integration has been demonstrated with much larger fragments (>12 kb) than λ-Red recombineering or CRIM.

While genomic mapping of expression level is useful for informing site-specific integration, it is often unclear whether high or low expression loci generate a desired phenotype, especially if the transposed gene or operon affects cellular metabolism. Transposome transformation produces a library of integrants with variable expression of the transposed gene or operon, which enables screening for a desired phenotype directly. Direct screening is especially important for complex pathways (*e.g.*,. biochemical production or growth on a non-native substrate) that may require nuanced and potentially nonintuitive expression of different constitutive genes (*i.e.*, high expression of a rate limiting gene but low expression of other genes).

Here, we investigate random chromosomal integration of an entire operon through Tn5 transposition into *E. coli* K-12 MG1655. The operon, *cscBKAR*, encodes a non-phosphotransferase (non-PTS) sucrose catabolism pathway found natively in *E. coli* W, a much less utilized strain that is genetically distinct from the canonical *E. coli* K-12. The

four genes responsible for chromosomal sucrose catabolism are *cscB*, *cscA*, *cscK*, and *cscR*, encoding a sucrose H^+/symporter or sucrose permease, an invertase, a fructokinase, and a regulator protein, respectively [119]. Unlike *E. coli* W, *E. coli* K-12 is predominantly cultured on glucose and has no native pathway for sucrose catabolism. Life cycle assessments of fermentable carbohydrates have established that sucrose is a significantly more sustainable feedstock than corn-based glucose in terms of both greenhouse gas emissions and energetics [70]. Chromosomal integration of sucrose catabolism into *E. coli* K-12 may yield strains with better sustainability metrics from growth and biochemical production on sucrose compared to glucose.

Investigation of plasmid expression of the *cscBKA* operon without the *cscR* regulator in *E. coli* K-12 has yielded strains with substantially lower growth rates on sucrose than K-12 on glucose [58]. Instability of episomal *csc* expression is also well documented, including an instance of spontaneous chromosomal integration at the *lacZ* locus in K-12 [58,101,120,121]. Site-specific chromosomal integration of the *csc* operon has been conducted using CRIM and knock-in knock-out vectors in an attempt to optimize K-12 growth on sucrose [58, 107, 122]. While generally producing faster strains than those carrying plasmids, the integrants from each study showed 20-30% slower growth rates than K-12 on glucose and 50% slower growth rates than *E. coli* W on sucrose [58, 120]. Adaptive laboratory evolution (ALE) has demonstrated that growth rate may be increased substantially by subsequent passaging of K-12 *csc* integrants, though these studies have identified common metabolic culprits that limit overall growth rate rather than focusing on the *csc* operon itself [122–124]. This presents an opportunity to exploit the chromosomal landscape to yield integrants with optimal expression of the sucrose pathway.

We generate a transposon library with a distribution of growth rates depending on chromosomal position and orientation. Our results indicate that the growth rate varies substantially between integration sites and reveal nonintuitive differences between plasmid and chromosomal expression. In addition, we use P1 transduction to transfer an optimized *csc*

fast-growth locus into two types of *E. coli* K-12 derivative strains to demonstrate the feasibility of carbon source switching: 1) ALE strains adapted on glucose; and 2) a chromosomally integrated isobutanol production strain.

3.2 Results and discussion

3.2.1 Overview of chromosomal integration of *csc* operon into *E. coli* K-12

The *cscBKA* gene cluster was amplified from *E. coli* W (ATCC 9637), a well-studied strain that grows quickly on sucrose [67, 119]. The repressor, *cscR*, was not included because derivative strains of *E. coli* W with *cscR* deleted were shown to have significantly higher sucrose uptake rates [125]. The *cscBKA* genes, mediating the first steps in sucrose catabolism, encode a sucrose H^+/symporter or sucrose permease, an invertase, and a fructokinase, respectively (Fig. 3.1A). The native operon has a bidirectional promoter between *cscK* and *cscA*. The amplified *csc* cassette was cloned into a plasmid pSAS30, a derivative of pET24-b, between a kanamycin resistance gene (*kanR*, neomycin phosphotransferase II) and an mNeonGreen (*mNG*) fluorescent reporter through Gibson Assembly (Fig. 3.1B). Next, the *csc* genes and resistance marker were amplified from the plasmid and randomly integrated into the K-12 MG1655 chromosome using Tn5 transposase. Integrants were screened on MacConkey agar and characterized using absorbance and curve fitting. Lastly, a desirable fast-growth integration locus was transduced into other strains of interest (Fig. 3.1C).

3.2.2 Characterization of strains carrying *csc* plasmids

Plasmid-carrying strains were initially developed to investigate the specific drawbacks of expressing *csc* genes from plasmid in light of the general issues reported in the literature. The *csc* genes were cloned into two plasmids differing only in the orientation of the *csc* operon; pAKB for *cscAKB* and pBKA for *cscBKA* (Fig. 3.2A). Both orientations have been reported in the literature [88, 122]. More importantly, Tn5 integration is not orientation

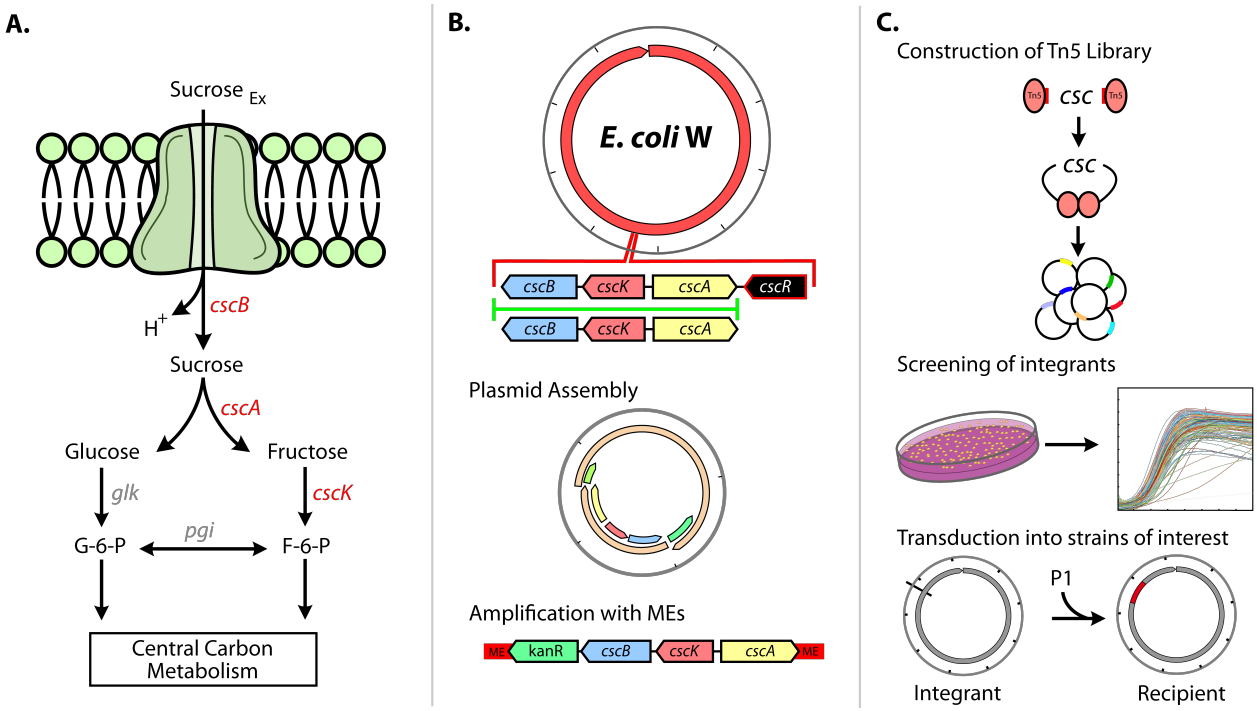


Figure 3.1: Overview of integrating sucrose pathway genes from *E. coli* W into K-12 derivative strains. (A) A diagram of non-PTS sucrose catabolism pathway in *Escherichia coli* W with relevant genes highlighted in red. (B) Workflow of integration construct assembly. The *cscBKA* operon is amplified from *E. coli* W without *cscR*, cloned into a plasmid and flanked by a kanamycin resistance gene and mNeonGreen fluorescence gene. (C) Integration of the *csc* operon and *kanR* by random Tn5 transposition into *E. coli* K-12 MG1655, subsequent screening of integrants and P1 transduction of a characterized locus into selected recipient strains

specific hence transposition of both *cscAKB* and *cscBKA* enables examination of possible strand effects and the position of the flanking *kanR* gene. Each plasmid was transformed into K-12 MG1655. Resulting strains were grown in liquid M9 media with 20 g/L sucrose as the sole carbon source and in the presence or absence of 50 μg/L kanamycin over five serial passages. Three biological replicates were included for each strain/condition with growth and *mNG* fluorescence monitored over time. It was observed that strains harboring pBKA se grew faster with kanamycin than without it, while strains harboring pAKB grew at similar rates independent of kanamycin addition (Fig. 3.2B). Strains harboring pBKA and growing with antibiotic had a shorter lag time compared to those without antibiotic. It is unclear why these trends arose, though it is suspected that partial loss of selective pressure from the

antibiotic to maintain the plasmid, metabolic burden from expressing genes on the plasmid, or a combination of these two could have contributed to the decrease in growth rate. It is also worth noting that the two strains with different *csc* orientations exhibited different growth profiles, potentially due to different expression levels of these genes.

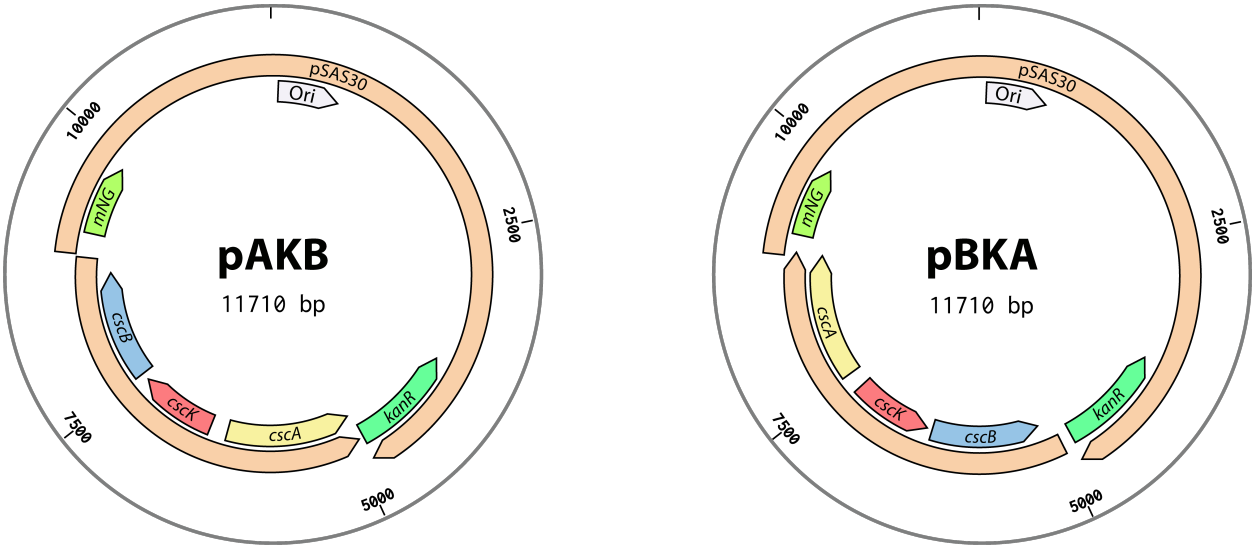


Figure 3.2: Characterization of plasmid-carrying *E. coli* strains. (A) Plasmid maps of pAKB and pBKA differing in the orientation of the *cscBKA* operon, both flanked at the 5' by *kanR* and at the 3' by mNG. (B) Growth curves of plasmid-carrying strains with and without 50 μg/mL kanamycin (n = 3) in M9 minimal medium with 20 g/L sucrose over 5 passages. (C) End-point *mNG* fluorescence of individual replicates over 5 passages

The plasmids also expressed *mNG* fluorescence that enabled investigation of the effect of antibiotic on plasmid gene expression over time (Fig. 3.2C). It was found that each strain/condition exhibited unique trends in the fluorescence signal. When the pAKB-harboring *E. coli* strain was grown without kanamycin, two replicate cultures showed a 5-fold increase in fluorescence, while the fluorescence declined in a third replicate. In the presence of kanamycin, the pAKB-harboring strain exhibited a slight decrease in fluorescence between passages one and two then exhibited a two-fold increase between passages two and five. The *E. coli* strain carrying pBKA, on the other hand, showed drastically different patterns in the fluorescence signal. In the absence of kanamycin, the strain reported a low fluorescence largely consistent across passages as well as replicates. In the presence of kanamycin, sur-

prisingly, fluorescence of all three replicate cultures decreased eight-fold between passage one and five. There was no correlation between end-point fluorescence and $\mu_{\max}$ of each replicate. The reasons behind these distinct observations are unclear. Nevertheless, these findings highlight the instability and inconsistency of both cellular growth and gene expression conferred by *csc* genes carried on plasmids, even with the supplementation of costly antibiotics.

3.2.3 Characterization of Tn5 integrant libraries by analysis of growth phenotype, integration loci, and gene expression

We sought to integrate the *csc* genes into the chromosome of *E. coli* K-12 in light of issues associated with plasmid gene expression. We explored transposase-mediated random integration, which in our previous work has revealed up to a 300-fold difference of reporter gene expression levels across the *E. coli* genome and allowed optimization of pathway gene expression in an extensive phenotype landscape [118, 126]. Tn5 transposons for *cscAKB* and *cscBKA* with a flanking *kanR* marker were transformed into *E. coli* K-12 MG1655. Screening of the resultant libraries was conducted via plating on MacConkey agar with kanamycin. Spot plating of sequential 2-fold dilutions of the transformation recovery on MacConkey agar with kanamycin indicated libraries of approximately 8,000 total integrants for *cscBKA* and 6,000 integrants for *cscAKB*. The sizes of these libraries suffice for this work (as indicated in analysis presented below) but might have been limited by relatively low transformation efficiency due to the length of the sequence ($\sim$ 5.2 kb) and the selection method which may have removed integrants not sufficiently expressing *kanR*.

After 48 hours of plate recovery, 188 integrants were randomly selected from each library for growth phenotype analysis using kinetic spectrophotometer reads of culture in 96-well plates over 16 hours. Of the colonies inoculated into M9 sucrose medium, 92.5% of *cscBKA* integrants grew while only 69.1% of *cscAKB* grew. Each transposon library presented a spectrum of growth rates for the integrants that grew (Fig. 3.3A, 3.3B). Maximum

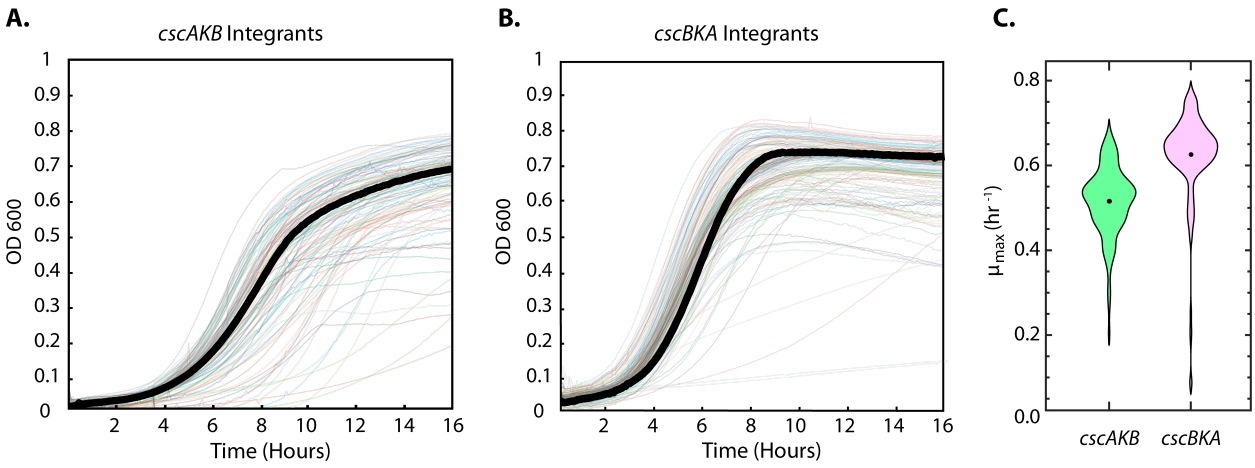


Figure 3.3: Growth phenotypes of Tn5 integrants. (A, B) Growth profiles of *cscBKA* and *cscBKA* integrants in M9 medium with 20 g/L sucrose. Opaque growth curves represent individual isolates and the dark black line shows the growth of the respective bulk Tn5 library. (C) Violin plots of the distribution of growth rates in each library. Dots represent the mean growth rate for respective *csc* orientation (n = 130, 174 for *cscBKA* and *cscBKA*, respectively).

specific growth rates of the *cscBKA* integrants were significantly higher ($\mu_{\max} = 0.62 \pm 0.11$ hr^{-1}, n = 174) than those of *cscAKB* integrants ($\mu_{\max} = 0.51 \pm 0.08$ hr^{-1}, n = 130) (Fig. 3.3C). Broadly, this indicates that the position of the kanamycin resistance gene with respect to the *csc* operon has strong effects on the expression of *csc* genes and thus the growth rate on sucrose. This difference may explain, in part, why the integration library size for *cscAKB* was smaller than that for *cscBKA*. Fig. 3.3C also illustrates a narrower distribution of *cscBKA* growth rates, with 74% of integrants exhibiting $\mu_{\max}$ between 0.55 and 0.70 hr^{-1}. Six integrants from each library representative of high, medium, and low growth rates were further characterized by high-confidence growth rate measurement, identification of integration loci, and gene expression analysis. Growth phenotypes were first verified alongside the two plasmid-carrying strains, pAKB and pBKA, and reference strains *E. coli* W and K-12 (Fig. 3.4A).

All selected integrants outperformed both pAKB (maximum specific growth rate: 0.33 $\pm$ 0.03 hr^{-1}) and pBKA (0.53 $\pm$ 0.03 hr^{-1}) except for B4 (0.47 $\pm$ 0.05 hr^{-1}). In particular, three integrants, B2 (0.70 $\pm$ 0.03 hr^{-1}), B1 (0.67 $\pm$ 0.03 hr^{-1}), and B5 (0.66 $\pm$ 0.05 hr^{-1})

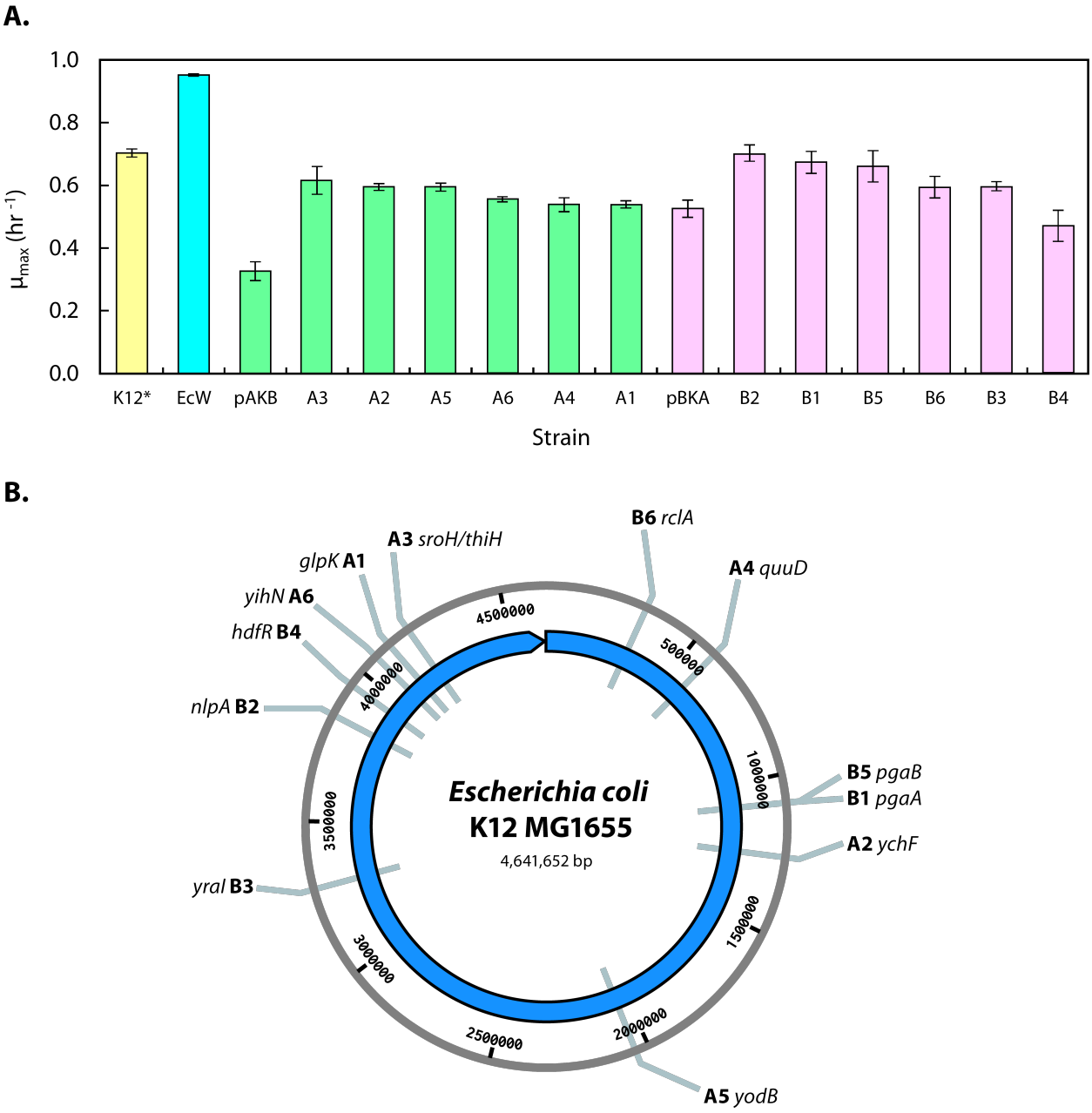


Figure 3.4: Isolation of select integrants from each Tn5 library with variable growth for footprinting and expression analysis. (A) Growth rates of isolated integrants, parent plasmids, and reference strains on M9 with 20 g/L carbon source (n = 4). *K-12 was grown on glucose, all others grown on sucrose. (B) Footprinting analysis of integration sites in *E. coli* K-12 chromosome. Each integrant is listed with genetic insertion site.

not only outperformed pBKA but exhibited a growth rate comparable to that of *E. coli* K-12 on glucose (0.70 ± 0.01 hr^{-1}). To our knowledge, these integrants mark the fastest growing *E. coli* K-12 base strains (*i.e.*, modified strains without undergoing adaptive laboratory

evolution) published so far. This result demonstrates that random transposition followed by simple phenotype screening is a viable and effective strategy for isolation of certain target phenotypes.

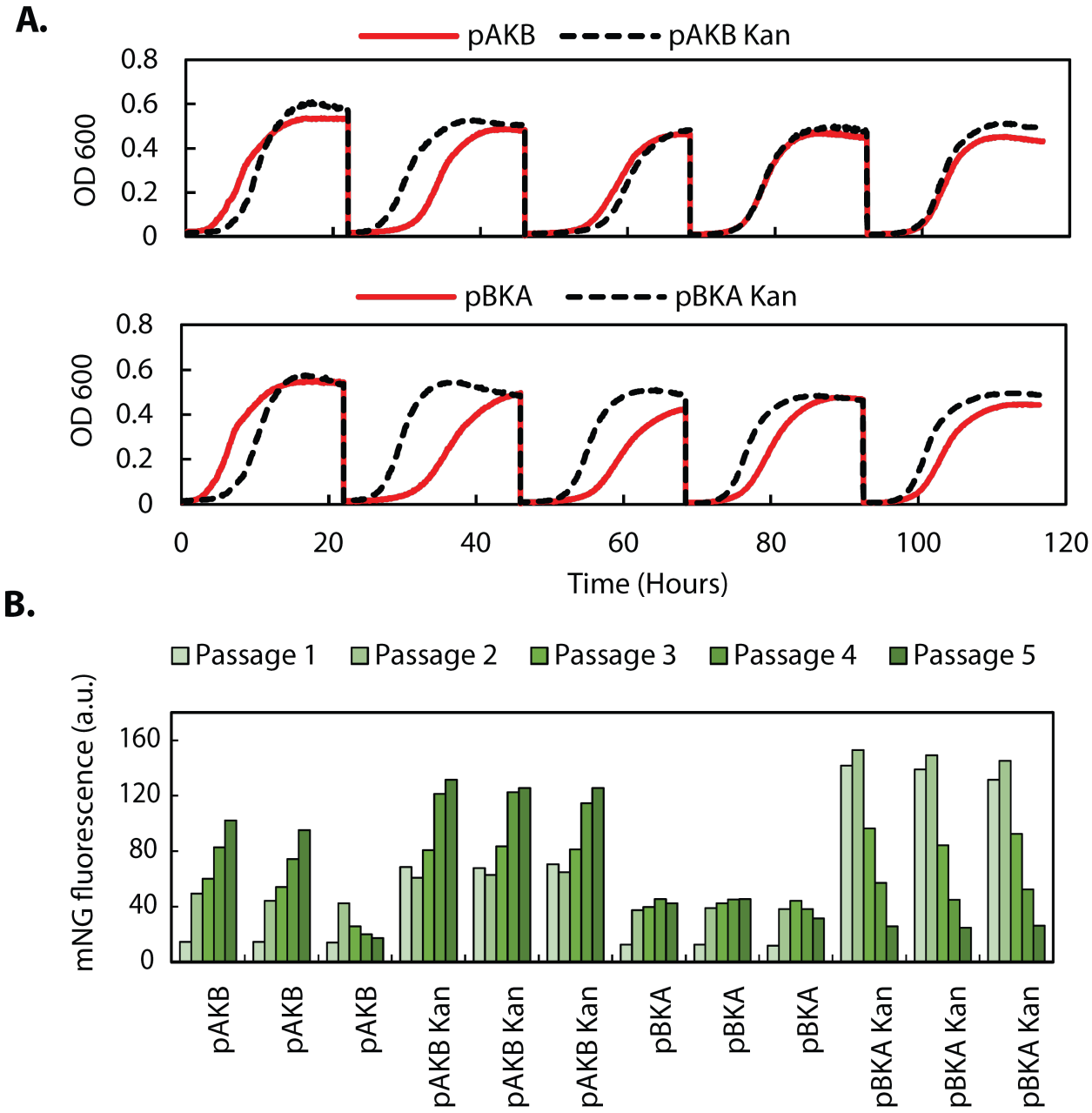


Figure 3.5: Maximum specific growth rates of *E. coli* K-12, *E. coli* W, and *E. coli* K-12 MG1655 *cscBKA*::*nlpA* on various concentrations of glucose, fructose, sucrose, and an equimolar mixture of fructose/glucose. Cultures were precultured in 20 g/L of the respective carbon source, inoculated into the respective concentrations of said carbon sources, and then passaged once for growth rate determination (n = 3).

E. coli W (0.95 ± 0.01 hr^{-1}) was significantly faster than all other strains. None of the selected integrants surpassed the growth rate of *E. coli* K-12 on glucose, which suggests the presence of a bottleneck in general carbon metabolism apart from uptake and initial processing of substrates that limits the growth of *E. coli* K-12 derivatives. To explore this hypothesis, *E. coli* K-12, W, and the integrant B2 were grown on sucrose, glucose, fructose, and equimolar mixture of fructose/glucose. Each strain exhibited similar growth rates independent of carbon source or concentration; *E. coli* K-12 and the integrant B2 grew at similar rates, yet slower than *E. coli* W (Fig. 3.5). Similarity of growth rates on multiple carbon substrates also supports the presence of a metabolic bottleneck.

Integration sites of selected integrants were determined using transposon footprinting, which involved cleaving gDNA using a four-base-pair endonuclease and subsequent Y-linker ligation for downstream PCR amplification and sequencing [118]. All integrants were found to be nonclonal (Fig. 3.4). Eleven of the twelve integration sites were intragenic with one site in an intergenic position (Table 3.1).

The sucrose catabolism phenotype is determined by a combination of *csc* gene expression and disruption to the local region on the chromosome where the integration occurs. Local gene disruption was primarily inactivation of non-essential genes for our integrants and the effects can be partially inferred based on previous gene knockout studies of *E. coli* K-12 [129]. B1 and B5 had the integration in *pgaA* and *pgaB*, respectively, in an operon encoding adhesin exopolysaccharide poly-beta-1,6-N-acetyl-D-glucosamine (PGA), a biofilm associated polymer. All genes of the *pgaABCD* operon are necessary for PGA production and optimal biofilm formation [130]. Knockout of constitutive *pga* genes does not affect growth under non-biofilm-forming conditions and may explain comparable growth rates between these two integrants in our study [131]. The fastest integrant, B2, was found to have the integration in the *nlpA* gene, which encodes a lipoprotein associated with methionine uptake and outer membrane vesicles but is ultimately non-essential for growth of *E. coli* [132]. [133, 134] On the other hand, the slowest growing integrant in this screen, A1,

Table 3.1: Integration sites for select *csc* integrants. * indicates genes whose function is poorly understood.

Name	**Int. coordinate**	**Int. gene**	**Gene Description** [127, 128]
A1	4,116,480	*glpK*	Glycerol kinase
A2	1,257,420	*ychF*	ATPase that binds the 70S ribosome and the 50S ribosomal subunit
A3	4,190,513	Int. *sroH/thiH*	Intergenic region
A4	574,591	*ybcQ**	Prophage anti-termination protein
A5	2,042,722	*yodB**	Cytoplasmic membrane protein
A6	4,062,263	*yihN**	Inner membrane protein; glycerophosphodiester uptake family
B1	1,090,138	*pgaA*	Exports the biofilm adhesin exopolysaccharide poly-beta-1,6-N-acetyl-D-glucosamine (PGA) across the outer membrane
B2	3,839,639	*nlpA**	Cytoplasmic membrane lipoprotein 28
B3	3,288,348	*yraI**	Fimbrial chaperone that may facilitate surface adhesion in specific niches
B4	3,947,617	*hdfR*	Negatively regulates the expression of the flagellar master operon
B5	1,088,464	*pgaB*	Outer membrane lipoprotein required for N-deacetylation and hydrolysis of PGA
B6	319,816	*rclA**	Reactive chlorine resistance A

had the *csc* genes integrated into *glpK*, which encodes for glycerol kinase and is necessary for growth on glycerol.

Next, we investigated the gene expression profile of the fast-growing B2 integrant (*cscBKA*::*nlpA*) in comparison to plasmid-carrying strains and *E. coli* W. Expression of the *cscBKA* operon was quantified using RT-qPCR with respect to a housekeeping gene,

ssrA. *ssrA*, which encodes tmRNA, was chosen due to strong correlations between cycle threshold number reflecting RNA concentration and cell culture densities during growth for both K-12 and W (Fig. 3.9). In addition, *kanR* expression was determined for the plasmid-carrying strains and B2. As shown in Fig. 3.6A, both pAKB and pBKA exhibited high expression of *kanR*, approximately 30-fold higher than that of B2. Interestingly, expression of the *csc* genes were mostly comparable across the two plasmid-carrying strains and B2.

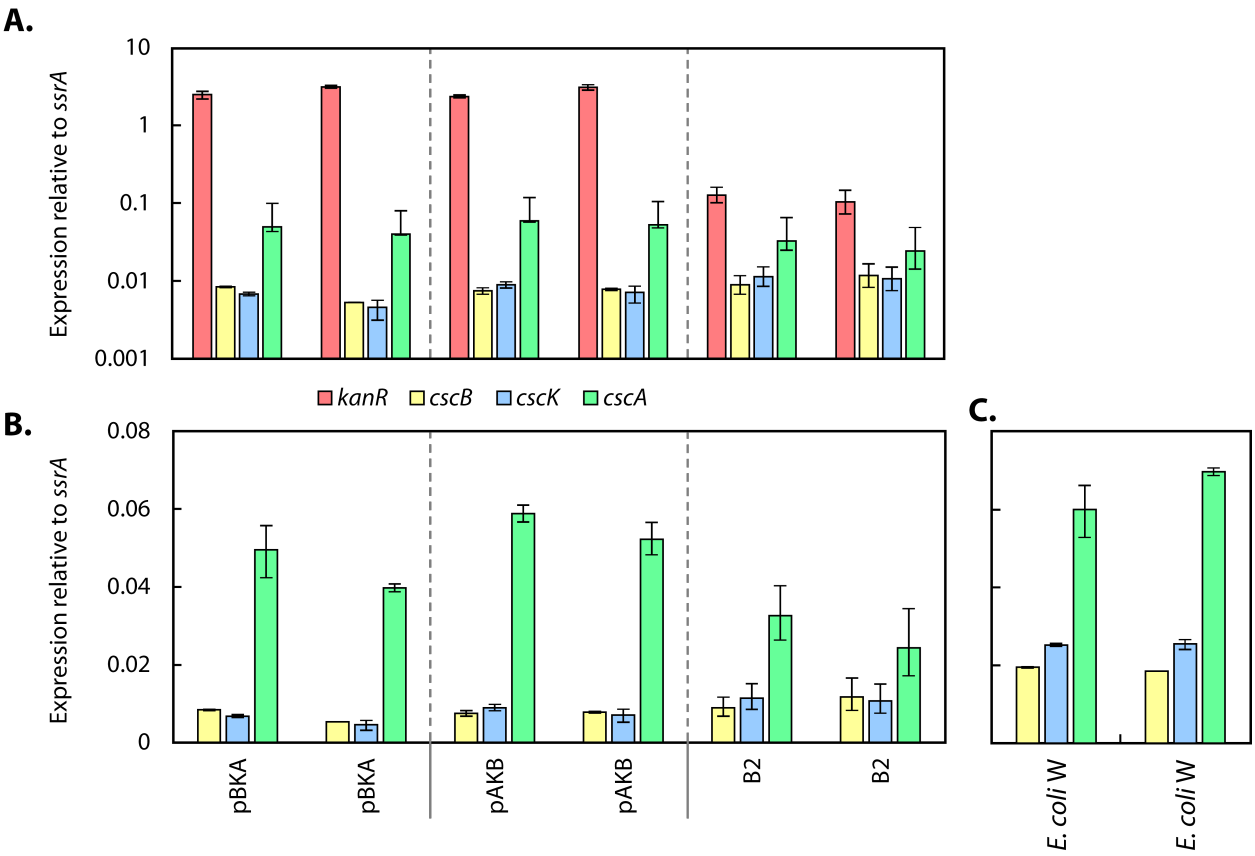


Figure 3.6: Expression of *csc* genes and *kanR* on plasmids, in integrated strain B2, and *E. coli* W. Two biological replicates are plotted for each strain. Error bars represent the standard deviation between technical replicates (n=3) (A) Expression from plasmids and in integrant compared to housekeeping gene *ssrA* on a logarithmic scale. (B) Expression of *csc* genes plot on a linear scale. (C) Expression in *E. coli* W compared to housekeeping gene *ssrA*, plot on a linear scale.

The overall profiles of *cscB*, *cscK*, and *cscA* expression in all three strains follow the same general trend in which *cscA* is most highly expressed while *cscB* and *cscK* are significantly lower. We note that this trend differs from results from previous investigations of *E.*

coli W *cscBKA* expression, which reported higher expression of the *cscB* and *cscK* genes over *cscA* despite similar minimal media [88]. This discrepancy might be due to the use of different reference genes and/or deviations in experimental procedures. Closer examination of *csc* gene expression in these strains (Fig. 3.6B) revealed that *cscB* and *cscK* expression levels were slightly higher in B2 compared to those in either plasmid-carrying strain, whereas *cscA* expression in B2 was lower. The remarkable fast-growth phenotype of B2 with respect to the plasmid-carrying strains may be due to reduced metabolic burden from overexpression of *kanR*, a more balanced relative expression profile across the three *csc* genes, or a combination of these two factors.

We also compared expression levels of the *csc* genes in the above strains to those in *E. coli* W (Fig. 3.6C). Relative to the same housekeeping gene *ssrA*, *E. coli* W appeared to exhibit higher expression of the *csc* operon compared to either the K-12 based integrant or the plasmid-carrying strains.

3.2.4 Transduction of *csc* genes at optimized integration locus into laboratory evolved strains further enhances growth on sucrose

The relative closeness of the maximum specific growth rates of the fastest growing integrants described above to that of *E. coli* K-12 MG1655 on glucose suggests that chromosomal integration of sucrose catabolism can attain but not surpass the maximum specific growth rate of the base strain due to limitation by other factors. Enhanced growth rates of *E. coli* K-12 derivative strains have been reproducibly demonstrated in adaptive laboratory evolution (ALE) studies wherein continuous culturing enables selection of mutants with faster growth rates in glucose minimal medium [123, 135]. *E. coli* strains passaged on minimal media tend to follow the same evolutionary trajectory with specific mutations. Predominant mutations included *rpoB* and *pyrE*/*rph*, which coincided with large shifts in the transcriptome likely due to the mutation of *rpoB*, a global regulators [123, 136]. Almost identical mutations were discovered when ALE was performed on an *E. coli* K-12 strain with *cscBKA* integrated into

the *attB* gene using site-specific integration [108, 122].

We hypothesized that integrating the *csc* genes at an optimized genomic locus we identified above into ALE derived strains which have overcome limiting factors in downstream metabolism conserved for different carbon substrates would lead to further improvement in growth rates on sucrose. Accordingly, we made use of three *E. coli* K-12 derivative strains previously generated through ALE and employed P1-transduction to transfer the *cscBKA*::*nlpA* locus into these ALE strains with growth rate limitations relieved considerably [137]. Each base strain (ALE-1, ALE-2, and ALE-3) represent evolved end-point clones of the *E. coli* K-12 MG1655 ALE experiments [137].

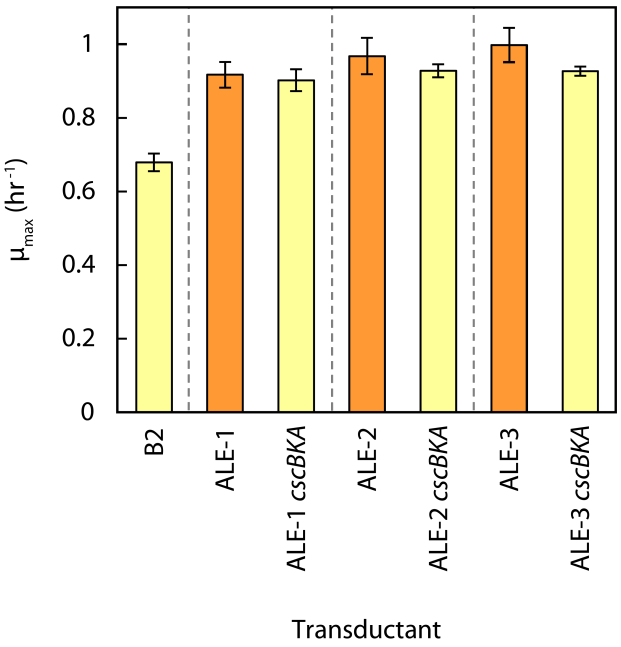


Figure 3.7: Transduction of *cscBKA*::*nlpA* locus into ALE strains with downstream growth limitations relieved leads to enhanced growth rates on sucrose (n = 3). ALE strains were grown on M9 glucose (20 g/L) while B2 and transductants were grown on M9 sucrose (20 g/L).

3.2.5 Transduction of fast-growth *csc* locus enables efficient production of isobutanol from sucrose in *E. coli* K-12

As shown in Fig. 3.7, the transductants, B2::ALE-1, B2::ALE-2, and B2::ALE-3 yielded maximum specific growth rates of 0.90 ± 0.03 hr^{-1}, 0.93 ± 0.02 hr^{-1}, and 0.93 ± 0.01 hr^{-1},

respectively, an increase of between 33% and 37% over the B2 integrant ($p < 0.0005$). Furthermore, the transductants exhibit growth rates comparable to those of their ALE base strains. These results indicate that *nlpA* is an ideal locus for chromosomal *csc* expression and mutations acquired through ALE were able to remove or alleviate certain metabolic bottlenecks limiting maximum growth rates on common carbon substrates. It is also noted that the growth rates of the transduced strains on sucrose were slightly lower than those of the respective ALE base strains on glucose. This may stem from sub-optimal fructose metabolism as the ALE strains were passaged on M9 minimal media with glucose as the sole carbon source. Nonetheless, these results demonstrate the effectiveness of transferring an optimized fast-growth locus into other *E. coli* K-12 strains for sucrose catabolism, which represents a potentially widely applicable strategy for enabling substrate switching in production strains.

Biofuels and biochemicals produced from microbial fermentation can displace petrochemical analogues and are attractive due to their potential to contributing to better environmental sustainability. Efficient isobutanol production was reported previously with a genetically engineered *E. coli* K-12 derivative strain through overexpression of five genes (*alsS*, *ilvCD*, *adhA*, *kivD*; Fig. 3.8A) from two plasmids [138]. The strain, JCL260, demonstrated 86% conversion efficiency of glucose to isobutanol and 12.7 g/L isobutanol production [138]. Saleski et al. recently expanded upon this work and created chromosomally integrated strains by generating and screening transposon libraries of integrants, yielding efficient production of isobutanol from glucose without episomal expression or costly supplementation of antibiotics [126]. Transducing the high-growth-on-sucrose locus from B2 into such an isobutanol-producing strain presents the opportunity to produce isobutanol from sucrose, thereby potentially increasing the sustainability of the process. We therefore employed P1 phage to transduce the *cscBKA*::*nlpA* locus of B2 into the chromosomally integrated production strain by Saleski et al. for switching of carbon sources (Fig. 3.8A) [126]. The resulting strain, C7 CIChE *cscBKA*::*nlpA*, produced 7.56 ± 0.25 g/L isobutanol after 48 hours of induction or approximately 76% isobutanol yield compared to the glucose-utilizing

strain (Fig. 3.8B).

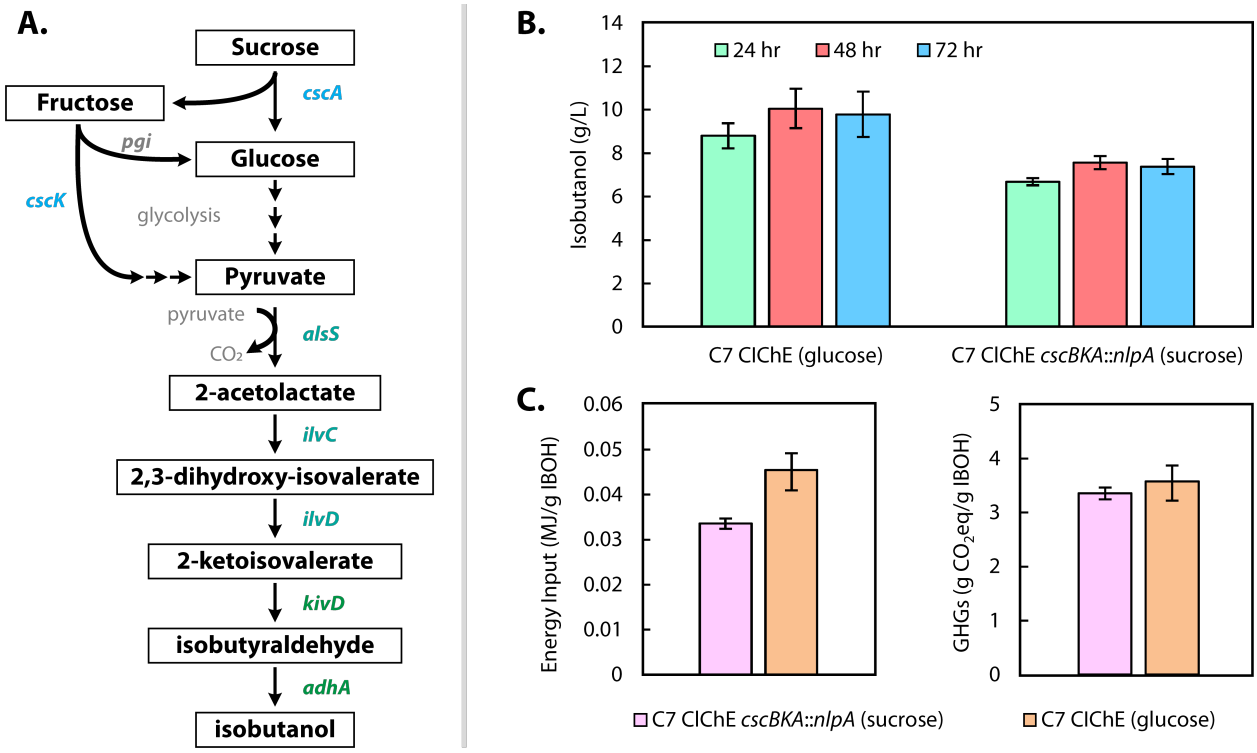


Figure 3.8: Transduction of fast-growth locus into an isobutanol production strain. (A) Overall pathway for isobutanol production from sucrose, with *cscA*, *cscK*, and five isobutanol-specific genes highlighted. (B) Production of isobutanol through chromosomally integrated pathway by the base strain on glucose and the transductant on sucrose. (C) Comparison in sustainability metrics of the two integrant strains on sucrose and glucose.

We also examined the effect of the change of carbon substrate on critical sustainability metrics, particularly energy inputs and greenhouse gas (GHG) emissions. Despite 24% lower isobutanol titer compared to the glucose base strain, C7 ClChE *cscBKA*::*nlpA* was found to exhibit a substantial sustainability advantage, requiring 33% less energy input while producing comparable GHG emissions on a per gram produced isobutanol basis (Fig. 3.8C). From a sustainability standpoint this unoptimized C7 ClChE *cscBKA*::*nlpA* strain is already competitive against its glucose counterpart. It may also impart a cost benefit depending on the relative prices of sucrose and glucose.

The results described here demonstrate that biofermentation of industrially relevant products may be enhanced through carbon source switching by simple P1 transduction of

the *csc* high-growth locus.

3.3 Conclusion

Random integration of a non-phosphotransferase sucrose catabolism operon into *E. coli* K-12 yielded multiple fast-growth integrants that outperformed plasmid-carrying strains and showed comparable growth rates to that of the *E. coli* K-12 base strain on glucose. These integrants demonstrate that transposon mediated integration and phenotypic screening can yield efficient and robust phenotypes. Further studies could expand upon this work by integrating *cscB*, *cscK*, and *cscA* individually into separate loci to further explore the combinatorial genotype and expression landscape. Yet our work indicated that further improvement of the most efficient *csc* integrant we identified was limited by other metabolic bottlenecks in *E. coli* K-12. Transduction of a *csc* fast-growth locus into *E. coli* strains generated from adaptive laboratory evolution on glucose led to further enhancement of growth on sucrose. Likewise, transduction of the same locus into a chromosomally expressed isobutanol-producing *E. coli* strain demonstrates that the fast-growth phenotype is readily transferrable and can be exploited to switch carbon substrate to sucrose in production strains, offering potential cost and sustainability benefits.

3.4 Materials and methods

3.4.1 Construction of *csc* plasmids

The *cscBKA* genes were amplified from *E. coli* W gDNA for integration into a target plasmid, pSAS30. pSAS30 is modified from pET-24b+, a low copy number plasmid, and encodes both kanamycin resistance and mNeonGreen (*mNG*) [118]. The pSAS30 plasmid was used previously for Tn5 integration of the *mNG* reporter for transcriptomic mapping of the *E. coli* chromosome [118]. Primers were selected amplify only the *cscBKA* operon from *E. coli* W (4059 bp), which functions under a native bidirectional promoter between

the *cscA* gene and the constitutively expressed *cscKB* genes. The primers were designed with 20 bp homology to a site nested between the *mNG* and *kanR* genes of pSAS30 for scarless integration with Gibson Assembly (NEBuilder HiFi DNA Assembly). Two amplicons, *cscBKA* and *cscAKB* were generated with opposite homology to determine whether plasmid could affect growth. The entire pSAS30 plasmid was PCR amplified using Phusion Polymerase and digested overnight with endonuclease DpnI prior to assembly. After assembly, plasmids pAKB and pBKA were transformed into *E. coli* DH5α competent cells via electroporation (2500 V, 5 ms). Successful transformants were screened using MacConkey agar plates with sucrose and 50 μg/mL kanamycin sulfate. MacConkey agar contains a pH indicator such that colonies appear purple when a carbon source is being fermented. The combined colorimetric assay and the antibiotic selected enabled isolation of colonies with the correct plasmid. Plasmids pAKB and pBKA were also extracted and cloned into *E. coli* K-12 MG1655. All strains used here and elsewhere are listed in Table 3.3.

Table 3.2: List of plasmids and strains used in this study.

Strain/Plasmid	**Genotype or relevant characteristics**	**Source**
E. coli K-12 MG1655	F$^-$ λ^- *ilvG* mutant *rfb*-50 *rph*-1	CGSC
E. coli W	WT strain, non-PTS *csc*	ATCC
E. coli DH5α	F– ϕ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA*1 *endA*1 *hsdR*17(rK–, mK+) *phoA supE*44 λ– *thi*-1 *gyrA*96 *relA*1	Invitrogen
pSAS30	*mNG*, neomycin phosphotransferase II, pET 24b-derived	[118]
pAKB	pSAS30::*cscAKB*	This study
pBKA	pSAS30::*cscBKA*	This study
A1	*cscAKB*::*glpK*	This study
A2	*cscAKB*::*ychF*	This study
A3	*cscAKB*:: Int. *sroH*/*thiH*	This study
A4	*cscAKB*::*ybcQ*	This study
A5	*cscAKB*::*yodB*	This study

Table 3.2: List of plasmids and strains used in this study.

Strain/Plasmid	**Genotype or relevant characteristics**	**Source**
A6	*cscAKB*::*yihN*	This study
B1	*cscBKA*::*pgaA*	This study
B2	*cscBKA*::*nlpA*	This study
B3	*cscBKA*::*yraI*	This study
B4	*cscBKA*::*hdfR*	This study
B5	*cscBKA*::*pgaB*	This study
B6	*cscBKA*::*rclA*	This study
C7 CIChE	JCL260 Δ*lysA*::FRT Δ*ilvD*::FRT Δ*ttdT*::*kan-alsS* Δ*yqiH*::*ilvCD* with *adhA-kivD* CIChE construct integrated in the *aslB* site	[126]
C7 CIChE *cscBKA*::*nlpA*	C7 CIChE; *cscBKA*::*nlpA*	This study
ALE-1	End-point *E. coli* K-12 MG1655 ALE strain	[137]
ALE-2	End-point *E. coli* K-12 MG1655 ALE strain	[137]
ALE-3	End-point *E. coli* K-12 MG1655 ALE strain	[137]
ALE-1::B2	ALE-1; *cscBKA*::*nlpA*	This study
ALE-2::B2	ALE-2; *cscBKA*::*nlpA*	This study
ALE-3::B2	ALE-3; *cscBKA*::*nlpA*	This study

All primers used here and elsewhere are listed in Table 3.3.

Table 3.3: List of primers used in this study. Underlined sequences represent flanking homology or mosaic ends.

Primer	**F/R**	**Application**	**Sequence (5' to 3')**	**Source**
Csc1	F	Amplification of *cscBKA* genes for Gibson Assembly BKA orientation	CCA GTA GCT GAC ATT CAT CCG GGG TCG ACA ATG TCC TGG AAA TCA GC	This study
Csc2	R	Amplification of *cscBKA* genes for Gibson Assembly BKA orientation	GCG GAA CAC GTA GAA AGC CAG TCC GTT AAC CCA GTA GCC AGA GTG C	This study
Csc3	F	Amplification of *cscBKA* genes for Gibson Assembly AKB orientation	CCA GTA GCT GAC ATT CAT CCG GGG TTT AAC CCA GTA GCC AGA GTG C	This study
Csc4	R	Amplification of *cscBKA* genes for Gibson Assembly AKB orientation	GCG GAA CAC GTA GAA AGC CAG TCC GCG ACA ATG TCC TGG AAA TCA GC	This study
Csc5	F	Amplification of pSAS30 backbone for Gibson Assembly	CGG ACT GGC TTT CTA CGT GTT C	This study
Csc6	R	Amplification of pSAS30 backbone for Gibson Assembly	ACC CCG GAT GAA TGT CAG CTA C	This study
Csc7	F	Assembly check *cscBKA* from *kanR*	CGA TTG TCT GTT GTG CCC AGT C	This study
Csc8	R	Assembly check *cscBKA* from *mNG*	ACG TAC ATC GGC TGG TTC TTC A	This study

Table 3.3: List of primers used in this study. Underlined sequences represent flanking homology or mosaic ends.

Primer	**F/R**	**Application**	**Sequence (5' to 3')**	**Source**
Csc9	F/R	Assembly check from *cscB*	TTG ACA GGG ACG GAA TTA GG	This study
Csc10	F	Amplification of *csc* with mosaic end	/5Phos/CTG TCT CTT ATA CAC ATC TCC TCG GTA CCA AAT TCC AGA	This study
Csc11	R	Amplification of *cscBKA* with mosaic end	/5Phos/CTG TCT CTT ATA CAC ATC TCG ACA ACA ACC GGA AAA AGT	This study
Csc12	R	Amplification of *cscBKA* with mosaic end	/5Phos/CTG TCT CTT ATA CAC ATC TGC CAG TCC GTT AAC CCA GTA	This study
Csc13	R	Footprinting nested in *kanR*	GCG CTC ATC TCT CCC TTA TG	This study
CviAII-5N-YAT5	-	Y- linker for footprinting	ACT ACG CAC GCG ACG AGA CGT AGC GTG	[118]
CviAII-YAT3	-	Y- linker for footprinting	/5Phos/ATC ACG CTA CGT CCG TGT TGT CGG	[118]
Csc14	F	qPCR *kanR*	CTC GTC CTG CAG TTC ATT CA	This study
Csc15	R	qPCR *kanR* check	AGA CAA TCG GCT GCT CTG AT	This study
Csc16	F	Amplification of *cscA* for qPCR (Sabri)	GTC CGG ACA TTC CCA CAT ATA G	[88, 139]
Csc17	R	Amplification of *cscA* for qPCR (Sabri)	AGG CAA CAC GGG GCA GAT CCT G	[88, 139]
Csc18	F	Amplification of *cscK* for qPCR (Sabri)	GCC GGG TTA CTC ACA GGT CTG	[88, 139]

Table 3.3: List of primers used in this study. Underlined sequences represent flanking homology or mosaic ends.

Primer	F/R	Application	Sequence (5' to 3')	Source
Csc19	R	Amplification of *cscK* for qPCR (Sabri)	TTC GCC GTT ACT GCA AGC GCT	[88, 139]
Csc20	F	Amplification of *cscB* for qPCR (Sabri)	ATC CGT CTT CAA ATA CAG CGT GG	[88, 139]
Csc21	R	Amplification of *cscB* for qPCR (Sabri)	CAG CAC AAT CCC AAG CGA ACT GG	[88, 139]
ssrA1	F	Amplification of *ssrA* for qPCR	ACG GGG ATC AAG AGA GGT CAA AC	[140]
ssrA2	R	Amplification of *ssrA* for qPCR	CGG ACG GAC ACG CCA CTA AC	[140]
hcaT1	F	Amplification of *hcaT* for qPCR	GCT GCT CGG CTT TCT CAT CC	[140]
hcaT2	R	Amplification of *hcaT* for qPCR	CCA ACC ACG CTG ACC AAC C	[140]
dld1	F	Amplification of *dld* for qPCR (Sabri)	AGC ACC CTG CGT CTC GAC AAG C	[88, 139]
dld2	R	Amplification of *dld* for qPCR (Sabri)	CAC GAC GAT CCA ATC ACC GAG TGC	[88, 139]
cysG1	F	Amplification of *cysG* for qPCR	TTG TCG GCG GTG GTG ATG TC	[140]
cysG2	R	Amplification of *cysG* for qPCR	ATG CGG TGA ACT GTG GAA TAA ACG	[140]

Site-specific integration onto the *E. coli* K-12 chromosome was attempted using λ-red recombineering, though it was ultimately unsuccessful. We suspect that a 6.1 kb fragment was too large for efficient integration. P1 transduction between strains W and K-12 was also

unsuccessful due to lack of homology surrounding the *csc* operon.

3.4.2 Tn5 integration and library screening

Tn5 transposase enabled integration of the *csc* operon into *E. coli* K-12. Both pBKA and pAKB were digested overnight with *ndeI* to linearize. The *cscBKA*/*cscAKB* operon and flanking *kanR* (5.2 kb) were amplified with 13 bp mosaic ends. The fragments were digested overnight with DpnI and electroporated into prepared *E. coli* K-12 MG1655 component cells using Tn5 transposase (Lucigen). Integrants were recovered for 1 hour in SOC liquid medium and plated on MacConkey Agar with 50 μg/mL kanamycin sulfate. The remaining liquid recovery was diluted into 100 mL of M9 sucrose medium and 50 μg/mL kanamycin sulfate, grown for 16 hours at 37 °C, centrifuged, resuspended in 50% glycerol, and cryopreserved. After 24 hours of growth on agar, integrants were randomly selected and grown overnight in 96 microwell plates (M9 20 g/L sucrose medium with kanamycin). Cultures were then passaged to new microwell plates for growth characterization with controlled initial cell concentration. Kinetic growth data were analyzed using a custom MATLAB script that identifies the most linear portion of a log-linear plot of optical density and time. Generally, this fell between optical densities of 0.05 and 0.25 (600 nm). Representative integrants across *cscBKA* and *cscAKB* growth rate distributions were selected for cryopreservation and further characterization.

3.4.3 Footprinting select integrants

Selected *csc* integrants were grown overnight in LB medium with kanamycin for gDNA extraction (Invitrogen). Harvested DNA was quantified and 2 μg was treated with endonuclease CviA II at room temperature for 2 hours. Digested gDNA was ligated using Quick Ligase (New England Biolabs) to an annealed Y-linker (CviAII-YAT3 with CviAII-YAT5) at room temperature for 10 minutes. The reaction was terminated using EDTA. Lastly, ligated fragments were removed from solution with paramagnetic beads (Axygen Biosciences)

designed for small amplicons. Fragments were washed repeatedly with 80% ethanol to remove residual unligated Y-linker and eluted with TE buffer. Samples were then amplified using Phusion Polymerase (New England Biolabs) and phosphorylated primers; one primer annealing to the Y-linker and one annealing to the *kanR* gene, specifically annealing 3' of the 13 bp mosaic sequence of the transposon. Sanger sequencing yielded high fidelity unique sequences that varied in length between 53 bp and approximately 540 bp.

3.4.4 Quantifying gene expression using qPCR

Overnight cultures of strains were inoculated into 10 mL M9 sucrose media in 125 mL baffled flasks at 37 °C with kanamycin where indicated. Cultures were harvested in duplicate at four distinct times during early log phase, treated with RNAprotect Bacteria Reagent (QIAGEN) and frozen at -80 °C. After all cultures were harvested, RNA was extracted using RNA-EZ and treated with DNASE I (Thermo Fisher). Purified RNA was then treated with Turbo DNASE to remove any residual DNA contamination. Samples were analyzed for concentration and purity ($\lambda_{260nm}/\lambda_{280nm}$) using nanodrop (Nanodrop 1000 spectrophotometer, Thermo Fisher). Next, RNA was reverse transcribed into cDNA using Multiscribe Reverse Transcriptase (Thermo Fisher) and stored at -20 °C for qPCR expression analysis. RT-qPCR was conducted using 7900HT Fast Real-Time PCR System (Thermo Fisher) and PowerUp SYBR Green Master Mix (Thermo Fisher) with an annealing temperature of 60 °C.

Transcriptomic studies have indicated that *hcat*, *ssrA*, and *cysG* are satisfactory reference genes for qPCR normalization in *E. coli* K-12 MG1655 [140]. Gene expression in *E. coli* W has been studied to a limited extent. Previous studies have used *dld*, which encodes a D-lactate dehydrogenase, as a housekeeping gene for *E. coli* W [88,139]. The *dld* sequence in *E. coli* K-12, however, is significantly different. As a result, a brief analysis of reference gene candidates was necessary.

We initially studied the linearity of the housekeeping gene expression with respect to

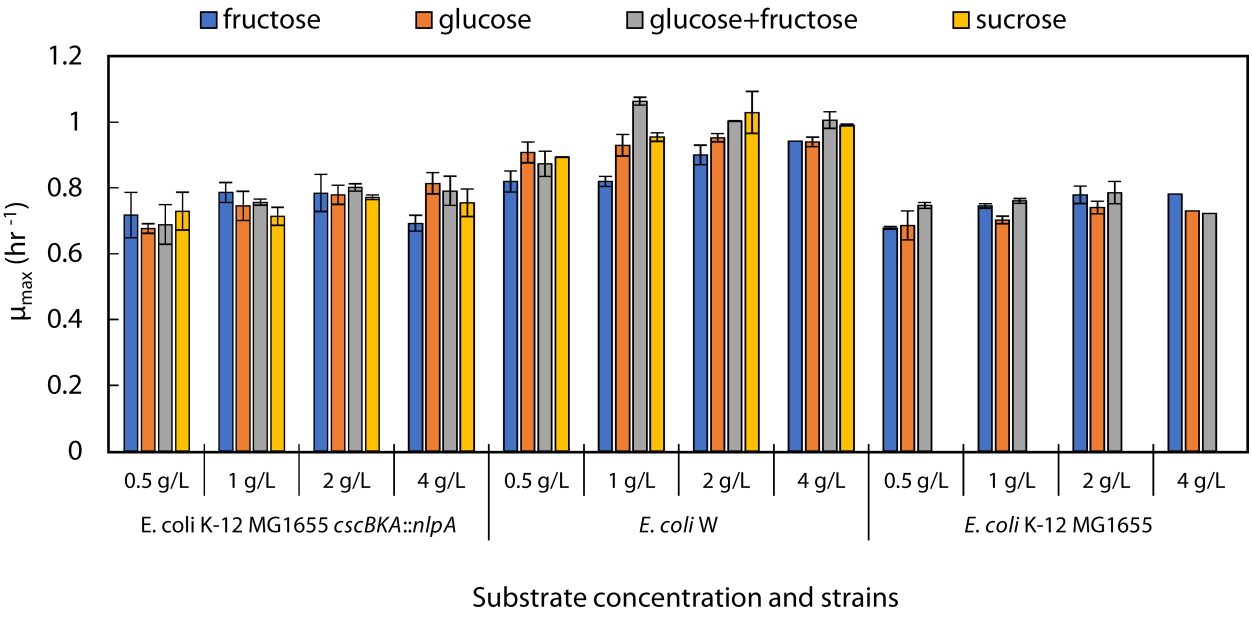


Figure 3.9: Investigation of three candidate housekeeping genes in *E. coli* strains W and K-12 for relative expression quantification via qPCR. Top row: relationships between total mRNA concentration and culture OD_{600}. Rows 2-4: relationships between cycle threshold number and culture OD_{600} for *dld* (d-lactate dehydrogenase), *cysG* (siroheme synthase), and *ssrA* (tmRNA). Each symbol represents one of two biological replicates. Error bars represent standard deviations of qPCR technical replicates (n = 3). Gene *dld* was only investigated for strain W, as it was employed in previous studies of this strain but the sequence in K-12 is significantly different rendering PCR amplification using the same primer set impossible

total RNA concentration and absorbance (λ = 600 nm) of culture at harvest. While RNA concentration and optical density values were strongly correlated for all strains ($R^2 > 0.94$), correlations between RNA concentration and cycle threshold varied significantly (Fig. 3.9). Housekeeping gene *ssrA*, which encodes tmRNA, was the best candidate for comparison between W and K-12.

Relative expression was calculated using intrastrain expression of *ssrA* and the gene of interest (*cscA*, *cscB*, *cscK*, *kanR*). A simple linear regression between cDNA gene amplification and serial dilutions between 2^{-1} to 2^{-7} indicated a strongly linear relationship ($R^2 > 0.97$).

3.4.5 Transduction of fast-growth rate locus into strains of interest

Transductions were completed using the P1 bacteriophage. Overnight K-12 B2 cultures were infected with P1 and added to Top agar. After overnight growth, Top agar was removed using a cell spreader, centrifuged, and treated with chloroform. The supernatant was then extracted for infection of target strains. ALE strains were provided by M. Antoinewicz [137]. Following standard P1 transduction protocol, ALE strains were infected with P1 for 30 minutes at 37 °C without shaking. Infection was terminated with the addition of 1 M Na_2Citrate and recovered for 3 hours at 37 °C. The culture volume was then spread on MacConkey agar plates with 50 μg/mL kanamycin sulfate, 20 g/L sucrose, and 1 mL of 1 M Na_2Citrate solution (P1 Plates). Plates were dried were incubated overnight at 37 °C. Transductants were screened by sanger sequencing of the *pyrE/rph* region for detection of an 82 bp deletion. Growth on sucrose and presence of this deletion confirmed successful transduction.

Transduction of the *nlpA* locus into C7 CIChE was also accomplished by infecting a C7 CIChE overnight culture with P1 phage harboring K-12 B2 DNA, recovery for 3 hours at 37 °C with the addition of sodium citrate, and subsequent plating on P1 Plates. Saleski et al. developed a chromosomally expressed isobutanol pathway using Tn5 integration of *ilvCD* and *alsS* followed by chemically inducible chromosomal evolution (CIChE) of the *adhA* and *kivD* genes. A novel strategy, syntrophic co-culture amplification of production (SnoCAP) was used to screen for phenotype. Although C7 CIChE already has kanamycin resistance colonies could still be colorimetrically identified. Transductants were repeatedly restruck on P1 Plates to account for possible contamination by the C7 CIChE base strain.

3.4.6 Isobutanol production experiments

An overnight culture of each isobutanol producing strain was passaged 1:100 to 20 mL M9 IPG (125 mL baffled flasks) at 37 °C and 250 rpm in a rotary shaker. M9 IPG contains M9 Minimal Salts (Sigma-Aldrich), 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 2.91 $(NH_4)2MoO_4$, 401.4

nM H_3BO_3, 30.3 nM $CoCl_2$, 9.61 nM $CuSO_4$, 51.4 nM $MnCl_2$, 6.1 nM $ZnSO_4$, 0.01 mM $FeSO_4$, 3.32 μM thiamine HCl, 36 g/L carbon source, and 5 g/L yeast extract. After 2 hours, 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) was introduced. Flasks were transferred to a 30 °C incubator shaker also maintaining 250 rpm. Samples were taken immediately and every 24 hours thereafter. Cell concentration was quantified using optical density (600 nm) in a Synergy H1 spectrophotometer (BioTek). Isobutanol and sugar concentrations were quantified using high performance liquid chromatography (Shimadzu) fit with a Rezex ROA-Organic Acid column (Phenomenex).

3.4.7 Assessment of sustainability metrics

Media GHG and energy calculations were calculated using the Greenhouse gases, regulated emissions, and energy in transportation (GREET) tool published by Argonne National Laboratory [141]. Only medium associated metrics were evaluated as all other process characteristics between glucose and sucrose biofermentation are identical. While a full life cycle assessment would normalize the net sustainability benefits of each media, substrate costs account for as much as 50% of total operating costs of a biofermentation scheme [142].

Chapter 4: Biodiversity Improves Life Cycle Sustainability Metrics in Algal Biofuel Production

Algal biofuel has yet to realize its potential as a commercial and sustainable bioenergy source, largely due to the challenge of maximizing and sustaining biomass production with respect to energetic and material inputs in large-scale cultivation. Experimental studies have shown that multi-species algal polycultures can be designed to enhance biomass production, stability, and nutrient recycling compared to monocultures. Yet, it remains unclear whether these impacts of biodiversity make polycultures more sustainable than monocultures. Here, we present results of a comparative life cycle assessment (LCA) for algal bio-refineries to compare the sustainability metrics of monocultures and polycultures of six fresh-water algal species. Our results showed that when algae were grown in outdoor experimental ponds, certain bicultures improved the energy return on investment (EROI) and greenhouse gas emissions (GHGs) by 20% and 16%, respectively, compared to the best monoculture. Bicultures outperformed monocultures by performing multiple functions simultaneously (*e.g.*,. improved stability, nutrient efficiency, biocrude characteristics), which outweighed the higher productivity attainable by a monoculture. Our results demonstrate that algal polycultures with optimized multi-functionality lead to enhanced life cycle metrics, highlighting the significant potential of ecological engineering for enabling future environmentally sustainable algal bio-refineries.

The majority of the work presented in this chapter was published in ACS *Environmental Science & Technology* as follows: Carruthers, David N., Casey M. Godwin, David C. Hietala,

Bradley J. Cardinale, Xiaoxia Nina Lin, and Phillip E. Savage. "Biodiversity improves life cycle sustainability metrics in algal biofuel production." *Environmental science & technology* 53(15) (2019): 9279-9288.

4.1 Introduction

Microalgae are promising feedstock candidates for renewable biofuel production due to their high photosynthetic efficiency, lipid content, and growth rates compared to terrestrial plants [143,144]. Algae-based fuels have the potential to increase energy security, reduce fuel cycle carbon emissions, and ultimately replace petroleum-based transportation fuels. But despite decades of intensive research and development, industrial algal biofuel production has been unsuccessful [44]. One major bottleneck is the difficulty in achieving the observed productivity and biomass characteristics from laboratory-scale studies in outdoor open pond raceways, which are generally deemed necessary for economic viability [48, 65]. In practice, production at scale has been hampered by low biomass yields, poor biomass quality, invasion by grazers, diseases, pond crashes [145] and unwanted algal species [146], which is a major barrier to achieving sustainability and profitability of algal biofuel production. Genetic engineering of algae and intensive management strategies such as chemical treatment with fertilizers or pesticides have shown some potential for overcoming these challenges; however, they tend to yield negative trade-offs due to either decreased algal growth or increased cost, and most have not been successfully demonstrated at scale [147–149].

In terms of biofuel production, algal communities that increase overall biomass yield, as opposed to just lipid content, are ideal candidates for hydrothermal liquefaction (HTL), which converts whole algal biomass to biocrude. Ecological engineering of multi-species polycultures has been proposed as an alternative means to overcome the challenges associated with algal cultivation. Based on extensive research on the role of biodiversity in natural ecosystems, we know that multi-species communities of plants or microalgae can yield more biomass, use abiotic resources more efficiently, and better resist threats such as disease,

grazers, and invasive species compared to a single species [50, 150]. Laboratory and field studies have tested this hypothesis experimentally and found that diversity has important impacts in terms of reducing temporal variance (increased stability), nutrient use, fuel quality, and the potential to optimize multiple functions simultaneously [49]. However, these multi-species communities rarely increase biomass yields relative to the most productive constitutive single species of the community [49]. While these studies suggest that ecological engineering of microalgal communities is a promising strategy to improve large-scale biofuel production [151], it remains unclear whether these improvements make polycultures more sustainable than monocultures in terms of environmental impacts. Answering this question requires a life cycle assessment (LCA) informed by data from experiments comparing mono- versus polycultures. An LCA is a methodology that considers the energy and materials embodied in a theoretical life cycle in order to predict whether a given scenario will have overall favorable performance in terms of well-defined sustainability metrics such as energy return on investment (EROI) and greenhouse gas emissions (GHGs).

Here we report results of a well-to-wheel LCA that analyzed the impact of algal biodiversity on biofuel production using both laboratory and outdoor pond cultivation data. The cohesive and comprehensive nature of the dataset closes the knowledge gaps between algal cultivation, biomass conversion, and life cycle assessment, minimizing concerns of inconsistency across models and increasing industrial practicality for a more holistic algal LCA. Our results show that certain multi-species communities can outperform the best monocultures in terms of sustainability metrics under conditions mimicking real-world cultivation due to the polycultures' superior potential for performing multiple functions well.

4.2 Materials and methods

4.2.1 Algal pond growth

The data used in this work stem from two published studies that have examined the role of algal biodiversity in biofuel production in both laboratory-based mesocosms [37,61,62,150], and field-based open pond systems [49]. For both of these studies, focal algal species were selected from among the high priority species determined by the U.S. Department of Energy's Aquatic Species Program and included within the Solar Energy Research Institute's microalgae collection [152]. We selected six of these species: *Ankistrodesmus falcatus* (code A), *Chlorella sorokiniana* (B), *Pediastrum duplex* (C), *Scenedesmus acuminatus* (D), *Scenedesmus ecornis* (E), and *Selenastrum capricornutum* (F). Algal cultures were grown photoautotrophically in BOLD 3N medium sparging air without carbon dioxide supplementation. The laboratory mesocosms (n = 180) were 9.5 L aquaria illuminated with fluorescent lights and diluted with 30% fresh medium every 7 days. Two distinct temperature treatments were employed to analyze the effect of temperature fluctuation on algal growth. The temperature of the tanks was either constant at 22 °C or variable between 17 °C and 27 °C on weekly cycles. Lab-based mesocosm experiments studied 37 combinations of algal species in two randomized blocks for each temperature treatment. Monocultures were replicated 6 times, bicultures 4 times, four-species cultures 4 times, and the six-species polyculture replicated 9 times for each temperature condition [62]. Mesocosms were grown for two weeks and then measured every seven days for a subsequent eight weeks. The outdoor ponds experiment was performed using 80 open ponds, each 1,100 L in volume and 0.5 m in depth, at the University of Michigan's Edwin S. George Reserve, located near Pinckney Michigan [49]. We used the four most productive species based on the laboratory mesocosms (A, B, D, and F) as monocultures, 6 two-species polycultures, and the four-species polyculture [49]. The open ponds experiment lasted for over 80 days, during which we measured biomass concentration weekly. Ponds were subject to daily temperature fluctuations, weather, and invasion by

unwanted species of algae, diseases, and grazers. A stochastic growth model was developed from empirical data to determine biomass concentration at harvest and areal productivity (Fig. 4.1). Harvest fractions were set to 70% at a weekly interval with a crash threshold of 50 mg biomass after harvest for determining the number of annual pond crashes (Fig. 4.2).

4.2.2 Stability

Despite repeated designation as a key technological barrier to algal biodiesel production, stability issues have only recently been considered in the design of algal biofuel systems. Stability of the feedstock cultures is particularly important for outdoor ponds, where biological and environmental forcings inevitably make biomass and productivity varied through time [153]. Algal culture stability is quantified as the coefficient of variation in biomass and productivity through time for a single culture. Factors that influence stability include environmental fluctuations, invasion by algal pest species, algal diseases, and grazers, as well as pond harvesting. Traditionally, pond productivity measurements are determined using a simple linear relationship:

$$Q_A = \mu * X * d \tag{41}$$

Where X is volumetric biomass density at harvest, d is pond depth, and μ is the specific growth rate of the algal culture. Areal productivity, Q_A, has the unit of $g/m^2/day$, which is convenient for pond sizing and infrastructure amortization. However, Equation 41 is a biologically absurd representation of algal growth as it ignores density dependence – that is, that per capita growth rates decrease as biomass density increases. In open ponds, density dependence arises because more dense cultures attenuate light required for photosynthesis and sequester nutrients required for growth, which has been discussed elsewhere [154]. Furthermore, Equation 41 allows for growth rates and cell densities to be independent of pond depth, which is not only implausible due to light attenuation, but implies that simply increasing the depth of a pond can lead to arbitrarily high areal productivity. Lastly,

Equation 41 ignores temporal variation and instability in the production of algal biomass, both of which are commonly observed risks associated with large-scale algal production with potentially monumental implications, particularly in cases where cultures crash. In lieu of Equation 41, we introduce a biologically realistic model of density-dependent algal growth that is based upon empirically derived maximum growth rates, maximum densities, and temporal variation.

$$B_t(t) = \frac{K}{1 + \left(\frac{K - B_0}{B_0}\right) * e^{-\mu * t}} \tag{42}$$

Here, K represents the carrying capacity, B_0 the initial biomass, and μ the specific growth rate. For the laboratory mesocosms experiments, we assumed a conservative maximum growth rate of 0.5 d^{-1} for all species compositions. For the outdoor pond experiment, we estimated maximum growth rates using time-series of in vivo chlorophyll-*a* fluorescence immediately following inoculation. For both the laboratory mesocosms and the outdoor ponds, we used a first-order autoregressive stationary model to simulate the carrying capacity K over time. The autoregressive model was parameterized using time-series of biomass for individual mesocosms or ponds. For each species composition, we performed 10,000 simulations of the growth over a. Each run was based on the autoregressive model from a single laboratory mesocosm or outdoor open pond, selected at random from the data. We modeled the biomass of the cultures at time steps of one week, for a total of 52 weeks. The carrying capacities for each week were generated from the autoregressive model, meaning that the value of K is auto-correlated through time. A threshold of 50 mg/L biomass defines a culture crash. If a culture fell below 50 mg/L at harvest, the pond would necessarily be drained, cleaned, and reinoculated, requiring an estimated two-week down time between culture crash, pond regrowth, and subsequent harvests. Culture biomass is average biomass accumulation over 52 weeks at harvest, intrinsically including increased water demand associated with pond crashes. A revised algorithm was developed to account for this productivity loss:

$$Q_A = h * \sum_{t=3}^{52} B'_t * d \tag{43}$$

$$B'_t = \begin{pmatrix} B_t & B_t \geq B_{threshold} \\ 0 & < B_{threshold} \end{pmatrix} \tag{44}$$

Where h is the harvest fraction, d is pond depth, and Q_A is the predicted areal productivity. In this work, we assumed a fixed harvest regime in which 70% of the culture is harvested every 7 days.

Parameterizing time-series models from the empirical data

For both the laboratory mesocosms and outdoor open ponds, we used replicated time series of algal biomass to parameterize a first-order autoregressive stationary model. This calculates estimates of the mean biomass, the 1st order autocorrelation coefficient, and the variance. Since these models are based on biomass at saturation, their output can be equated to carrying capacities that fluctuate through time. The first-order correlation is not only more realistic, but directly impacts the modeled productivity [155].

Simulating density-dependent growth with varying carrying capacities, harvests, and crashes

Density-dependent growth of the algae was modeled using the logistic growth function in Equation 42. For each species composition, we performed 10,000 simulations of the growth in a pond over a year. Each run was based the autoregressive model from a single aquarium or outdoor pond, selected at random from the available data. We modeled the biomass of the cultures at time steps of one week, for a total of 52 weeks. The carrying capacities for each week were generated from the autoregressive model, meaning that the value of K is auto-correlated through time. In some cases, these autoregressive models forecast carrying

capacities less than zero. Because this is biologically unrealistic, we used a floor of 10 mg/L for carrying capacities. The initial biomass density in week 1 was 10 mg/L. From the biomass density on the first day of the week ($B_{0,t}$), we forecasted the biomass density after 7 days using the logistic growth model. Then, depending upon the biomass density, the culture was either harvested or declared crashed. Cultures were not harvested in weeks 1 through 3 to allow for initial growth. This scheme is depicted in Fig. 4.1 below.

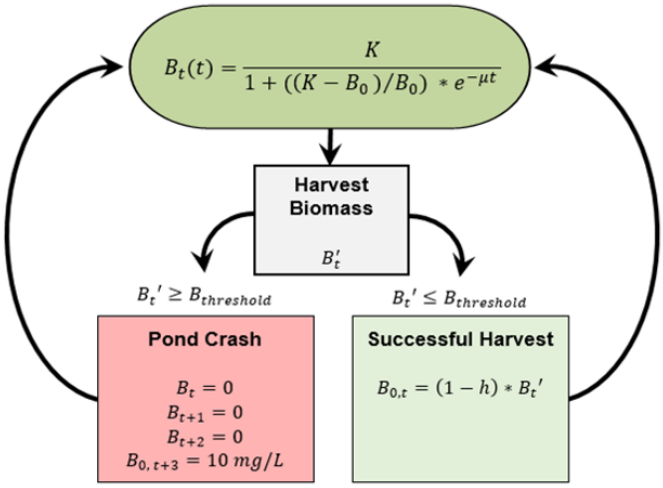


Figure 4.1: Flow chart for determination of algal productivity and number of crashes per year.

When the culture is harvested, the biomass density is decreased by a fixed proportion (*e.g.*,. 70%) immediately before the next week, and the logistic model is used again. A 70% harvest fraction was selected through the sensitivity analysis of harvest proportion and harvest time as depicted in Fig. 4.2. If the biomass density is below the threshold at the end of the week, the culture is declared crashed and there is no harvest. The crash threshold employed in this study was 50 mg/L density after harvest. When a crash occurs, the biomass is zero for the week in which the crash occurred and the next two weeks. This time represents inactivity while the pond is drained and re-established. After three weeks of zero biomass, the culture is re-started with an initial biomass density of 10 mg/L, and the simulation continues.

Productivity from density-dependent growth model:

The mean annual productivity (mg/L/week) of a simulated pond is equal to the product of the harvested proportion (*e.g.*,. 0.7) and the mean of all biomass densities above the threshold. By assuming a pond depth of 0.17 m for the laboratory mesocosms and 0.5 m for the cattle tanks, we get a mean annual productivity in units of $g/m^2/day$ as dry mass. For each simulated year, we also counted the number of discrete crash events and the total number of weeks that the culture was crashed. Fig. 4.2 provides an example of this harvest strategy over a year with 5 crash events.

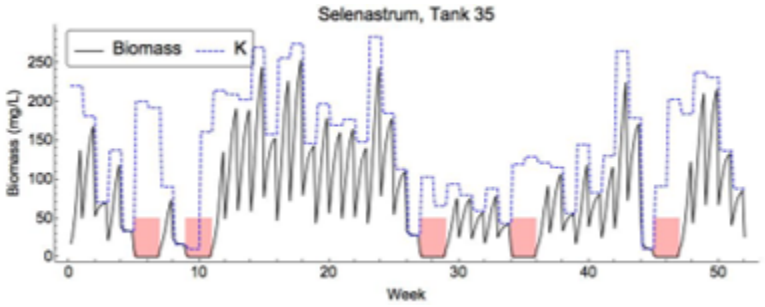


Figure 4.2: Modeled biomass through time based on open pond number 35 of *Selenastrum capricornutum* assuming a weekly harvest of 70%. The blue dashed line shows the carrying capacity through time and the black line represents biomass. When the culture biomass is less than 50 mg/L at the end of a seven-day growing period, the culture is declared crashed shown in pink.

4.2.3 Hydrothermal liquefaction

Many strategies have been developed to produce usable fuel from algae, ranging from transesterification of fatty acid methyl esters, liquid-liquid extraction, and hydrothermal liquefaction (HTL). More recently, focus has shifted towards HTL of algal biomass, a strategy that employs high temperature (200-400 °C) and pressure (2-20 MPa) to convert whole algal biomass into biocrude rather than only lipids into biodiesel [156]. The co-products of HTL include solid biochar, nutrient-rich aqueous phase, and gases. Laboratory HTL experiments were conducted using harvested biomass at 5% dry mass loading at 350 °C for 20 minutes [61]. The 5% mass loading was scaled to the 15% from algal dewatering for more energetically

favorable biomass conversion. Slurries above 15% demonstrate non-Newtonian behavior at operating temperatures and were not considered [40]. Heating values of biocrude samples were calculated from elemental composition using the Channiwalla formula [157]. Upgrading methodology then followed Frank et al. (2013) for carbon mass balance efficiency during deoxygenation and denitrogenation, which differ between cultures due to variable biocrude oxygen and nitrogen content [158]. Upgraded biocrude is assumed suitable as a diesel fuel.

4.2.4 First and remaining dewatering

The process flow includes open pond growth and harvesting, followed by an initial dewatering that employs dissolved air flocculation with chitosan coagulant to 60 g/L. A disc-stack centrifuge raises the total solid mass fraction of the algal slurry to 150 g/L, appropriate for HTL. Equations 43 and 44 inherently increase water demand from pond crashes, but do not account for pumping new media into the ponds. As a result, crash recovery media was incorporated as a function of annual average pond crashes in kWh/g_{afdw} algae. Approximately 95% of total process volume resides within the growth and dewatering stages such that most recycling flows are quantified within these first two stages.

4.2.5 Recycling

Flows of carbon, nitrogen, phosphorus, and water are depicted in Fig. 4.3 within the process framework. Values denote normalized nutrient cycle across the entire pool of cultures on a per gram afdw algae basis. Nitrogen and phosphorus recycling were 76 ± 4% and 80 ± 9% respectively, with 48 ± 5% total biomass carbon in the biocrude. First dewatering and remaining dewatering provide a net return of 70 ± 7% total water, with approximately 26 ± 6% water lost due to evaporation. Overall, our empirically-derived nutrient recycling rates corroborate those found throughout the literature [65, 159], but there are substantial differences among the different species compositions in terms of their elemental content and the proportion of biomass nutrients that can be recovered following HTL.

Many strategies have been pursued at length to demonstrate the viability of the aqueous co-product (ACP) as a valuable organic substrate for either heterotrophic algal regrowth or bacterial regrowth [22, 37, 159], Previous work has shown that recycling of the ACP to the algal cultures is feasible at the dilution factor that would arise in our process (approximately 0.5% total pond volume) [37], While recycling of ACP by polycultures can lead to increased growth compared to the best monocultures, this effect has only been documented in small-scale laboratory cultures and thus we would not be able to parameterize its effect on biomass production or reactivity during HTL [37]. As heterotrophic growth is not representative of cultivation conditions in this study, we instead considered co-product treatment using a recently patented strategy by Genifuel Corporation pursuing catalytic hydrothermal gasification (CHG) of the aqueous phase at 300-350 °C and 10-20 MPa. CHG involves relatively complex chemistry including reactions of pyrolysis, steam-reformation, methanation, hydrogenation, and the water-gas shift to produce methane, hydrogen gas, water, and carbon dioxide [45]. This strategy has produced yields at estimated 0.32 L $CH_4/g_{ACPsolids}$, combustion of which provides substantial benefit to energy return and has been incorporated in assessments from Argonne National Laboratories [158, 160]. Integration of a CO_2 capture system could reduce gross GHG emissions and likely increase biomass accumulation across experimental conditions, this again was not considered because the experiments that inform this LCA did not use CO_2 supplementation. It was further assumed that any residual phosphorus and nitrogen in the biocrude cannot be recycled as part of the upgrading process. Although not considered in this study, hydrogen gas co-product from CHG could be recycled for upgrading HTL biocrude [144].

4.2.6 Infrastructure

Infrastructure data was sourced from Lundquist et al. and Canter et al. amortized to pond surface areal productivity and denoted on a per gram afdw algae basis [155, 161]. Although these studies consider algal lipid extraction, infrastructure energetic and environ-

mental accounting is dominated by open pond construction, which accounts for an average of 85% of total infrastructure energy emissions [161]. Infrastructure was scaled linearly with areal productivity as in previous studies. Transportation of CO_2 via flue gas bubbling was not considered within this framework in congruence with experimental conditions. Future mechanistic study and ecological engineering of diverse algal polycultures with optimized multi-functionality could provide an alternative pathway towards environmentally sustainable and economically viable algal biofuels.

4.2.7 Life cycle assessment overview

A computational module of algal cultivation and life cycle inventory was established based on our empirical data, which we called the Algal Hydrothermal liquefaction Module (AHM). The AHM is broken down into six stages: algal pond growth, first dewatering, remaining dewatering, biomass conversion, upgrading, and infrastructure. The biomass conversion phase is divided into hydrothermal liquefaction (HTL) and catalytic hydrothermal gasification (CHG). To compare algal polycultures and monocultures, we employed the AHM to simulated a hypothetical bio-refinery system (Fig. 4.3a), using a large experimental dataset generated previously through multidisciplinary work involving six algal species (Fig. 4.3b) that were grown alone or in mixed species consortia in replicated laboratory mesocosms (Fig. 4.3c) and open outdoor ponds (Fig. 4.3d) [25, 37, 61, 62].

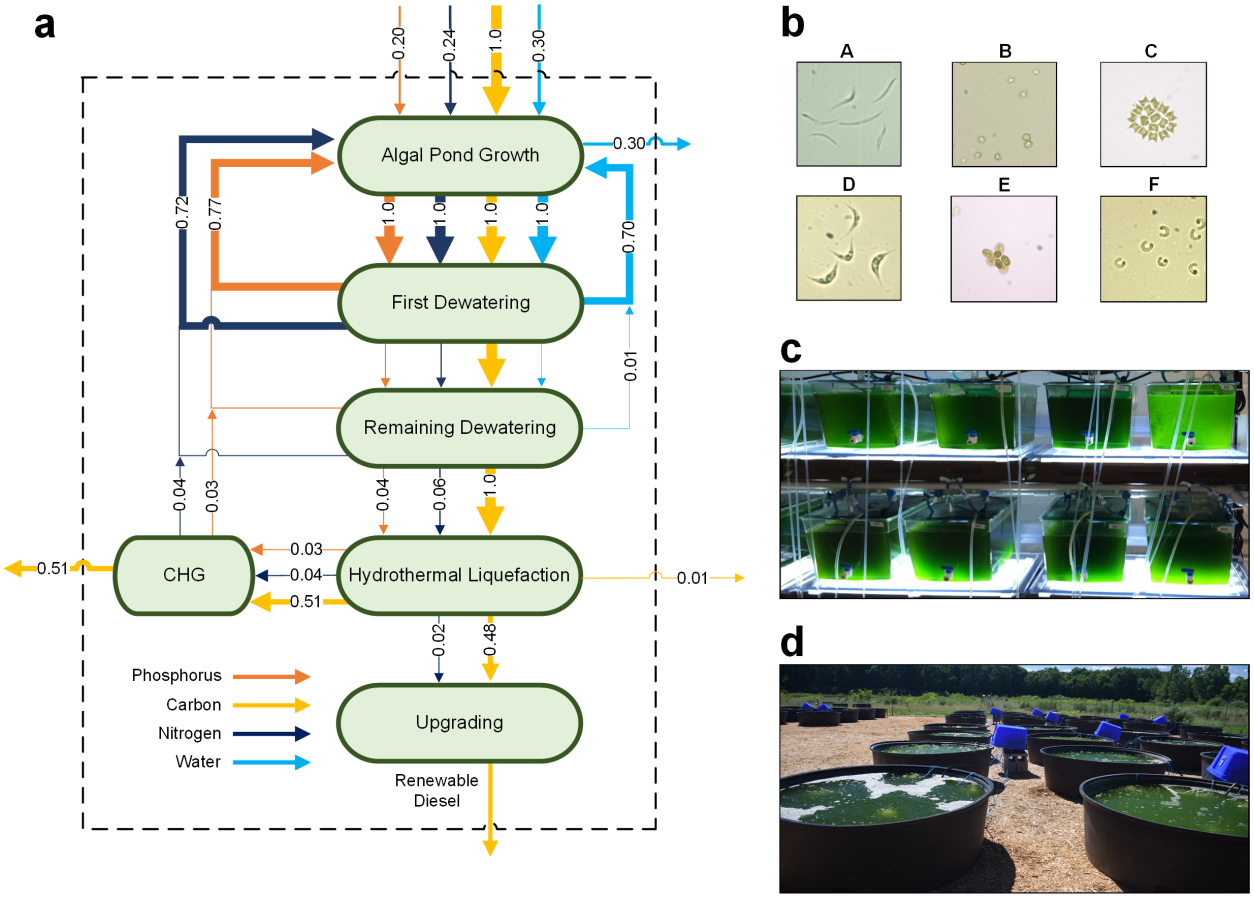


Figure 4.3: Life cycle assessment (LCA) of algal biofuel production using data from laboratory and outdoor cultivation experiments. (A) Schematic of hypothetical algal biorefinery in LCA, with flows of phosphorus, carbon, nitrogen, and water depicted as fractions of the amount leaving the first unit of Algal Pond Growth. Values shown, corresponding to line widths, represent the proportion of P, C, N, and water in the pond that flow through each step depicted in the diagram (dimensionless) and are the average fractions across all experimental conditions and culture compositions. (B) Representative microscopic images of the six microalgal species examined in this study: *Ankistrodesmus falcatus* (A), *Chlorella sorokiniana* (B),*Pediastrum duplex* (C), *Scenedesmus acuminatus* (D), *Scenedesmus ecornis* (E), *Selenastrum capricornutum* (F). (C) A photo of 9.5-L laboratory mesocosms. (D) A photo of 1,100-liter cattle tanks simulating outdoor pond cultivation.

The assessment employs the Argonne National Laboratory's GREET, a Microsoft Excel model used to study the environmental and energetic costs of many fuel and vehicle technologies [63]. For algal biofuels, GREET interfaces with a life cycle inventory, the Algal Process Description (APD), which provides general input characteristics of algal species on a per gram afdw algae basis, accounting for fertilizer demand, water consumption, electricity and other relevant resource inputs. While the APD was fundamental to initial algal LCA

development, it does not reflect biological reality due to overgeneralized parameters and assumptions. For example, the APD allows for the growth rate of the algae to be manipulated independently from their density and pond depth, though these parameters are inextricably linked due to biological principles [162, 163]. To make the process description more realistic, we developed a life cycle inventory called the Algal HTL Module (AHM) based on the APD, which tracks inputs and outputs throughout each stage of biofuel production, readily incorporating data from our collaborative experiments to GREET 2016b for LCA calculations.

The AHM calculations use a functional unit of 1 g_{afdw} algae and, where necessary, employ realistic and documented assumptions to scale experimental data to an industrial-sized facility. The AHM quantifies properties and functions of the algal biorefinery system, including biomass concentration, productivity, crop stability over time, biocrude and biomass elemental composition, efficiency of abiotic nutrient use, and biocrude yield. Fig. 4.4 illustrates the various assumed, specified, and empirical inputs as well as their incorporation in the multi-layered calculations within AHM (see Table G.4 for details).

We apply the AHM inventory and the GREET 2016b impact assessment model for comparative analysis firstly of idealized laboratory mesocosm experiments under two temperature regimes and secondly of more realistic outdoor pond experiments. Monte Carlo simulations were conducted using a truncated normal distribution from empirical data (mean and variance) to analyze error propagation over 1,000 iterations for each set of experimental conditions and species combinations in GREET 2016b. A process flow diagram of the AHM is depicted in Fig. 4.3a within a dashed system boundary. We used our comprehensive experimental dataset to parameterize elemental content at each stage, formulating high confidence mass balances for translation to an industrial sized facility. Balances for carbon, nitrogen, phosphorus, and water are denoted by the width and valuation within each flow. To provide a thorough description of the AHM, Fig. 4.4 differentiates between assumed, specified, and empirical process inputs as well as their incorporation within the GREET

impact assessment.

Values from the AHM were then utilized as inputs to the Greenhouse gases, Regulated Emissions, and Energy in Transportation (GREET) impact assessment model developed by the Argonne National Laboratory [165] using a functional unit of 1 MBTU (10^6 MBTU) transportation energy in compression ignited direct injection diesel vehicles [63]. The GREET utility is a versatile tool updated biannually that computes greenhouse gas emissions (CO_2, CH_4, N_2O), non-renewable energy usage, as well as six standard pollutants: carbon monoxide, mononitrogen oxides, sulfur oxides, volatile organic compounds, and particulate matter (PM_{10} and $PM_{2.5}$) for many fuel/vehicle systems [165]. GHGs are normalized to CO_2eq using global warming potentials established by the Intergovernmental Panel on Climate Change AR5 [166].

Operating cohesively as an inventory and impact assessment, the AHM and GREET compose the overall LCA. We focus on three metrics that are key to determining the sustainability of a transportation fuel technology and represent the output of the GREET model: energy return on investment (EROI), greenhouse gas emissions (GHGs), and water intensity (WI). Only direct effects from the algae pathway are included with indirect effects like land use changes, infrastructure preparation, water transportation to site and similar operational activities omitted [165]. Sustainability baselines are defined such that an energy return below 1.0 is a net loss of energy and thereby energetically unsustainable, whereas GHGs and WI are analyzed either as absolute values ($kgCO_2$eq and L H_2O) or as percentages relative to compression ignited direct injection diesel vehicles fueled with conventional diesel. To be considered a sustainable fuel, a biofuel must have an EROI higher than 1.0 and GHGs must be lower than conventional diesel, estimated at 99.5 $kgCO_2$eq per MBTU on a well-to-wheel basis [165].

In most cases there is a negative correlation between GHG and EROI: increasing productivity increases the return on investment and simultaneously decreases the GHG footprint by decreasing non-renewable energy consumption [167]. Rather than using arbitrary values

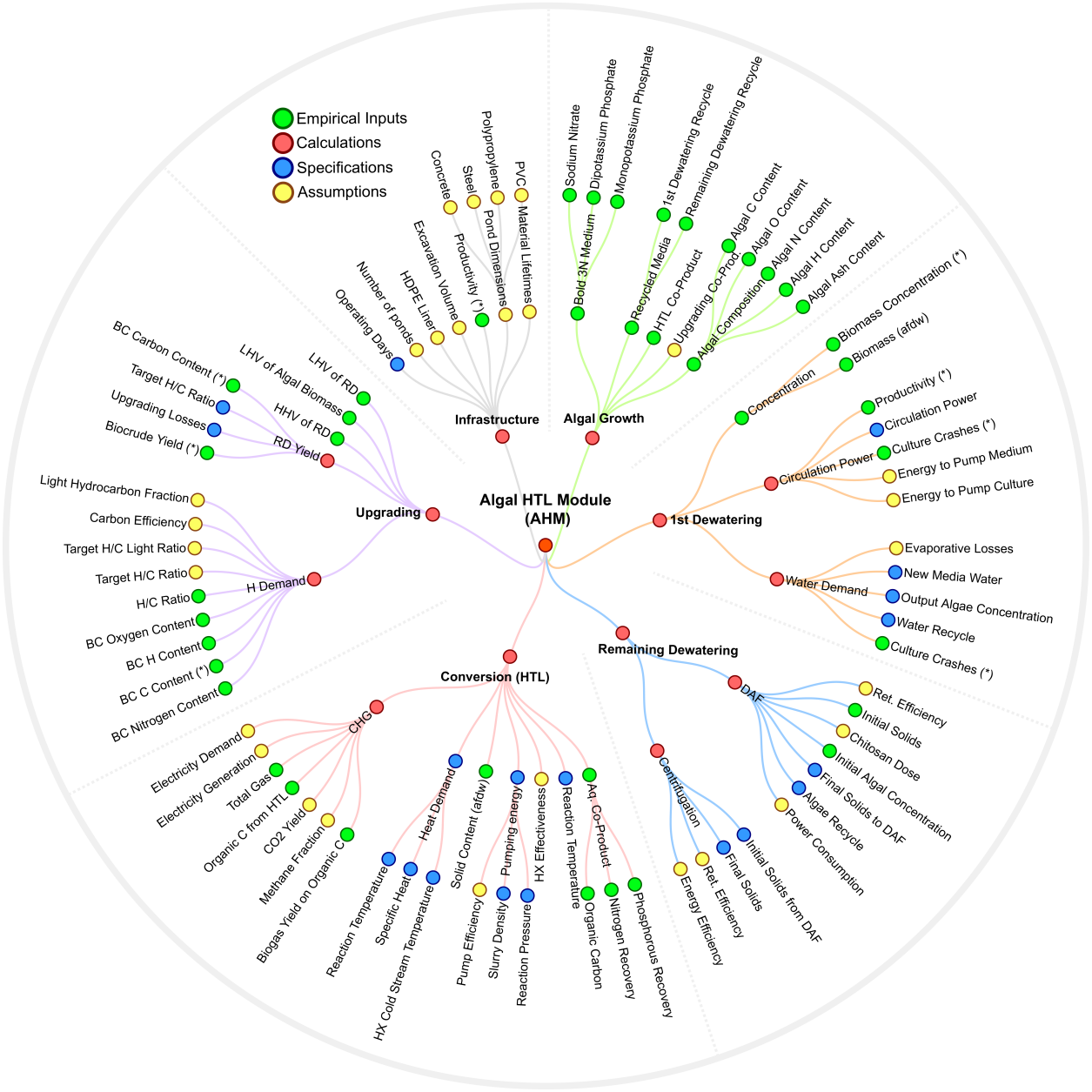


Figure 4.4: A Reingold-Tilford tree illustrating how information is sourced and processed within the AHM. The tree is subdivided into 41 empirical inputs (green), 19 process specifications (blue), and 28 assumptions (beige). Internal calculations (red) combine the values from the other categories to derive LCA metrics. Inputs marked with a (*) were used in multiple calculations. Plot developed using Data-Driven-Documents [164]

.

to conduct a sensitivity analysis of parameters, as is commonplace among LCAs, we examined empirical variation in Monte Carlo simulations using a truncated standard distribution. Using this method, values of inventory inputs are varied simultaneously for an accurate representation of the dataset. Regression coefficients were generated by using 5,000 Monte Carlo simulations with the highest EROI to quantify the influence of variation in inventory inputs on impact assessment outputs. We used this approach because life cycle metrics are inherently nonlinear at low values of inputs like productivity and biomass concentration, which reduces fit dramatically.

4.3 Results and discussion

4.3.1 Life cycle assessment of mesocosm algal growth

Based on laboratory data, a single highly productive monoculture was the top performer in maximizing EROI and minimizing GHGs

We first applied the LCA framework described above to data collected from a set of experiments where a total of 37 algal cultures, including 6 monocultures and 31 select polycultures, were grown in the laboratory at a constant temperature of 22°C and then converted to biocrude oil through HTL. Fig. 4.5 summarizes results for two LCA metrics, EROI and GHGs.

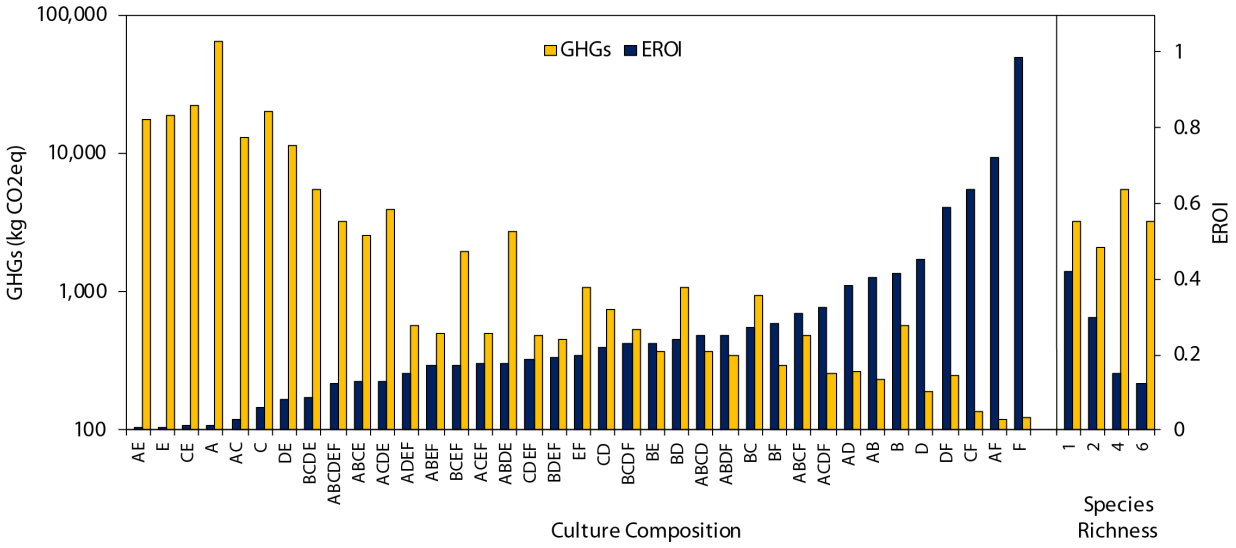


Figure 4.5: GHGs and EROI of 37 algal cultures under constant temperature laboratory conditions. GREET model return using empirical data from laboratory experiments by culture composition (sorted by increasing EROI) and by mean species richness.

In this assessment, a monoculture, *S. capricornutum* (F), and this species in bicultures with *A. falcatus* (AF) or P. duplex (CF) were the best performers, generating EROIs of 0.99, 0.72, and 0.64 as well as GHGs of 122, 120, and 135 $kgCO_2eq/MBTU$, respectively. Water intensity, which represents the net water usage per MBTU fuel produced, was determined

to be 172, 165, and 150 L/MBTU, respectively (Fig. 4.6).

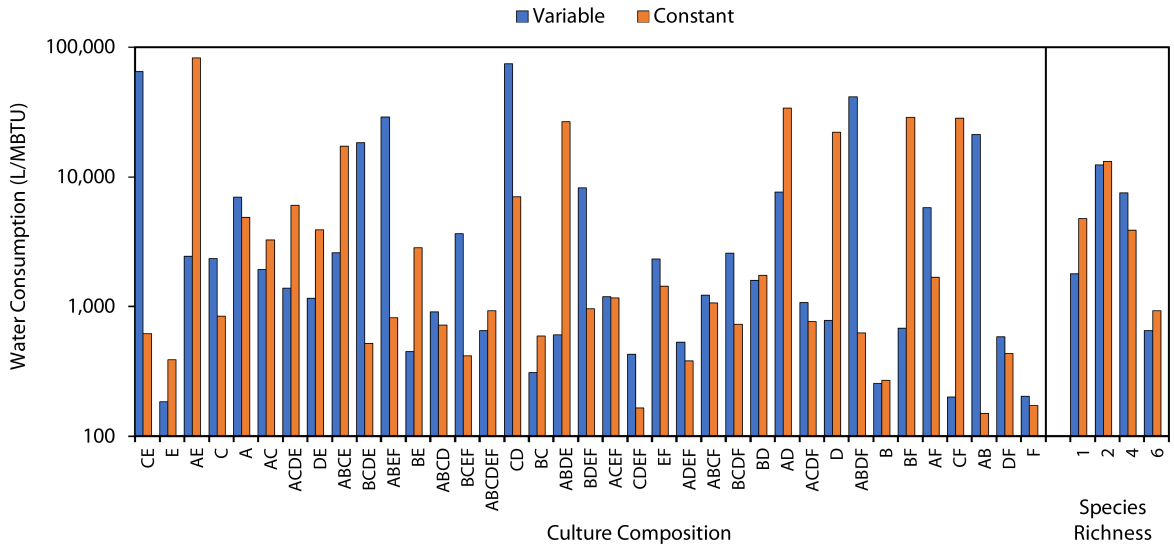


Figure 4.6: WI of algal cultures under constant and variable temperature mesocosm conditions from the GREET model by culture composition and species richness.

Closer examination of the results revealed that *S. capricornutum* (F) cultivation yielded the best sustainability metrics due to its ability to achieve a high biomass concentration of 0.37 g/L, which was nearly three times higher than the average of all cultures under these experimental conditions (0.13 ± 0.08 g/L), and substantially better than the second-best culture AF (0.30 g/L). In another set of laboratory experiments, the same algal cultures were grown under variable temperature conditions, where the temperature cycled between 17 and 27 °C on a weekly basis [62]. We also performed an LCA based on data collected from these experiments (Fig. 4.6 4.7) and our results followed the same general pattern obtained using constant temperature data, with the best performers in EROI again being *S. capricornutum* and its bicultures.

It was noted that the variable temperature condition had no significant impact on biomass concentration, biocrude yield, or stability [62]. Nonrenewable energy usage was partitioned by stage of production to identify areas for optimization and illustrated in Fig. 4.10.

Algal growth and first dewatering dominated process energetics with a minimum fraction of approximately 60% of the total energy for *S. capricornutum*. The subsequent factors

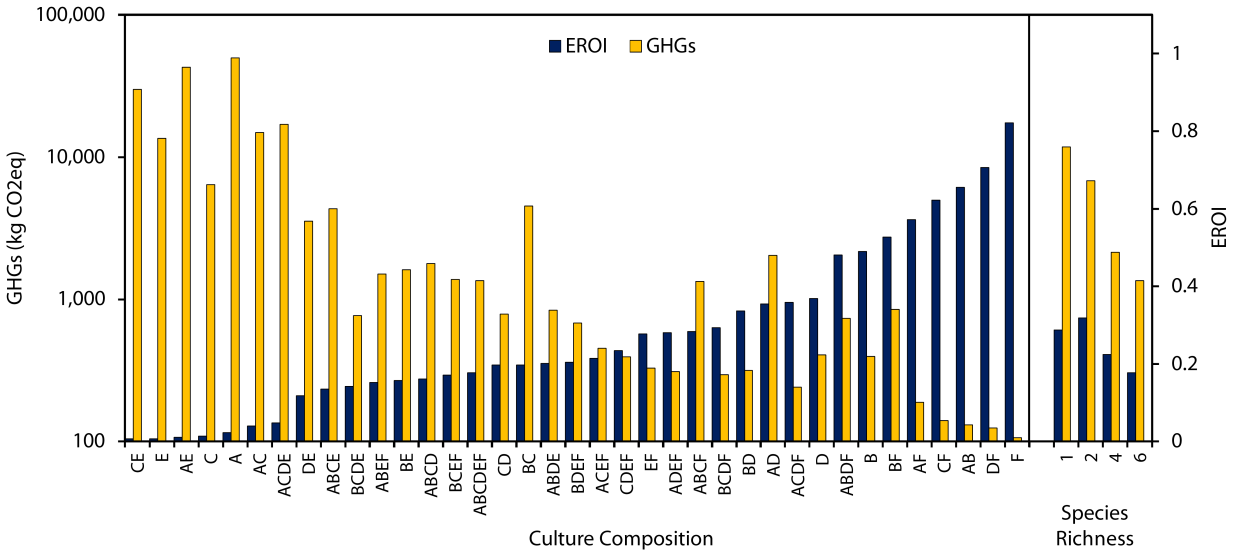


Figure 4.7: GHGs and EROI of 37 algal cultures under variable temperature conditions from the GREET model by culture composition and species richness.

were infrastructure (14.3%), transportation and distribution (11.3%) and biocrude conversion (4.1%). These results confirm that even for the best performing algal species, the greatest gains to advancing biofuel sustainability could be made by optimizing cultivation conditions and minimizing upstream energy inputs [168].

4.3.2 Life cycle assessment of outdoor pond algal growth

Select bicultures outperformed the top monoculture in outdoor ponds

We next used data collected from an outdoor cultivation experiment, where a total of 11 promising algal cultures, including 4 monocultures, 6 bicultures, and 1 four-species culture, were grown in 1,100 L open ponds (6-8 replicates). It should be pointed out that scaling of experimental data to large-scale cultivation introduces uncertainties, particularly related to how parameters like productivity and biomass accumulation change in actual raceway pond cultivation compared to experiments [169]. Here, we have made an effort to minimize these uncertainties by incorporating empirical data from experiments performed under conditions that mimic real-world conditions. Specifically, the outdoor pond cultures were subject to

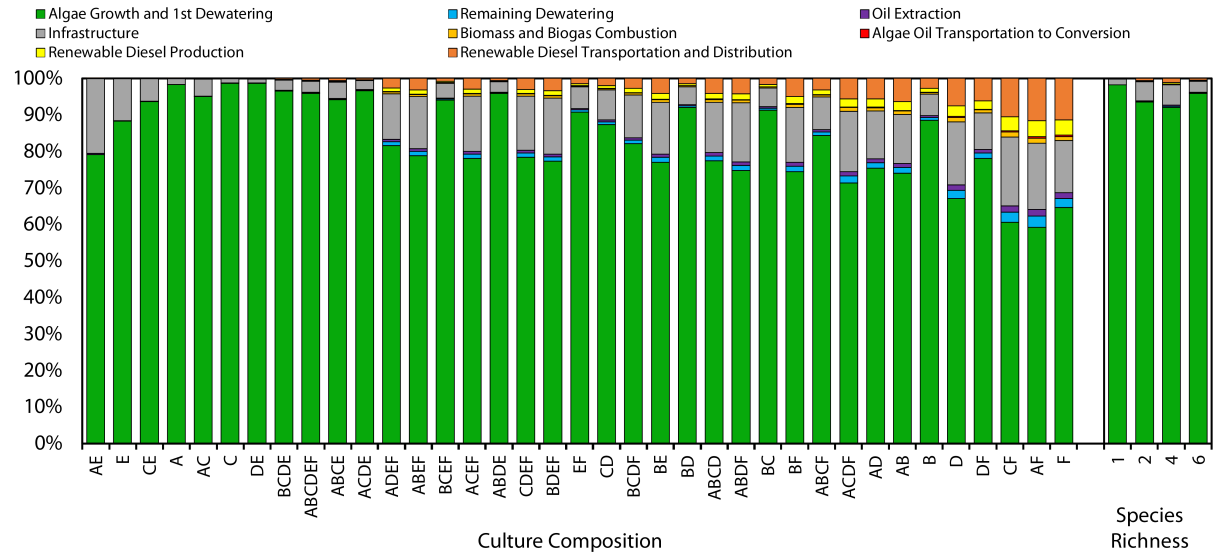


Figure 4.8: Breakdown of nonrenewable energy demand from GREET, normalized and plotted by culture composition and by species richness under the constant temperature mesocosm condition.

environmental realities including fluctuations of photosynthetically active radiation (PAR), invaders, temperature change, evaporation, etc. These data on biomass cultivation were supplemented with elemental data and biocrude yields from variable temperature mesocosm experiments to capture characteristics of outdoor pond cultivation, encompassing physiological changes in elemental composition due to varying temperature as well as mitigated growth due to exposure to naturally occurring invaders, variable sunlight, and other environmental factors.

Fig. 4.9 shows LCA results on two metrics based on outdoor condition. Intriguingly, in this analysis, bicultures BF and AF outperformed *S. capricornutum* (F) in terms of EROI, GHGs, and WI ($p < 0.05$). EROI for cultures BF, AF, and F were 0.60, 0.54, and 0.51, respectively, representing a significant 20% improvement by the best biculture BF over the best monoculture F. For GHGs, cultures BF, AF, and F led to values of 137, 154, and 163 $kgCO_2eq/MBTU$, respectively, marking a 16% improvement by biculture BF in comparison to monoculture F. Water intensity (WI) results followed those of GHGs with values of 268, 364, and 444 L/MBTU (Fig. 4.10).

The apparent success of bicultures BF and AF is contrary to what was observed in the

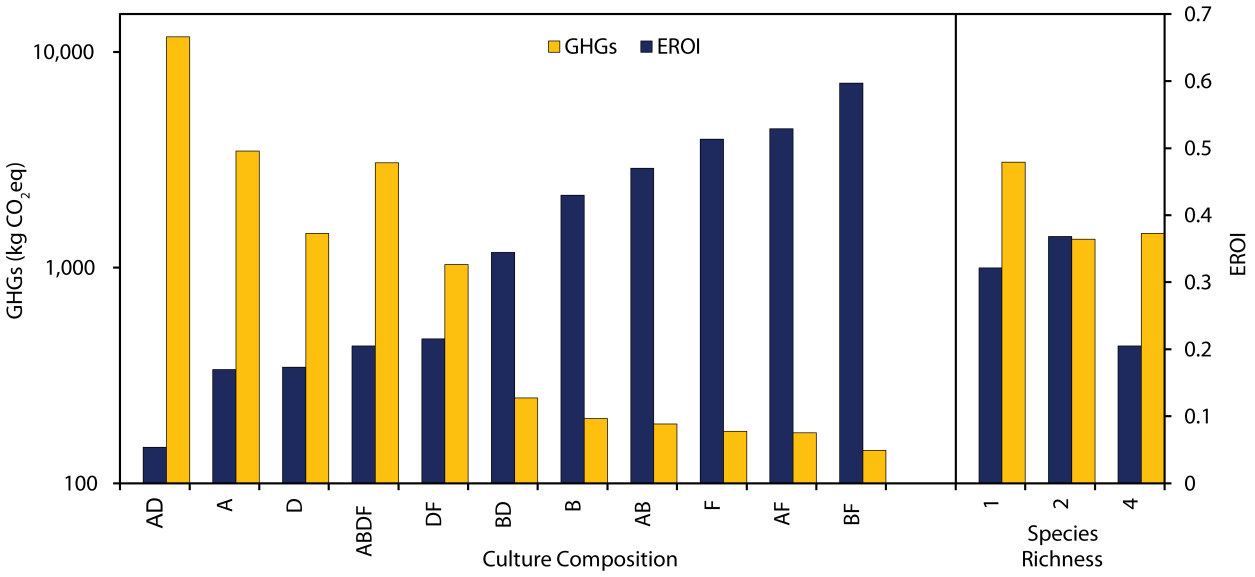


Figure 4.9: Life cycle metrics of EROI and GHGs from outdoor pond cultivation experiments. GREET model return using empirical data primarily from outdoor pond cultivation experiments for varied cultures (sorted by increasing EROI) and by mean species richness..

laboratory experiments in which *S. capricornutum* outperformed the next best culture in EROI by a 27% margin. The average biculture also outperformed the average monoculture and the four-species culture in each life cycle metric ($p < 0.05$). We noted that despite lower performance in EROI, GHGs, and WI compared to two bicultures BF and AF, *S. capricornutum* maintained the highest biomass concentration and productivity of all cultures with 0.200 g/L and 9.24 g/m^2/day, far greater than BF and AF at 0.143 g/L and 6.45 g/m^2/day and 0.16 g/L and 7.54 g/m^2/day, respectively. However, the extent of these advantages of *S. capricornutum* under outdoor conditions were considerably smaller than those under laboratory conditions. In terms of biomass productivity, *S. capricornutum*'s advantage over the bicultures AF and BF decreased from 48% (laboratory) to 19% (outdoor) and from 137% to 59%, respectively. With regard to biomass concentration at harvest, *S. capricornutum*'s advantage over the bicultures AF and BF decreased from 45% (laboratory) to 21% (outdoor) and from 126% to 48%, respectively.

The disparity between *S. capricornutum*'s superiority in biomass concentration/productivity and its inferiority in LCA metrics, compared to bicultures BF and AF, indicated that the

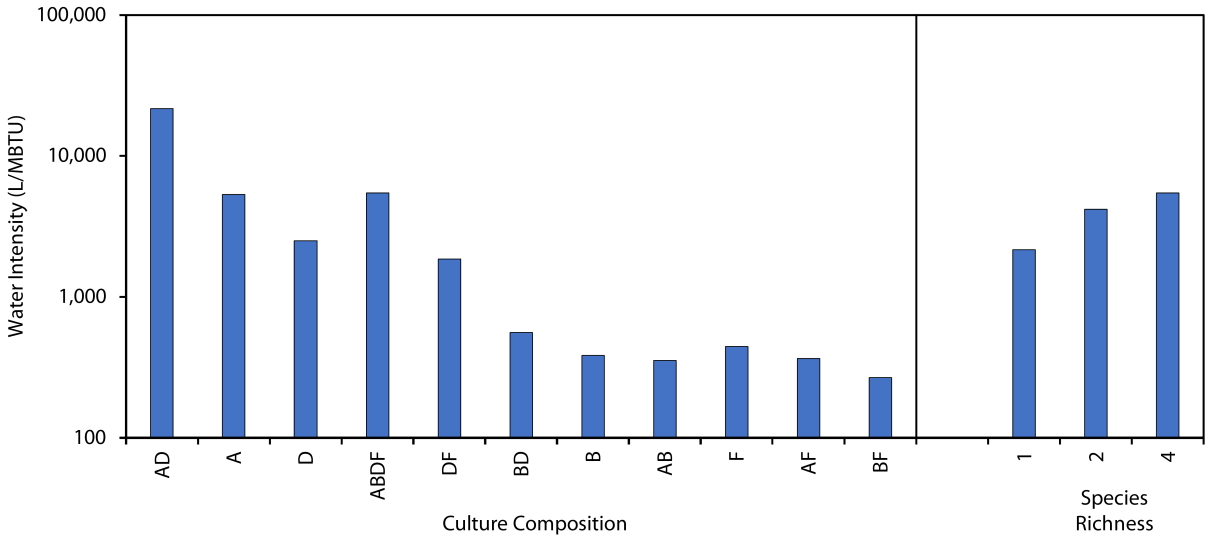


Figure 4.10: WI of algal cultures under outdoor pond conditions from the GREET model by culture composition and species richness.

positive impact of biomass production was outweighed by other aspects of *S. capricornutum* that negatively impacted LCA metrics. For instance, we noted that *S. capricornutum* demanded significantly higher phosphorus inputs at 2.93 mM P/g_{afdw} algae, compared to 1.7 mM P/g_{afdw} for BF and 2.1 mM P/g_{afdw} for AF. Higher P requirements adversely affected overall life cycle metrics. Intriguingly, this high demand for phosphorus was not apparent under the constant temperature mesocosm condition in the laboratory, suggesting physiological differences between algal growth under the constant temperature condition and the outdoor pond condition. We further hypothesized that select polycultures outperformed the best monoculture by maintaining multiple functions in a more balanced manner, an effect of biodiversity referred to as multi-functionality [170, 171]. We previously found that several of our polycultures have high potential for multi-functionality [49]. In particular, the two most productive species *C. sorokiniana* and *S. capricornutum*, have disparate N and P nutrient use efficiencies [150]. When grown together, these species simultaneously optimize N-efficiency, P-efficiency, and biomass production better than either species can as a monoculture. Thus, while *S. capricornutum* dominated in terms of stability, productivity, and biomass concentration across all experimental conditions, these factors were outweighed by

nutrient use efficiency under outdoor pond conditions, compared to bicultures AF and BF.

4.3.3 Quantification of multi-functionality in outdoor pond growth

Superior multi-functionality of polycultures leads to more favorable life cycle metrics

To systematically dissect how multi-functionality benefited bicultures BF and AF, compared to the best monoculture F, we conducted a multilinear regression analysis to quantify how different aspects of performance impact LCA metrics. Independent input variables included nutrient recycling, elemental composition, biocrude yields, biomass concentration, and frequency of crashes from outdoor pond experiments. Dependent inputs such as productivity and H:C ratios were excluded as, for example, biomass concentration is strongly correlated with calculated productivity (Equation 43) ($R^2 = 0.99$). Each input was normalized using Z-scores such that regression coefficients provide relative weights for individual inputs with respect to EROI ($\sigma_{\text{EROI}}/\sigma_{\text{Input}}$):

$$\left(\frac{Y_i - \overline{Y}_i}{\sigma_i}\right) = \beta_0 + \sum_{j \in i} \left(\beta_j * \left(\frac{X_j - \overline{X}_j}{\sigma_j}\right)\right) + \epsilon \tag{45}$$

The multilinear regression yielded a good fit to the data ($R^2_{\text{EROI}} = 0.78$) using the workflow depicted in Fig. 4.11. The product of an input regression coefficient and its Z-score yielded its contribution to the EROI. The aggregation of these contributions then yielded the overall sustainability impact. Predicted EROI values corroborated actual EROI from GREET within reasonable deviations. GHGs ($R^2_{\text{GHGs}} = 0.75$) and WI ($R^2_{\text{WI}} = 0.92$) regressions were also conducted and were inversely correlated to EROI (Table G.2, G.3).

Results from the regression analysis showed that biomass concentration, "Biomass Density" in Fig. 4.11, is the single most important aspect of performance; however, biocrude yield, biocrude and algal elemental compositions, and nutrient use efficiencies also have sig-

AHM Inputs	Z-Scores $(X_{Input} - \mu_{Input})/\sigma_{Input}$			β_{Input}	Component Contribution (σ_{EROI})		
Culture	**BF**	**AF**	**F**	**--**	**BF**	**AF**	**F**
Biomass Density	0.58	0.91	1.63	1.15	0.66	1.04	1.88
Algal P	-0.17	1.38	1.76	0.05	-0.01	0.07	0.09
Algal N	0.52	-0.97	-1.22	0.01	0.01	-0.01	-0.01
Biocrude Yield	0.10	-0.67	-0.71	0.20	0.02	-0.13	-0.14
Biocrude C	0.14	-1.07	-1.22	0.31	0.04	-0.33	-0.38
Biocrude N	0.03	-0.86	-0.75	-0.04	0.00	0.03	0.03
Exhausted media N (as NO_3^-)	0.20	0.39	0.17	0.37	0.07	0.14	0.06
Exhausted media N (as NH_4^+)	-0.37	-0.28	-0.27	0.00	0.00	0.00	0.00
Exhausted media P (as PO_4^+)	0.80	0.16	-1.93	0.65	0.52	0.10	-1.25
ACP (as PO_4^+)	-0.13	0.29	1.32	0.14	-0.02	0.04	0.18
ACP (as NH_4^+)	0.54	-0.59	-0.92	0.11	0.06	-0.07	-0.10
Average Crashes (per yr)	-0.69	-0.81	-0.75	-0.14	0.09	0.11	0.10
$\sum\sigma_{EROI}$ **Components**					**1.45**	**1.00**	**0.46**
					EROI Results		
Predicted EROI					**0.65**	**0.56**	**0.46**
Actual EROI					**0.60**	**0.53**	**0.51**

Figure 4.11: Quantification of input impacts on EROI in LCA. Using Monte Carlo simulations based on parameter distributions, AHM inputs for each species combination were normalized using Z-scores of input parameters to analyze input effect on EROI through multilinear regression. Red table cells correspond to inputs and components with a negative impact on EROI. The component contribution indicates how much an input contributes to variation in EROI (σ_{EROI}) and, beneath it, a comparison of the predicted EROI to actual EROI from AHM-GREET calculation.

nificant effects on sustainability metrics. Changes in dry algal biomass P ($\beta = 0.05$) and N ($\beta = 0.01$) have minimal effects on EROI whereas changes in biomass concentration ($\beta = 1.15$) and recovered phosphorus ($\beta = 0.65$) have large effects (Table G.1). Despite the importance of individual aspects of performance such as biomass concentration, the net impact of different algal culture compositions depends upon the balance of positive impacts from above-average functions/traits and negative impacts from below-average functions/traits. As a result, culture compositions that have high Z-scores for multiple functions can outperform compositions that perform only one or two functions well but perform other functions poorly. In our system, the best single species (*S. capricornutum*) had a strong positive impact on LCA metrics based on its ability to achieve high biomass concentration (+1.88 σ_{EROI}), but this advantage was largely negated by poor phosphorus nutrient use efficiency, which decreased EROI by -1.25 σ_{EROI}. Conversely, two bicultures that had lower biomass production (BF and AF) exhibited high efficiency of nutrient recycling from the exhausted medium and aqueous co-product (ACP) from hydrothermal liquefaction, accounting for 0.63

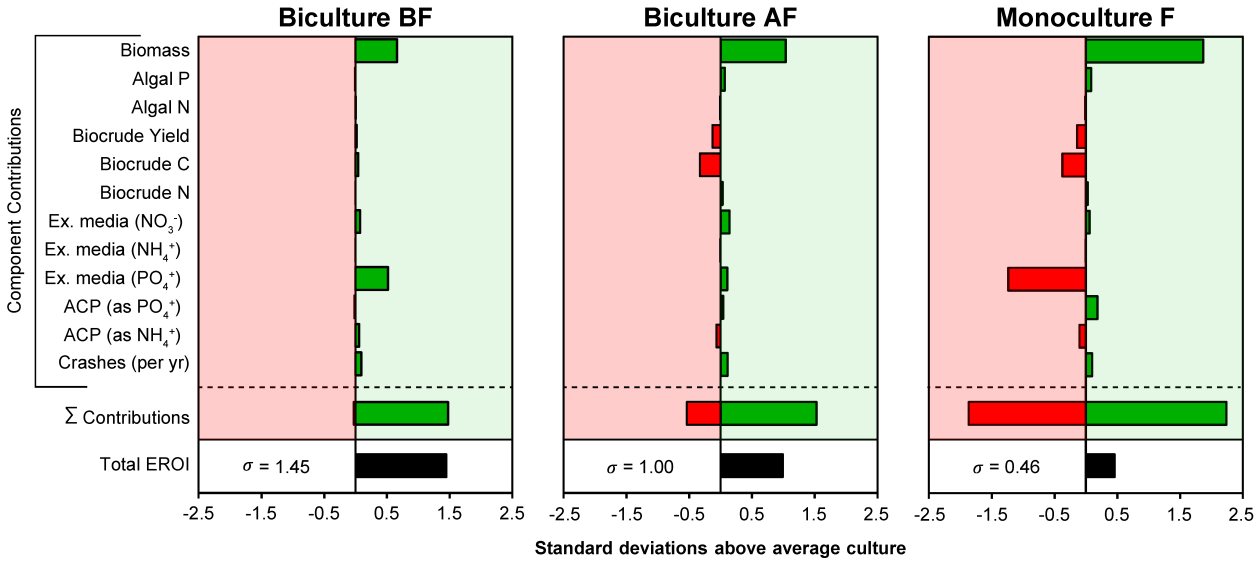


Figure 4.12: Schematic of the EROIs of monoculture F, biculture AF, and biculture BF from accumulated positive and negative contributions from all inputs of the LCA calculation (exact values in Fig. 4.11, under column header "Component Contribution").

and 0.21 σ_{EROI}, respectively, (compared to -1.11 σ_{EROI} for species F). The result of these impacts is that the best single species, F, has an EROI of 0.51 whereas the best polyculture, BF, has an EROI of 0.60 (an increase of 18%). Amongst these inputs, *S. capricornutum* and many monocultures alike tend to have imbalanced impact contributions such that, as in this case, adept biomass accumulation but poor phosphorus nutrient use efficiency ultimately yield worse sustainability metrics.

Fig. 4.12 illustrates how this effect of multi-functionality led to the superior performance of bicultures BF and AF, over the best monoculture F. Thus, using data collected under realistic outdoor open pond conditions, our results highlight how biodiversity improved life cycle sustainability metrics in algal biofuel production. Our study shows that biodiversity can improve sustainability of algal biofuel due to the ability of polycultures to optimize multiple functions simultaneously that influence LCA metrics. This is a key finding for the development of sustainable biofuel systems as it shows that polycultures can help overcome the intense tradeoffs that are characteristic of single-species cultivation.

Contrary to some previous LCAs that suggest algal biofuels could attain EROI > 3.0,

our assessment did not indicate that algal biocrude production would be energetically favorable. We believe that this discrepancy is due to the fact that we used real-world cultivation data from our experiments, whereas many other assessments have been based on hypothetical cultivation. For example, a recent review of well-to-pump LCAs showed that assumptions about lipid yields (m^3/ha/day) varied by more than 2000%, resulting in predicted greenhouse gas emissions ranging from a net negative value of -90.7 gCO_2eq/MBTU to over five times the emissions of diesel at 506 $kgCO_2eq$/MBTU [172]. Recent efforts to addresses these issues have culminated in meta-analyses and harmonization reports that attempt to standardize assumptions. [65, 173–175] However, one such meta-analysis noted that of 54 assessments reviewed, nearly half did not explicitly name a target algal species despite inherent variation in species specific properties presented here and throughout the literature [18, 61, 150]. Discrepancies may be more appropriately addressed through multidisciplinary collaboration to close the knowledge or expertise gaps between algal cultivation, biomass conversion, and life cycle assessment by reducing the number of unrealistic assumptions and by increasing continuity across algal processing stages.

In this work in particular, biomass concentration and areal productivity were not both considered independent variables as indicated in Equation 42. To demonstrate the importance of biomass and productivity modeling, we conducted a simple sensitivity analysis for varied productivities and biomass concentrations, using *S. capricornutum* data from the constant temperature mesocosm experiment, which had empirical values of 6.25 g/m^2/day and 0.37 g/L, respectively (Fig. 4.13).

If biomass concentration and productivity are arbitrarily specified to be 0.5 g/L and 13.2 g/m^2/day, values commonly assumed in previous LCA studies, *S. capricornutum* would have an EROI of 2.41 and GHGs of 33 $kgCO_2eq$/MBTU, which are more favorable than the values presented above in Results. This dramatic contrast raises a valid and long-overdue concern over the readiness of algal biofuel technologies. It also highlights the need for more aggressive interdisciplinary research and development efforts to ultimately realize the full

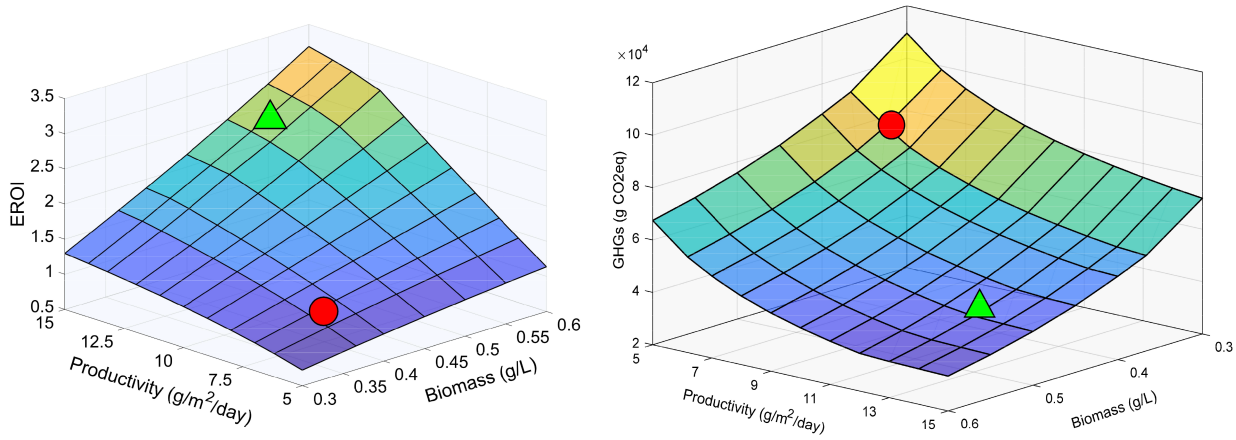


Figure 4.13: Effects of biomass density and areal productivity on EROI and GHGs. Red circles denote values employed in this study for the best performing condition and culture, *S. capricornutum* at constant temperature mesocosm with 0.37 g/L biomass accumulation and 6.2 g/m^2/day areal productivity. Green triangles represent results under the common assumptions of 0.5 g/L biomass accumulation and 13.5 g/m^2/day areal productivity.

environmental and economic benefits of algal biofuels.

4.3.4 Tuning of harvest scheduling

Algal biofuel life cycle assessments have explored sensitivity to major bottlenecks like biocrude yield, composition, and conversion efficiencies. However, few LCAs have explored pond crashes and fluctuations due to upstream cultivation despite the labeling of pond crashes as a major barrier to industrial algal biofuel production [169]. Crashes are most widely attributed to invasion by biological contaminants. Invaders include rotifers, other species of algae, pests, and diseases like chytrid infection, which are complicated further by seasonality. Earthrise Farms, a company that pioneered algal cultivation in the 1980s reported that contaminant algal strains like *Chlorella* and some species of *Spirulina* resulted in an estimated annual productivity loss of 15 to 20% [176]. Likewise, half of their 15 ponds of 5,000 m^2 surface area were declared crashed after one month of cultivation due to excessive contamination by green algae [176]. More recently, Sapphire Energy's Las Cruses Test Site, which harbored 48 ponds of 1.1 acre surface area and 20 ponds of 2.2 acre surface area under continuous operation for over two years reported that losses to grazers like rotifers, cilates,

and amoeba resulted in approximately 20 ± 10% of total pond productivity [169].

While crashes have been marked as a leading barrier to development of algal production, few attempts to model stability with respect to ecological physiology have been made. A more common strategy is the use of industrial agricultural management strategies, which have proven somewhat effective. Second generation sequencing of the V6 and V6 rRNA domains for bacterial and eukaryotic species, respectively, can readily characterize biological contaminants for tailored management treatments, though these methods are significantly more expensive. Furthermore, broad elimination of grazer populations may contribute to decreased stability of the algal cultivar. Ecological engineering has the potential to reduce deleterious effects of biological contamination through multifunctionality in which several parameters akin to efficient biofuel production and increased pond stability are optimized simultaneously [62, 177]. To date, only a few LCAs have explored the effects of algal pond crashes on overall life cycle metrics. Pond crashes remain difficult to characterize not only due to their stochastic nature, but also the manner with which pond composition is reestablished.

Here, we employ the AHM as well as the autoregressive algal growth model to explore the impact of harvest scheduling of two polycultures, AF and BF, and the best performing monoculture, F. As described previously, the autoregressive algal growth model was developed using Wolfram Mathematica for iterative calculation of parameter fits from empirical pond data [155]. The model was verified for validity and homoscedasticity with a comparative weekly biomass output over two years shown in Fig. 4.14. Results were then bootstrapped to a logistic equation.

The model employed here returned values of average biomass at harvest, areal productivity, and mean crashes per year depending on harvest fraction, harvest interval, and crash thresholds. The nature of the model means that biomass and productivity are inherently coupled. More complex models for algal growth that incorporate nutrient availability, light penetration, and heat transfer have been explored in more detail elsewhere. [37,154,178–180] While such models are more rigorous, these parameters were not considered due to limita-

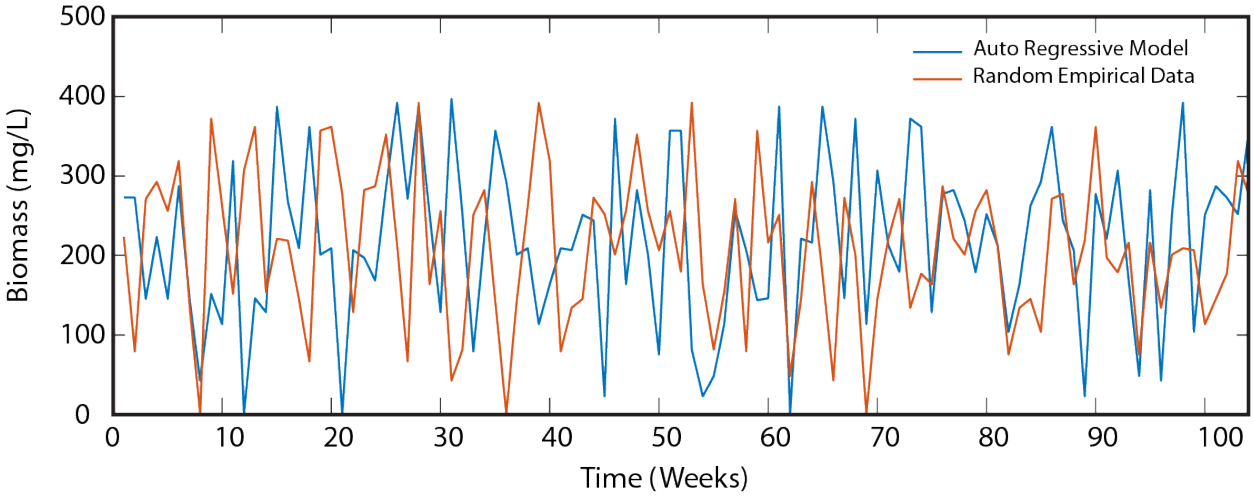


Figure 4.14: A 104 week plot comparing variation in the autoregressive model data to random empirical data for *S. capricornutum.*

tions in empirical data collection. These models also tend to overlook stochastic processes like pond contamination, accumulation of autoinhibitive dissolved organic matter, and events that could ultimately lead to pond crashes. Instead, we used the Equation 42 to approximate the many factors that directly impact algal growth and resolved a piecewise function for the determination of algal pond collapse. While the model does not explain the relative impacts of different environmental and nutrient factors on the algal cultures, it effectively replicates the variation and overall mean observed in the field.

The LCA detailed in the previous sections used a crash threshold of 50 mg/L as a mechanism accounting for these crashes based on observational evidence from open pond experiments. If a culture is below the threshold after a given harvest, the biomass is calculated as zero for the week in which the crash occurred as well as the following two weeks for draining and re-inoculation. After three weeks of zero biomass the culture is re-started with an initial biomass density of 10 mg/L and the simulation continues. However, the crash threshold is an arbitrary biomass concentration and, again, may be subject to a myriad of factors. By using normal distributions for carrying capacity and growth rates, this threshold is probabilistic in nature.

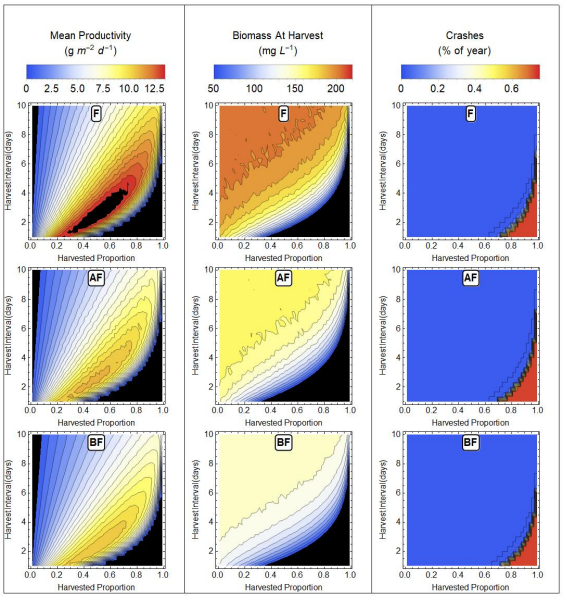
In terms of harvesting, the LCA also used a set harvesting schedule of 0.70 fraction every

7 days. This, again, was based on field experiments rather than computational optimization. To better understand how harvest fraction and interval affect algal growth (biomass, productivity, and annual pond crashes), harvest fractions were varied between 0.02 and 0.99 and harvest interval between 1 and 20 days at intervals of 0.02 and 0.33 days, respectively. We also wanted to consider how harvest schedule for optimal algal growth might be affected by varying the crash threshold. Fig. 4.15 illustrates effects of different crash thresholds of 5 mg/L (Fig. 4.15a), 45 mg/L (Fig. 4.15b), 85 mg/L (Fig. 4.15c), and 125 mg/L (Fig. 4.15d) affected (left to right) mean productivity, biomass at harvest, and % time crashed per year for cultures (top down) F, AF, and BF.

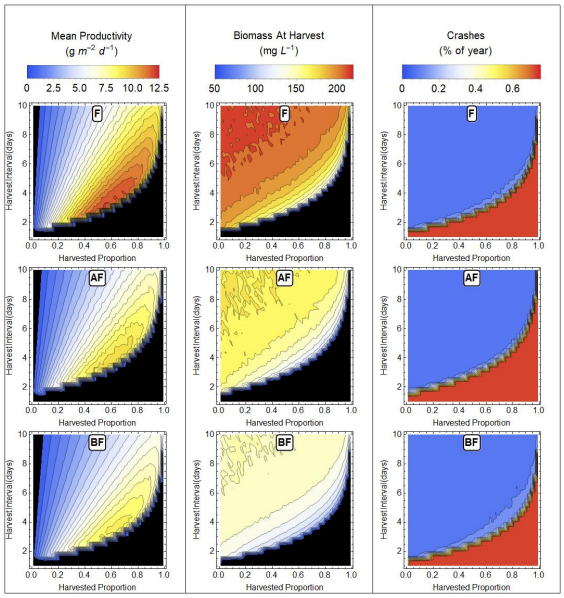
Increasing the crash threshold resulted in higher crashes per year with lower biomass at harvest and mean productivity. However, all three cultures are remarkably stable over the crash threshold variation. Differences in the frontier (black area) of crashed vs. non-crashed culture indicates that certain harvest schemes are fundamentally inviable. High fraction harvests that occur frequently, for example, lead to high pond crash rates. Infrequent, low fraction harvests are likewise not viable. Comparing the crash frontier between panels indicates that higher crash threshold shifts scheduling towards longer harvest intervals with larger harvest fractions (semi-batch). At lower crash thresholds, these cultures are more optimized towards semi-continuous harvesting. In the extreme case of 5 mg/L crash threshold *S. capricornutum* (F), this means harvesting 0.5 pond fraction every 2 – 3 days.

Lastly, we analyzed the effects of pond crash threshold on LCA metric EROI of BF, as illustrated in Fig. 4.16 with different crash thresholds of 5 mg/L (Fig. 4.16a), 45 mg/L (Fig. 4.16b), 85 mg/L (Fig. 4.16c), and 125 mg/L (Fig. 4.16d).

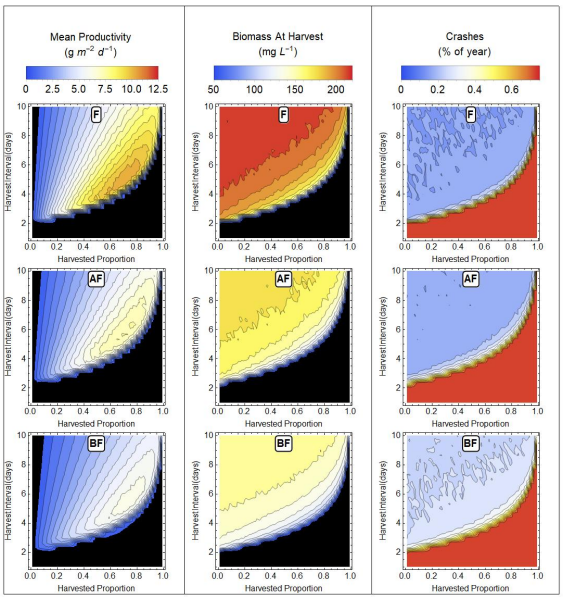
These findings are perhaps unsurprising by nature of the logistic growth model. Lower crash threshold makes fluctuations on growth rate and carrying capacity less detrimental to harvest schemes. Similarly, cultures with the highest mean biomass accumulation at harvest will have higher resilience to pond crashes as defined in this model. Nonetheless differences in harvesting regimes suit different crash threshold conditions. If crash threshold



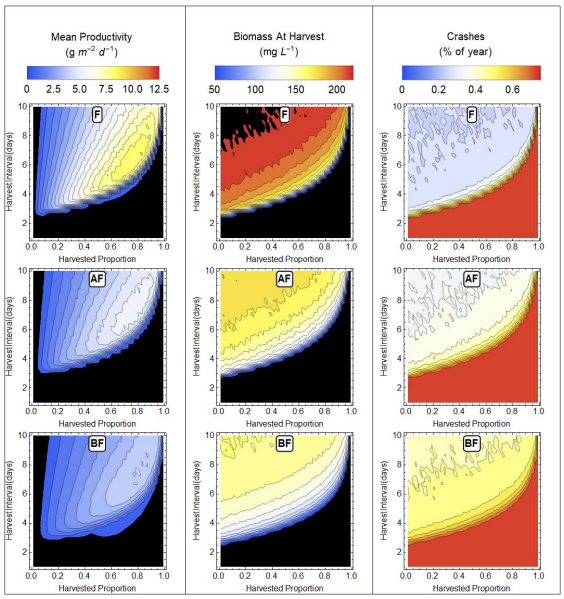
(a) Crash threshold of 5 mg



(b) Crash threshold of 45 mg

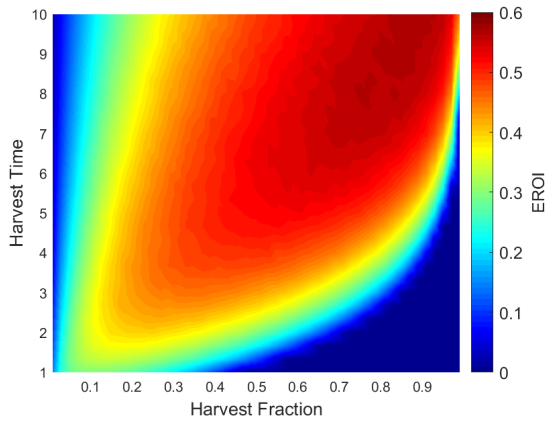


(c) Crash threshold of 85 mg

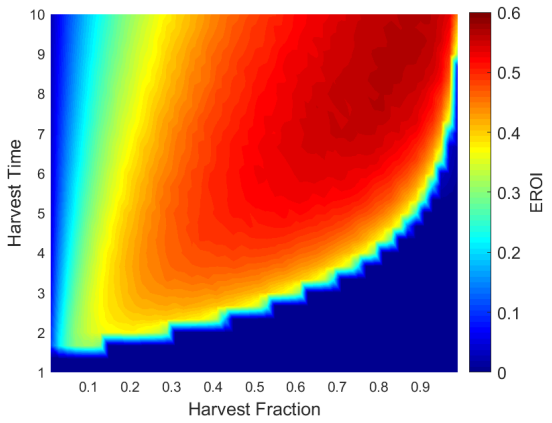


(d) Crash threshold of 125 mg

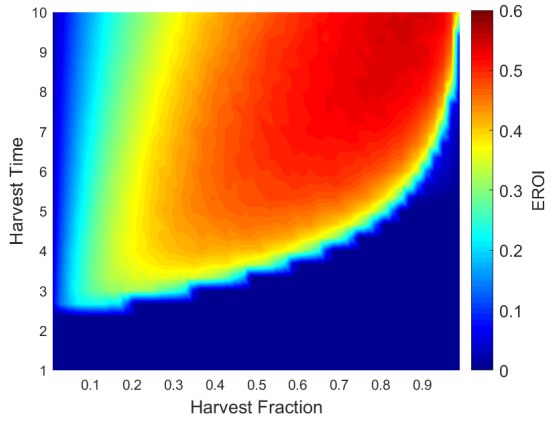
Figure 4.15: Analysis of shifting crash threshold on mean aerial productivity, mean biomass at harvest, and average crashes per year.(a) Pond crash threshold set to 5 mg, (b) Pond crash threshold set to 45 mg, (c) Pond crash threshold set to 85 mg, (d) Pond crash threshold set to 125 mg



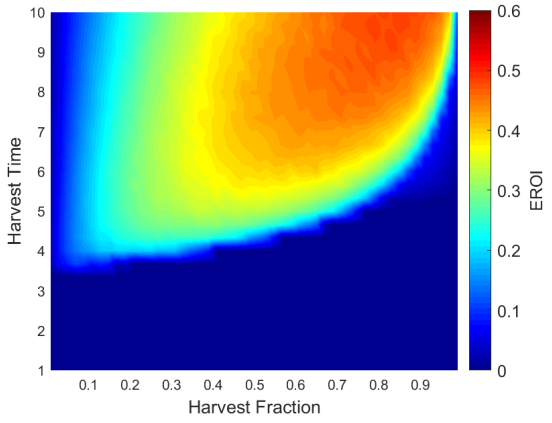
(a) BF EROI with crash threshold of 5 mg



(b) BF EROI with crash threshold of 45 mg



(c) BF EROI with crash threshold of 85 mg



(d) BF EROI with crash threshold of 125 mg

Figure 4.16: Analysis of shifting crash threshold on EROI. (a) Pond crash threshold set to 5 mg, (b) Pond crash threshold set to 45 mg, (c) Pond crash threshold set to 85 mg, (d) Pond crash threshold set to 125 mg

conditions may be identified (*e.g.*,. severe weather event, low cell/CFU count invader, etc.), pond harvesting can be changed to avoid pond crashes to increase biomass and productivity, which in turn benefit life cycle metrics. Redefining the pond crash threshold could also enable growers to better test pond resilience. Inclusion of parameters beyond those defined in the simplistic logistic growth model could enable a better optimization of harvest schemes.

4.4 Conclusion

The superior potential for multi-functionality of certain polycultures compared to the single best species suggests that the widely employed strategy of selecting strains based on potential productivity or lipid content is inadequate for determining the sustainability or suitability of an algal feedstock. Specifically, choosing a species based on its productivity may maximize the rate of biofuel production, but that single measure of performance often comes at the expense of other important aspects of performance (*e.g.*,. nutrient use efficiency, biomass composition). Notwithstanding improvements to downstream technologies, our assessment shows exploiting biodiversity as a powerful approach for improving energetic efficiencies. However, the development of renewable fuels from algae faces the challenge of scaling from small experiments to large commercial cultivation systems. Our assessment suggests that biodiversity can benefit sustainability metrics based on data from 1,100 L scale outdoor cultivation experiments, which were subject to PAR fluctuations, invaders, temperature, evaporation, and other environmental factors. However, it remains to be answered whether those benefits would prevail at larger scales of cultivation (*e.g.*,. million liter raceway ponds). In particular, it is uncertain whether the increase of productivity potentially achievable in shallow raceway ponds would outweigh the increased risk of contamination due to greater surface area [163]. On the other hand, our ponds were considerably deeper than commercial cultivation systems and were not supplemented with carbon dioxide, both of which reduce overall productivity. Future studies in large raceway ponds for long periods that explore multi-species polyculture cultivation as well as downstream processing will be

required. Design of cultivation systems that optimize depth, carbon dioxide supplementation, and growth medium could serve to enhance productivity beyond what we have achieved so far in open ponds, which would increase the EROI of all species combinations. Ecological engineering of algal communities presents enormous potential in optimizing multiple functions simultaneously under realistic cultivation conditions, thereby addressing a major barrier for the production of algal biofuel as a sustainable alternative to petroleum-based fuels.

Chapter 5: Concluding Remarks and Future Directions

The underlying theme of this dissertation has been harnessing microbial consortia for bioproduction of sustainable chemicals. Microbial consortia may be genetically engineered as showcased through strain engineering in the *csc* random integration work (Chapter 3) and the tripartite platform (Chapter 2). Consortia may also be developed through community engineering as showcased through the combinatorial analysis of algal polyculture characteristics (Chapter 4). Collectively, these strategies have shown how synthetic communities can impart beneficial characteristics to bioproduction platforms and, ultimately, increase the viability of certain platforms in terms of their relative sustainability, energetics, and production to current practices.

5.1 Optimization of the tripartite platform

While an interesting strategy for bioproduction with minimal nutrient inputs, the tripartite platform requires further optimization before becoming comparable to current practices.

5.1.1 Genetic modification of culture constituents

There are a number of opportunities for optimizing the tripartite platform. For example, the system employs two notoriously energy intensive processes, nitrogen fixation and photosynthesis, to produce nutrients that are readily usable by a vast library of microbial strains. Genetic modification of these processes is remarkably complex. Optimization of

photosynthesis, namely the Calvin cycle or the rubisco enzyme, has long been a holy grail of bioproduction platforms. Many strategies instead focus on efficient carbon storage mechanisms or sinks that can increase photosynthetic efficiency. Faster growing strains of *S. elongatus* have been generated that could have immediate utility in this system. A recent work describes an engineered strain of *Synechococcus elongatus* 2973 with a peak sucrose productivity of 1.9 g/L/day and maximum titers as high as 8.1 g/L approaching 90% photosynthetic efficiency [181]. Assuming these values are translatable to the tripartite system and that the relative loss of productivity/titer between BG-11 and the Union Medium is comparable to that of *S. elongatus* (*cscB* + SPS), this would enable approximately 2-fold higher sucrose production in our system. A 2-fold increase in sucrose could relieve the sucrose production bottleneck. On the other hand, optimization of alginate or PHB production in *A. vinelandii* has produced strains with significant growth defects [182–184]. Tuning nitrogenase production is complicated by the fact that there is no scientific consensus on the entire pathway of nitrogen fixation as well as the myriad of associative genes. Most studies have focused purely on the manipulation of the the *nifL*/*nifA* regulatory elements. Gains analogous to *S. elongatus* made in *A. vinelandii* could also spur sustained, high dilution production in this framework.

5.1.2 System stability

A consequence of the dual photosynthesis-BNF coculture that is alluded to in Chapter 2 is the triculture's susceptibility to invasion. The system is always threatened by potential contaminants as well as "cheaters". Indeed, contamination led to the spontaneous generation of the stable photosynthetic consortium. Cheaters are characterized as cells or strains that mutate to benefit at the cost of another strain. In monoculture, for example, there is selective pressure for *A. vinelandii* or *S. elongatus* to withhold beneficial and energetically intensive nutrients (sucrose, NH_4^+) and also to use those excreted/secreted nutrients in the surrounding medium. Although the synthetic symbiosis described here and elsewhere

necessitates nutrient exchange and thereby does not exhibit this monocultural pressure, cheaters can arise in any community that operates at evolutionary timescales. Alternatively, it is also possible that continual passage of *A. vinelandii* and *S. elongatus* cocultures could yield beneficial mutations that increase their combined propensity for supporting production strains.

5.1.3 Physical platform changes

Improved physical tuning of compartmentalized conditions by, for example, differentiating gaseous delivery of $CO_2/N_2/O_2$ could enable enhanced constituent growth and production. Membrane separated systems are frequently used in co-cultures, but rarely in industrial biofermentation due to membrane fouling and maintenance logistics. In order to be effective in industrial biofermentation, most robust growth or compartmentalization systems must be engineered.

Other conditions like tuning *S. elongatus* culture irradiance with increases in cell density for optimal PAR could yield enhanced growth by the photosynthetic organism. A benefit of using the LED system described in Chapter 2 is that the LEDs are connected to an Arduino, which means light intensity can be programmed to change over time. A specific opportunity would be tuning light intensity to the Beer-Lambert law such that the intensity increases in parallel to the optical density of the culture. This may require a different method for measuring real time optical density, but could provide significant benefits to the *S. elongatus* strain by increasing light availability. Another strategy would be the use of LEDs that are more tuned to PAR spectrum to increase the amount of useful light.

5.1.4 Modular addition of production strains

The tripartite platform simply requires production strain candidates to utilize sucrose and ammonia. Transduced strains can be readily cultured inthe tripartite system due to this unique modularity. Specifically, the development of a fast-growth, transducible *cscBKA* locus

essentially enables any *E. coli* K-12 strain to be a potential bioproduction candidate assuming no modification in or around the *nlpA* gene. The model used in this system, however, assumes minimal interactions between strains. In the case of the isobutanol production strain, there will definitely be inhibition to the constituent strains of *S. elongatus* and *A. vinelandii* due to a lack of isobutanol tolerance. It is realistic to expect that other products would have considerable inhibition in the system and would require modification to increase tolerance or avoidance of such products altogether. It is similarly realistic to expect that strains producing beneficial chemicals like amino acids may not be viable due to uptake by the other strains.

5.1.5 Corroboration of the kinetic mode with empirical data

The kinetic model demonstrated good corroboration with the growth and production of tripartite constituents using monoculture data. Certain observations were poorly explained. This is particularly evident for the *A. vinelandii* overshoot during the first several days of culture as well as the accumulation of ammonia for *E. coli* K-12 *cscBKA*:: *nlpA* and *B. subtilis* 168. Generation of high confidence polyculture parameters could improve the overall fit of the model. Additional maintenance terms that better explain, for example, the initial accumulation of PHB in *A. vinelandii* would also be beneficial and improve model corroboration with empirical data.

5.1.6 Conclusion

The platform must address many challenges before efficient bioproduction is possible. While the subset of experiments discussed in Chapter 2 are important first steps in platform development, the opportunities outlined here are essential for platform viability. With further research and development, this framework could prove competitive with fossil fuel analogues in terms of greenhouse gas emissions, energetics, and overall cost.

5.2 Random chromosomal expression of heterologous genes in *E. coli* K-12

The results of this study are unique both to sustainable biofermentation in K-12 derivatives and for molecular biology. The integration of the *cscBKA* operon randomly into the K-12 chromosome led to a uniquely nonintuitive conclusion that fastest growing strains were expressed in low or non-expressing genes. Originally, the study aimed to integrate genes exclusively into high and low expression loci on the hypothesis that high expressing loci would yield high growth strains while low expressing loci would yield low growth strains. However, our conclusion, which is mildly limited by the few colonies screened, suggests that the opposite is true: the fast-growth phenotype is characterized by low expression relative to plasmid. Expression also appears lower than *E. coli* W, though is complicated by the fact that comparative reference genes between *E. coli* K-12 and *E. coli* W are not well-studied. The conclusion of this work poses more questions about interplay between native/heterologous gene expression as well as the interplay between chromosomal expression and metabolic flux.

There are a number of opportunities for better characterization of the transposon library, namely methodologies that do not screen for faster growing integrants. More robust screening strategies include cultivation on a lower concentration of antibiotic or longer recovery time under ideal conditions. Longer recovery is complicated slightly by the tendency of a polyclonal library, though can still be screened with sufficient sequencing throughput. Although it has never been explored, it is possible that piecewise integration of *cscK*, *cscB*, and *cscK* into *E. coli* K-12 could yield positive benefits. This strategy would remove the bidirectional promoter and potentially allow for better overall tuning. It is unclear whether that would actually improve growth performance.

The transduction of the K-12 B2 integrant (*cscBKA*::*nlpA*) into a chromosomally encoded isobutanol production strain with minimal reduction in overall efficiency on sucrose compared to native production on glucose demonstrates its versatility in carbon substrate

switching. A full comparative LCA of isobutanol production on sucrose vs. glucose production would better showcase the benefits of this strain. Growth of the B2 integrant in triculture with *A. vinelandii* and *S. elongatus* also demonstrates how transduced strains are viable candidates for growth in the tripartite system. Transduction could also be extrapolated biologically to other production strains (recombinant protein, biofuels, biopolymers, etc.) to yield potential reductions in cost and environmental or energetic burden. Furthermore, a more thorough, standalone LCI for bioproduction platforms based on the work published by Argonne National Laboratory could provide an excellent tool for top-down strain development.

5.3 Optimization of algal biofuel production through the use of polycultures

The life cycle assessment of the algal polyculture biorefinery presented the unique finding that certain polycultures could outperform monocultures. This gain, however, arose within the context of outdoor pond cultivation under realistic conditions. An underlying finding is that highly productive laboratory strains do not necessarily translate to outdoor cultivation, which has important implications to the development of biofuel platforms. It is also important to note that some of the susceptibility to invasion in the field could stem from coevolution of invaders with specific strains of microalgae used for production. This could have positive or negative effects on resistance to invasion. It remains unclear whether any laboratory-scale experiments can be effectively extrapolated to performance under real cultivation conditions in outdoor ponds. Our results quite clearly show that this is often not the case.

Appendix A: Strains explored in the tripartite study

Table A.1: List of strains explored in the tripartite study including growth conditions, characteristics, and designations.

Strain Name	Strain Designation	ATCC	Modifications	Genome Size	Accession	Gram stain	Relevant Phenotype	Culture Temp.	pH	Minimal Media
Synechoccocus elongatus	PCC 7942	33912	*cscB* + *lacI*p, SPS + *lacI*p	2,695,903	CP000100	-	Produces and excretes sucrose	30 - 35	8	BG-11
Azotobacter vinelandii	DJ	1303	*nifL*::*ptII*(KanR)	5,365,318	CP001157	-	Inactive nitrogenase repression, kanamycin resistance (3 mg/L), overproduces and secretes ammonia	30	7.0 - 8.0	Burks N-free
Escherichia coli	W	9637	-	4,900,968	CP002185	-	.	37	7	M9 sucrose
Escherichia coli	K-12 MG1655	47076	Transformed with pAKB	4,641,652	NC000913	-	Grows on sucrose, mNG fluorescent	37	7	M9 glucose
Escherichia coli	K-12 MG1655	47076	Transformed with pBKA	4,641,652	NC000913	-	Grows on sucrose, mNG fluorescent	37	7	M9 glucose
Corynebacterium glutamicum	534	13032	-	3,309,401	NC003450	+	Produces L-glutamate, L-glutamic acid	37	8	CG XII
Corynebacterium glutamicum	KY 10256	21526	-			+	Produces l-lysine	30	8	CG XII
Bacillus subtilis	168	23857	-	4,215,606	NC000964	+	.	30	8	
Bacillus licheniformis	CD-2	9945a	-	4,376,305	CP005965	+	Produces glutamyl polypeptide	37	8	
Saccharomyces cerevisae	S288c	204508	-	12,157,105	GCF000146045	+	.	30	7	

Appendix B: Culture apparatus tuning details

Peristaltic pumps were tuned and analyzed for precision and accuracy in dispensing media for continuous flow experiments (Fig. 2.7, 2.9, 2.11).

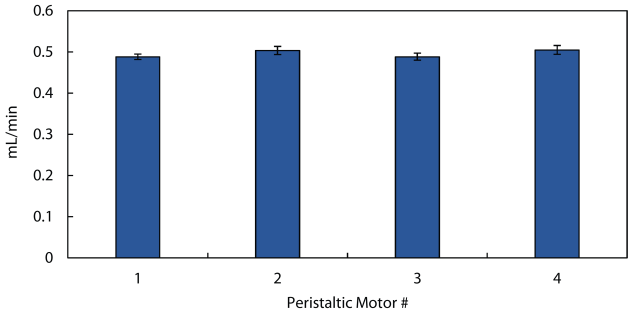


Figure B.1: Peristaltic motor dispensing DI H_2O (n = 10).

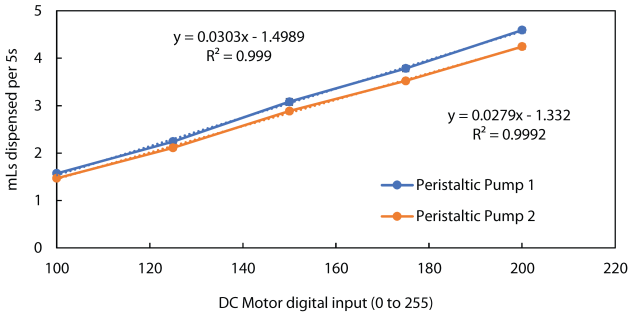


Figure B.2: DC motor calibration over a dynamic range of digital inputs (n = 5). The motors are highly linear with respect to digital input.

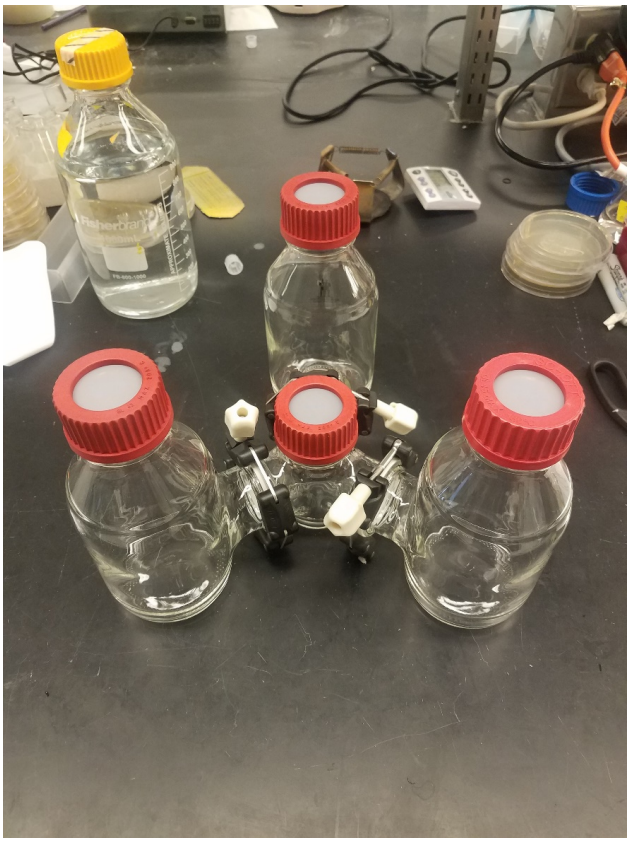
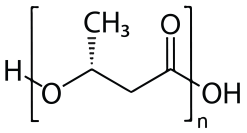


Figure B.3: Depiction of specialty separated tripartite culture vessel. The three main chambers (top, left, and right) are modified 500 mL Pyrex bottles. The middle chamber is a modified 100 mL Pyrex bottle. Flanges tighten the glass bottle connection, which consists of two plastic rings with a rubber washer between.

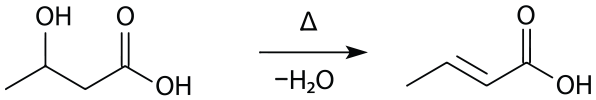
Appendix C: Characterization of polyhydroxyalkanoates produced by *A. vinelandii*

C.1 Analysis of natural PHB, commercial PHB, and crotonic acid

Both polyhydroxybutyrate granules (Sigma-aldrich) and trans crotonic acid (Sigma-aldrich) were used as standards to determine cellular PHB content. PHB is a variable chain length polymer of a monomeric 3-hydroxybutyric acid used by a number of bacterial species to store excess carbon under other nitrogen, phosphorus, or oxygen limitation [183]. It serves both as carbon storage and as a (Fig. C.2A) [185].



(a) PHB

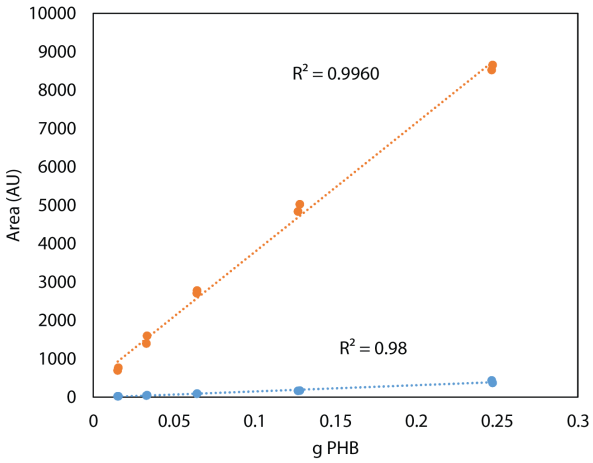


(b) Distillation of 3-hydroxybutyric acid to crotonic acid.

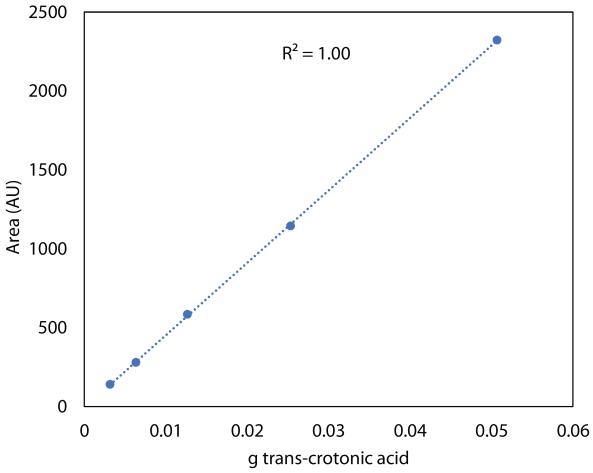
Figure C.1: Polyhydroxybutyrate structure and distillation of 3-hydroxybutyric acid to crotonic acid.

Fig. C.2B shows conversion of 3-hydroxybutyric acid to crotonic acid. This reaction occurs after acid hydrolysis of PHB polymer and may be quantified using either standard curves of commercially available PHB (Fig. C.2A) or crontonic acid (Fig. C.2B). PHB chain length and purity vary substantially between organisms and while literature tend to source PHB from commercial sources, these are generally denoted as "natural origin" rather than % purity, which makes HPLC confidence challenging. Standard curves were generated using both "natural origin" commercial PHB and pure crotonic acid (Fig. C.2).

Both HPLC standards were highly linear, though maintained significantly different slopes. Furthermore, PHB indicated two highly linear peaks with respect to hydrolyzed "natural origin" PHB, which could indicate the presence of some underlying impurity likely due to a different hydrolyzed polyhydroxyalkanoate (*e.g.*, polyhydroxyvalerate, polyhydroxybutyrate-



(a) HPLC standard curve of hydrolyzed PHB, which produced two highly linear peaks with respect to the mass of PHB hydrolyzed.



(b) HPLC standard curve of trans-crotonic acid, producing a single highly linear peak with respect to acid concentration.

Figure C.2: HPLC standard curves of hydrolyzed PHB and crotonic acid.

co-valerate, etc.). Elucidation of polymer characteristics and components would require mass spectrometry and was not investigated in this study.

C.2 Generation of PHB by *A. vinelandii* in monoculture and in tripartite culture

A. vinelandii was cultivated in Burk N-free medium (20 g/L) in a 4 L biofermentation unit (Bioflo 3000, NEB). Biofermentation conditions included 20% culture dilution per day and bubbling 1 L/min filtered air. Periodically 100 mL aliquots were desiccated, hydrolyzed, and analyzed by HPLC (Fig. C.3).

Fig. C.3 shows that the trans-crotonic acid samples consistently produced approximately 72% quantification of PHB content compared to the commercial "natural origin" samples. Deviation here stems from chemical reaction efficiency of PHB hydrolysis as well as the unknown purity of the commercial PHB samples. Nonetheless, there was between 35% to 60% PHB accumulation by weight in *A. vinelandii* under our biofermentation conditions. These values corroborate with under studies that have investigated PHB accumulation in *A. vinelandii* DJ derivatives.

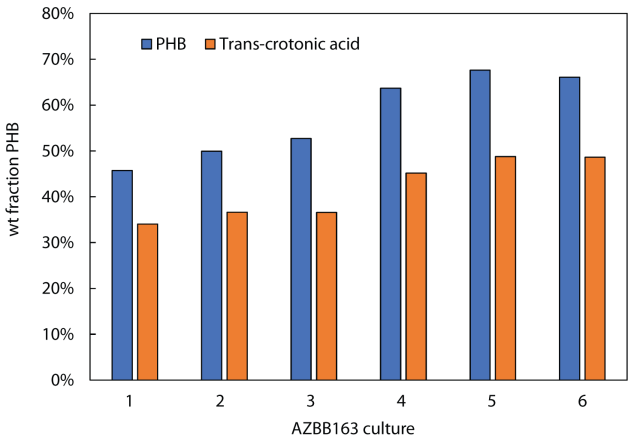
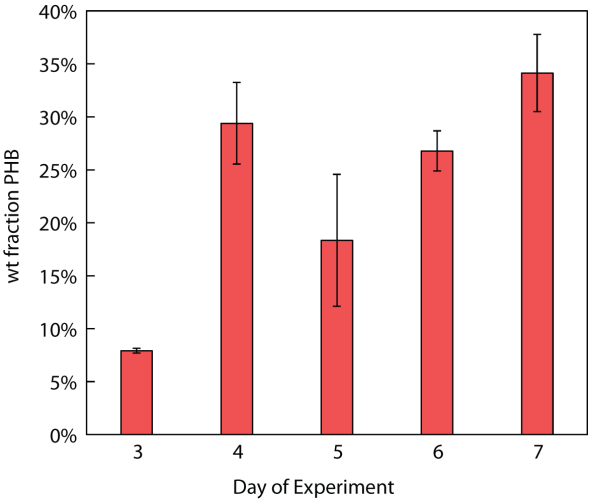


Figure C.3: Aliquots of 100 mL *A. vinelandii* (AZBB163) culture compared to both PHB and trans-crotonic acid standards.

Lastly, PHB content of the biofilm formed in the tripartite *A. vinelandii* partition was investigated (Fig. C.4).

Fig. C.4 shows that PHB content of the biofilm increased between days 3 and 7. Qualitatively further biofilm formation was not observed after day 7. This as well as the lower weight fraction compared to the biofermentation conditions are likely due to carbon limited conditions.

(a) Image of biofilm formation and accumulation in the 500 mL *A. vinelandii* partition



(b) Plot of wt fraction PHB of biofilm samples (technical replicates, n = 3) between days 3 and 7 of the tripartite system.

Figure C.4: Biofilm formation and HPLC analysis during a spatially separated tripartite experiment.

Appendix D: Curve fitting of monocultures using Monod kinetics

All cultures were fit as monocultures using Monod kinetics and empirical absorbance data. Cultures were well-described by Monod kinetics if absorbance was below 0.80 (λ = 750 nm). Absorbance data were used as a proxy for biomass. Fig. D.1 depicts curve fitting of *C. glutamicum* ATCC 13032 under sucrose limited conditions.

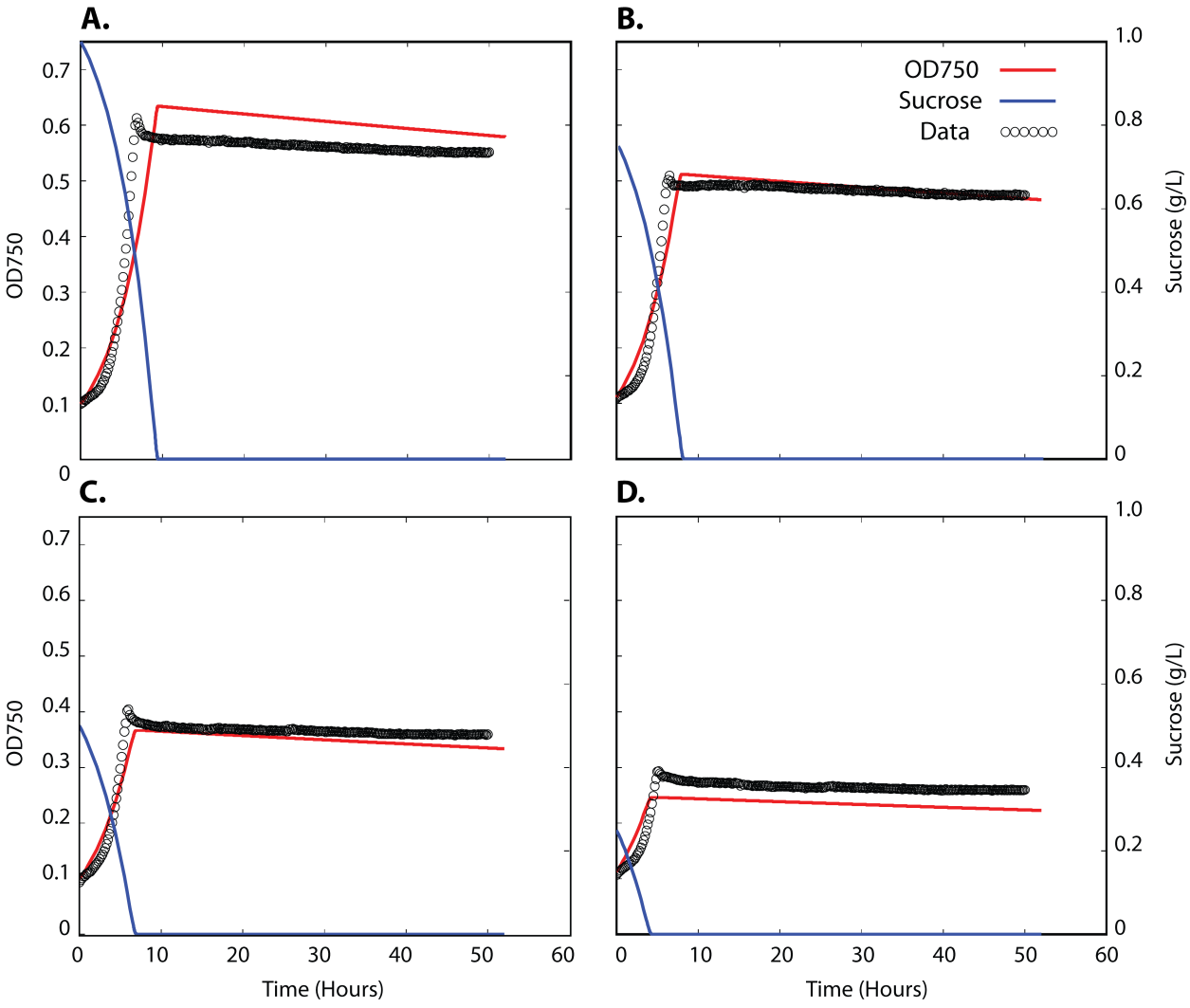


Figure D.1: MATLAB Levenberg-Marquardt fits of *C. glutamicum* 13032 growth at various sucrose concentrations in the union media. (A) 1.0 g/L sucrose, (B) 0.75 g/L sucrose, (C) 0.5 g/L sucrose, and (D) 0.25 g/L sucrose (n = 3).

Cultures were growth in a 96-well microplate in union medium at 30C with orbital shaking for approximately 50 hours. *C. glutamicum* parameters were optimized by an automated MATLAB script using Levenberg-Marquardt fits, which minimize sum of square error, for sucrose concentrations between 1.0 g/L and 0.25 g/L. The fits, depicted in Fig. D.1, are good and are representative of the best fit parameters for all empirical data and sucrose conditions. Deviations between empirical data and curve fitting may stem from physiological differences in culture growth at different carbon concentrations.

Appendix E: Synopsis of flow cytometry data

Cocultures of *S. elongatus* and *A. vinelandii* were studied extensively using flow cytometry (attune) after three days of growth in union minimal media at standard culture conditions. Front and side scatter (FSC, SSC) indicate somewhat differentiated cell types between bacteria, though cells are most easily differentiated using blue light (BL3-A, 488 nm), which clearly shows a tightly clustered population of fluorescent *S. elongatus* cells due to chl-*a* (Fig E.1).

Unfortunately, *A. vinelandii* also shows significant fluorescence under excitation at 488 nm due to siderophore production. Fig. E.1 illustrates that only a subpopulation of *A. vinelandii* cells appear to be fluorescent, spanning between 10 to 10^3 fluorescent units while there is a clustered population with no fluorescence. Although not shown, initial growth of *S. elongatus* or *A. vinelandii* tends to yield smaller cells and contributes to cell size differentiation over time. *E. coli* cells labeled with mNG or YFP also fluoresce under blue/green light, though are generally smaller than *Azotobacter*. Nonetheless, *A. vinelandii* and fluorescent *E. coli* cell populations maintain substantial overlap that make culture composition analysis difficult. Likewise, weakly fluorescent *S. elongatus* cells overlap with smaller, siderophore producing *A. vinelandii* cells.

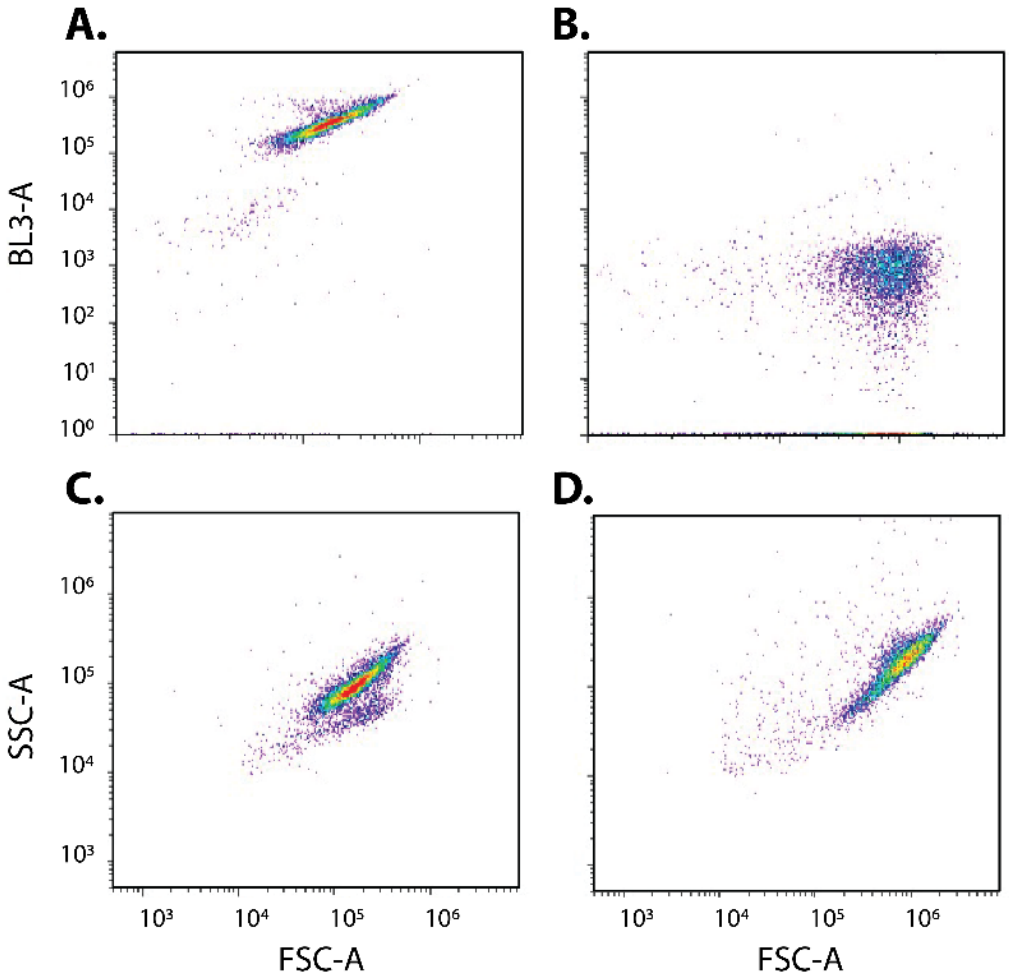


Figure E.1: FCS plots of *S. elongatus cscB* + SPS and *A. vinelandii* AZBB163. (A) *S. elongatus* population with blue laser and front scatter. (B) *A. vinelandii* population with blue laser and front scatter. (c) *S. elongatus* population with side and front scatter. (d) *A. vinelandii* population with side and front scatter.

Appendix F: Sensitivity analysis outputs

Sensitivity data were also plotted on a heatmap by scaling columns (outputs) and rows (inputs) to visualize interplay between inputs and outputs. Fig F.2, shows how certain inputs are especially negative (AZBB163, Y_x) or positive (*S. elongatus*, μ_{max}) such that they dominate the remaining outputs. Spearman's rank correlation coefficients are also listed in Fig. F.3.

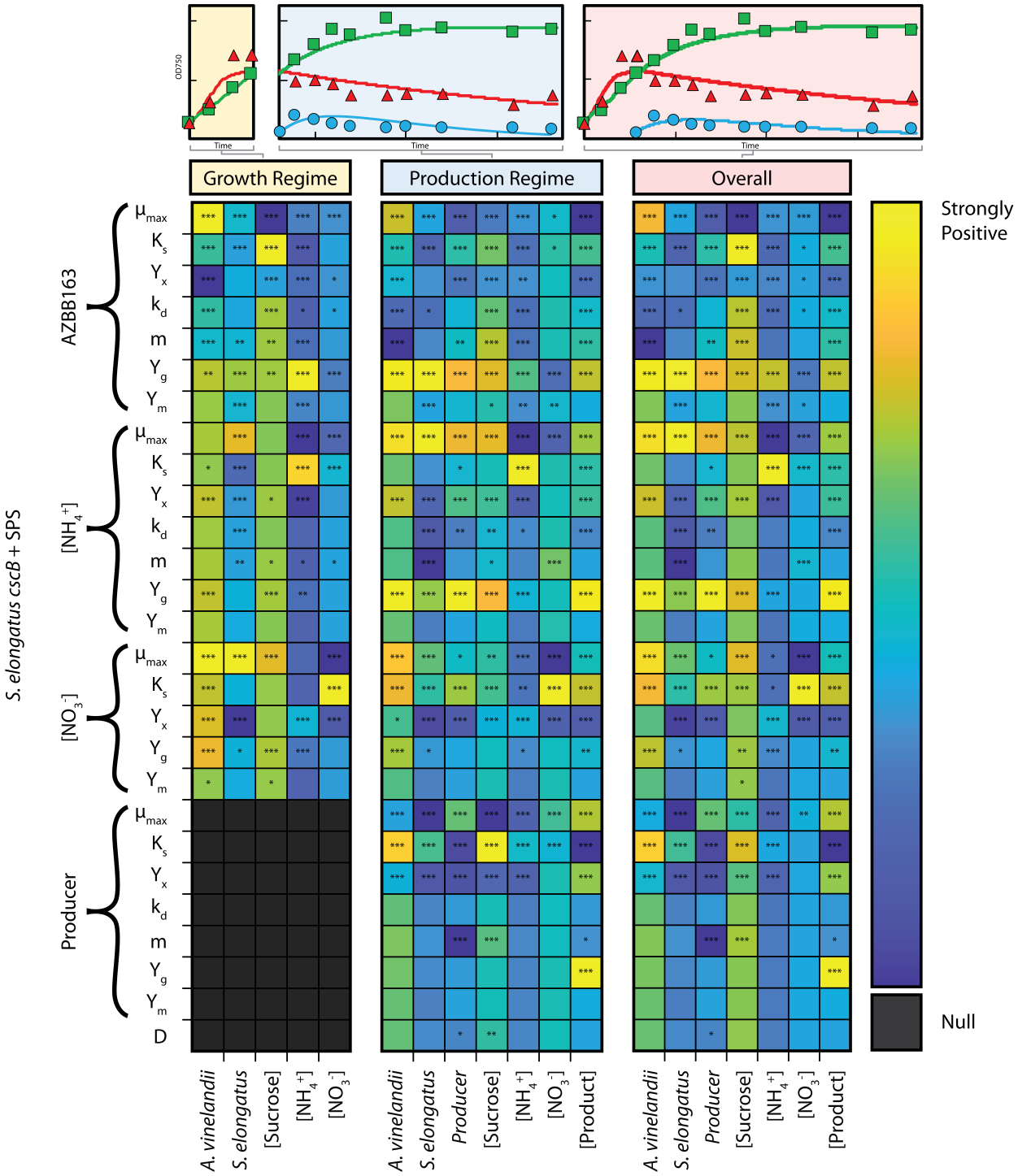


Figure F.1: Heatmap of the tripartite correlation analysis scaled by column (output) Spearman's rank correlation coefficients.

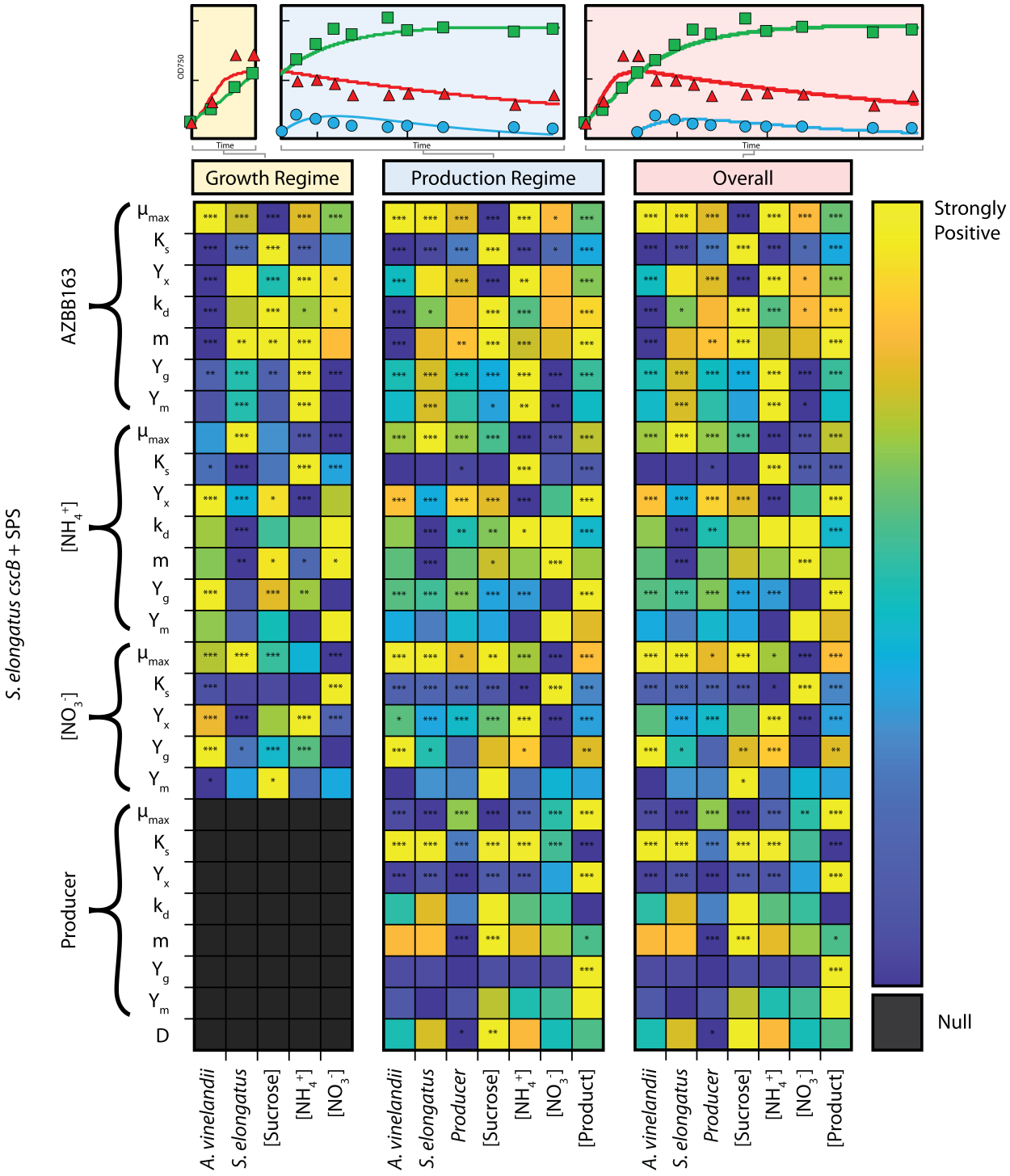


Figure F.2: Heatmap of the tripartite correlation analysis scaled by row (input) Spearman's rank correlation coefficients.

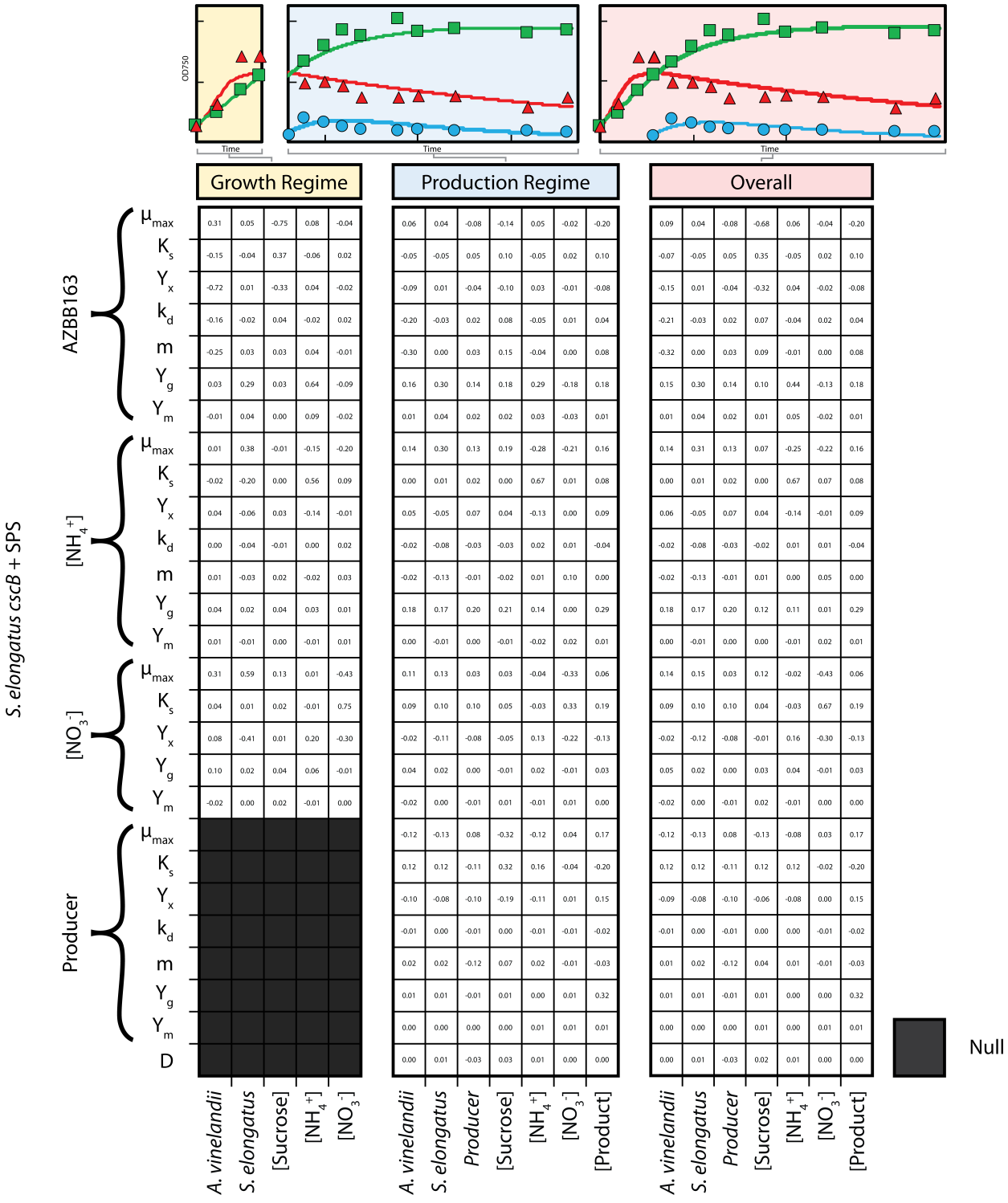


Figure F.3: Tripartite correlation analysis with using Spearman's rank correlation coefficients.

Appendix G: Statistics and description of the Algal HTL Module

Table G.1: Coefficients for energy return on invested (EROI) life cycle metric. Residual standard error: 0.4705 on 4987 degrees of freedom, multiple R-squared: 0.7792, adjusted R-squared: 0.7787, F-statistic: 1467 on 12 and 4987 DF, p-value: < 2.2e-16. Sig. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Input	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Sig.**
(Intercept)	0.000	0.007	0	1	
Biomass	1.152	0.010	113.3	< 2e-16	***
DM_P	0.048	0.011	4.3	2.18E-05	***
DM_N	0.011	0.011	1.0	3.28E-01	
BC	0.198	0.008	26.0	< 2e-16	***
BC_C	0.310	0.009	33.6	< 2e-16	***
BC_N	-0.039	0.008	-4.7	2.56E-06	***
NO_3	0.367	0.007	52.4	< 2e-16	***
PO_4	0.646	0.009	72.4	< 2e-16	***
NH4	0.003	0.007	0.5	0.639	
ACP_{PO4}	0.139	0.008	17.6	< 2e-16	***
ACP_{NH4}	0.113	0.009	12.1	< 2e-16	***
Crash	-0.136	0.007	-19.2	< 2e-16	***

Table G.2: Coefficients for greenhouse gas emissions (GHGs) life cycle metric. Residual standard error: 0.5038 on 4987 degrees of freedom,Multiple R-squared: 0.7468, Adjusted R-squared: 0.7461 , F-statistic: 1225 on 12 and 4987 DF, p-value: < 2.2e-16. Sig. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Input	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Sig.**
(Intercept)	0.000	0.000	0.000	0.000	
Biomass	-1.128	0.011	-103.6	< 2e-16	***
DM_P	-0.064	0.012	-5.3	1.29E-07	***
DM_N	-0.015	0.012	-1.3	0.203751	
BC	-0.176	0.008	-21.6	< 2e-16	***
BC_C	-0.323	0.010	-32.7	< 2e-16	***
BC_N	0.030	0.009	3.5	0.000566	***
NO_3	-0.344	0.007	-46.0	< 2e-16	***
PO_4	-0.643	0.010	-67.3	< 2e-16	***
NH_4	-0.003	0.008	-0.4	0.72086	
ACP_{PO4}	-0.121	0.008	-14.3	< 2e-16	***
ACP_{NH4}	-0.095	0.010	-9.6	< 2e-16	***
Crash	0.135	0.008	17.8	< 2e-16	***

Table G.3: Coefficients for water intensity (WI) life cycle metric. Residual standard error: 0.2858 on 4987 degrees of freedom, multiple R-squared: 0.9185, adjusted R-squared: 0.9183, F-statistic: 4684 on 12 and 4987 DF, p-value: < 2.2e-16. Sig. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Input	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Sig.**
(Intercept)	0.000	0.004	0.0	1	
Biomass	-0.989	0.006	-160.1	< 2e-16	***
DM_P	0.003	0.007	0.4	0.657066	
DM_N	-0.044	0.007	-6.6	4.21E-11	***
BC	-0.066	0.005	-14.3	< 2e-16	***
BC_C	-0.267	0.006	-47.6	< 2e-16	***
BC_N	-0.004	0.005	-0.8	0.396229	
NO_3	-0.106	0.004	-25.0	< 2e-16	***
PO_4	-1.082	0.005	-199.6	< 2e-16	***
NH_4	-0.014	0.004	-3.3	0.000848	***
ACP_{PO4}	-0.217	0.005	-45.2	< 2e-16	***
ACP_{NH4}	-0.051	0.006	-8.9	< 2e-16	***
Crash	0.034	0.004	7.9	2.41E-15	***

Table G.4: List of all inputs and calculations in the Algal HTL Module. The values are either explicitly stated or symbolized with "V" as they differ between species combinations and experimental conditions. Sources are listed as either "E" for empirical inputs, "A" for assumption, "S" for specification, or "C" for calculation. Sources with reference numbers are empirical inputs or assumptions from relevant literature. Many of the variables are analogous to those employed in the original Algal Process Description published by ANL.

Variable	Description	Process	Value	Units	Source
Biomass	Dry Mass Composition	Culture Parameter	V	g Biomass/L	E [62]
DM_P	Dry Mass Composition	Culture Parameter	V	g P/g-algae	E [150]
DM_N	Dry Mass Composition	Culture Parameter	V	g N/g-algae	E [150]
BC	Biocrude Composition	Culture Parameter	V	g biocrude/g_{algae}	E [61]
BC_C	Biocrude Composition	Culture Parameter	V	g C/$g_{afdw-algae}$	E [61]
BC_H	Biocrude Composition	Culture Parameter	V	g H/$g_{afdw-algae}$	E [61]
BC_N	Biocrude Composition	Culture Parameter	V	g N/$g_{afdw-algae}$	E [61]
BC_O	Biocrude Composition	Culture Parameter	V	g O/$g_{afdw-algae}$	E [61]
H/C Ratio	Biocrude Composition	Culture Parameter	V	mol H/mol C	E [61]
LHV	Biocrude Composition	Culture Parameter	V	MJ/kg	[157]
HHV	Biocrude Composition	Culture Parameter	V	MJ/kg	[157]
NO3	Exhausted Media Composition	Culture Parameter	V	g N/$g_{afdw-algae}$	E [150]
NH4	Exhausted Media Composition	Culture Parameter	V	g N/$g_{afdw-algae}$	E [150]
PO4	Exhausted Media Composition	Culture Parameter	V	g P/$g_{afdw-algae}$	E [150]
ACP_{PO4}	Aqueous Co-Product Composition	Culture Parameter	V	g P/$g_{afdw-algae}$	E [37]
ACP_{NH4}	Aqueous Co-Product Composition	Culture Parameter	V	g N/$g_{afdw-algae}$	E [37]
Algae mass	Other	Culture Parameter	V	g afdw-algae	E [61]
Solid mass	Other	Culture Parameter	V	g_{solid}	E [61]
Average Crashes (per yr)	Algal Productivity	Culture Parameter	V	#	E
Productivity EST.	Algal Productivity	Culture Parameter	V	g/m^2/day	E
Ash	Dry Mass Composition	Culture Parameter	V	g ash/$g_{afdw-algae}$	E [61, 150]
AL_C	Dry Mass Composition	Culture Parameter	V	g C/$g_{afdw-algae}$	E [61]
AL_H	Dry Mass Composition	Culture Parameter	V	g H/$g_{afdw-algae}$	E [150]
AL_N	Dry Mass Composition	Culture Parameter	V	g N/$g_{afdw-algae}$	E [61]
AL_O	Dry Mass Composition	Culture Parameter	V	g O/$g_{afdw-algae}$	C
$Total_{mass}$	Other	Culture Parameter	V	$g_{afdw-algae}$	E [61]
BC_{mass}	Biocrude Composition	Culture Parameter	V	g biocrude	E [61]
AQ_{mass}	Aqueous Co-Product Composition	Culture Parameter	V	g ACP	E [61]
AQ_Y	Aqueous Co-Product Composition	Culture Parameter	V	g ACP/g-algae	E [61]
$Extra_{mass}$	Other	Culture Parameter	V	g	E [61]
HTL_{mass}	Other	Culture Parameter	V	g	E [61]
SO_Y	Other	Culture Parameter	V	g-solid/g-total	E [61]
SO_{AY}	Other	Culture Parameter	V	g_{solid}/g_{algae}	E [61]
BC_{AY}	Other	Culture Parameter	V	g BC/g afdw-algae	E [61]

Rec_{mass}	Other	Culture Parameter	V	g	E [61]
Rec_{Y}	Other	Culture Parameter	V	g-rec/g-total	E [61]
Carbon	Atomic weight	Culture Parameter	12.01	g/mol	S
Nitrogen	Atomic weight	Culture Parameter	14.01	g/mol	S
Phosphorous	Atomic weight	Culture Parameter	30.97	g/mol	S
$NaNO_3$	Molecular weight	Culture Parameter	84.99	g/mol	S
K_2HPO_4	Molecular weight	Culture Parameter	174.2	g/mol	S
KH_2PO4	Molecular weight	Culture Parameter	136.1	g/mol	S
$NaNO_{33}$	BOLD 3N	Culture Parameter	0.745	g/L	[179]
K_2HPO_4	BOLD 3N	Culture Parameter	0.0749	g/L	[179]
KH_2PO_4	BOLD 3N	Culture Parameter	0.176	g/L	[179]
Nitrogen	BOLD 3N	Culture Parameter	0.00882	N mol/L	[179]
Phosphorus	BOLD 3N	Culture Parameter	0.00172	P mol/L	[179]
New N source	Nutrient supply summary	Culture Parameter	V	g NaNO3/g-algae	C
New K_2HPO_4 Source	Nutrient supply summary	Culture Parameter	V	g K_2HPO_4/g-algae	C
New KH_2PO_4 Source	Nutrient supply summary	Culture Parameter	V	g KH_2PO_4/g-algae	C
New N per cycle	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
New N per cycle	Nitrogen recycling	Recycle stream	V	%	C
N in the Pond	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in the Pond	Nitrogen recycling	Recycle stream	V	%	C
N in Dry Mass	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Dry Mass	Nitrogen recycling	Recycle stream	V	%	C
N in Dewatering (Exhausted Media)	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Dewatering (Exhausted Media)	Nitrogen recycling	Recycle stream	V	%	C
N in Biocrude	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Biocrude	Nitrogen recycling	Recycle stream	V	%	C
N in ACP	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in ACP	Nitrogen recycling	Recycle stream	V	%	C
N in Upgrading	Nitrogen recycling	Recycle stream	0	g N/g-algae	A
N in Upgrading	Nitrogen recycling	Recycle stream	0	%	A
N in Recycling Stream	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Recycling Stream	Nitrogen recycling	Recycle stream	V	%	C
N in Dry Mass Losses	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Dry Mass Losses	Nitrogen recycling	Recycle stream	V	%	C
N in Biocrude Losses	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Biocrude Losses	Nitrogen recycling	Recycle stream	V	%	C
N in Upgrading Losses	Nitrogen recycling	Recycle stream	V	g N/g-algae	C
N in Upgrading Losses	Nitrogen recycling	Recycle stream	V	%	C
New P per cycle	Phosphorus recycling	Recycle stream	V	g P/g-algae	C

New P per cycle	Phosphorus recycling	Recycle stream	V	%	C
P in Pond	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in Pond	Phosphorus recycling	Recycle stream	V	%	C
P in dry mass	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in dry mass	Phosphorus recycling	Recycle stream	V	%	C
P in dewatering (Exhausted Media)	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in dewatering (Exhausted Media)	Phosphorus recycling	Recycle stream	V	%	C
P in biocrude	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in biocrude	Phosphorus recycling	Recycle stream	V	%	C
P in ACP	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in ACP	Phosphorus recycling	Recycle stream	V	%	C
P in Upgrading	Phosphorus recycling	Recycle stream	0	g P/g-algae	A
P in Upgrading	Phosphorus recycling	Recycle stream	0	%	A
P in recycling stream	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in recycling stream	Phosphorus recycling	Recycle stream	V	C	
P in dry mass losses	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in dry mass losses	Phosphorus recycling	Recycle stream	V	%	C
P in biocrude losses	Phosphorus recycling	Recycle stream	V	g P/g-algae	C
P in biocrude losses	Phosphorus recycling	Recycle stream	V	%	C
P in upgrading losses	Phosphorus recycling	Recycle stream	0	g P/g-algae	A
P in upgrading losses	Phosphorus recycling	Recycle stream	0	%	A
New NaNO3 per cycle	Recycling Summary	Recycle stream	V	g NaNO3/g afdw-algae	C
New K2HPO4 per cycle	Recycling Summary	Recycle stream	V	g K2HPO4/g afdw-algae	C
New KH_2PO_4 per cycle	Recycling Summary	Recycle stream	V	g KH_2PO_4/g afdw-algae	C
H/C Ratio	Biocrude Composition	Upgrading	V	g/g-algae	C
$(H/C)_{RD}$	Renewable Diesel Target H/C	Upgrading	2	g/g-algae	[165, 186]
(H/C) Light hydrocarbons	H/C of light hydrocarbons	Upgrading	2.67	g/g-algae	[165, 186]
Carbon in renewable diesel	Carbon Efficiency	Upgrading	0.95	#	[165, 186]
Carbon in light hydrocarbons	Carbon Efficiency	Upgrading	0.95	#	[165, 186]
RD Yield	Upgrading yield from biocrude	Upgrading	V	g RD/g biocrude	C
H Added	Hydrogen demand for upgrading to RD	Upgrading	V	g H/g RD	C
H_2 Added	Hydrogen demand for upgrading to RD	Upgrading	V	g H_2/g RD	C
LHV of H_2	Property	Upgrading	290	BTU/ft3	S
Hydrogen Density	Property	Upgrading	2.6	g/ft^3	S
H_2	Unit Conversion	Upgrading	111.5	BTU/g H_2 using LHV	C
H_2/RD	Hydrogen for upgrading biocrude	Upgrading	V	BTU/g RD	C
H_2/RD	Hydrogen for upgrading biocrude	Upgrading	V	BTU H_2/lb RD	C
MJ to BTU	Unit Conversion	Upgrading	947.8	BTU/MJ	C
LHV of RD	RD Lower Heating Value	Upgrading	V	BTU/kg	C

Air Transport Energy	Open Pond	Growth and first dewatering	3.04E-05	kWh/g-algae	[165]
Water use for evaporative cooling	Open Pond	Growth and first dewatering	0	L/g-algae	S
Energy for evaporative cooling	Open Pond	Growth and first dewatering	0	kWh/g-algae	S
Culture concentration (w/v)	Open Pond	Growth and first dewatering	V	g afdw-algae/L	C
Energy to pump water to site/culture	Pond / Reactor Common Parameters	Growth and first dewatering	1.23E-04	kWh/L	[187]
Energy to pump culture	Pond / Reactor Common Parameters	Growth and first dewatering	2.50E-05	kWh/L	[187]
Energy cost factor	Pond / Reactor Common Parameters	Growth and first dewatering	V	#	C
CO_2 total supply rate	Pond / Reactor Common Parameters	Growth and first dewatering	0	g/g-algae	S
Recycled carbon	Pond / Reactor Common Parameters	Growth and first dewatering	0	g C/g afdw-algae	C
Specific productivity	Pond Parameters	Growth and first dewatering	V	$g/m^2/d$	[187]
Circulation power for pond	Pond Parameters	Growth and first dewatering	48	kWh/ha/d	[187]
Avg. pan evaporation	Pond Parameters	Growth and first dewatering	0.3	in/day	[165]
Lake/pan correction factor	Pond Parameters	Growth and first dewatering	0.75	%	[165]
Evaporative losses from pond	Pond Parameters	Growth and first dewatering	5.715	$L/m^2/d$	C
Evaporative losses from pond	Pond Parameters	Growth and first dewatering	V	L/g-algae	C
CO_2 loss	Pond Parameters	Growth and first dewatering	0	#	S
Total circulation power for pond	Pond Parameters	Growth and first dewatering	V	kWh/g	C
First dewatering algae concentration	First Dewatering Parameters	Growth and first dewatering	10	g afdw-algae/L	C
First dewatering water loss	First Dewatering Parameters	Growth and first dewatering	0	L/net g	C
First dewatering returned to process	First Dewatering Parameters	Growth and first dewatering	1	Fraction	S
Algae retention efficiency	First Dewatering Parameters	Growth and first dewatering	0.9	Fraction	[158]
Growth media	Water flows	Growth and first dewatering	V	L/g afdw-algae	C
Total water	Water flows	Growth and first dewatering	V	L/g afdw-algae	C
Water sent to $R_{dewatering}$	Water flows	Growth and first dewatering	0.1	L/g afdw-algae	C
Water separated in first dewater	Water flows	Growth and first dewatering	V	$L/g_{afdw-algae}$	C
Water sent to waste	Water flows	Growth and first dewatering	0	$L/g_{afdw-algae}$	S
Water returned for media make-up	Water flows	Growth and first dewatering	V	$L/g_{afdw-algae}$	C
Water returned from $R_{dewatering}$	Water flows	Growth and first dewatering	V	L/g afdw-algae	C
New water introduced for media make-up	Water flows	Growth and first dewatering	V	L/g afdw-algae	C
New water introduced for cooling	Water flows	Growth and first dewatering	0	L/g afdw-algae	S
Energy for water for media	Water flows	Growth and first dewatering	V	kWh/g afdw-algae	C
Energy for new water for cooling	Water flows	Growth and first dewatering	0	kWh/g afdw-algae	S
Energy to move culture to dewatering	Water flows	Growth and first dewatering	V	kWh/g afdw-algae	C
Water consumption	Water flows	Growth and first dewatering	V	L/g afdw-algae	C
Energy for spray cooling circulation	Water Consumption	Growth and first dewatering	0	kWh/g afdw-algae	S
Summary of water movement	Summary of water movement	Growth and first dewatering	V	mL/g afdw-algae	C
Summary of GFD electricity	Summary of GFD electricity	Growth and first dewatering	V	kWh/g afdw-algae	C
Initial solids	Centrifugation	Remaining Dewatering	6	%	S
Final solids	Centrifugation	Remaining Dewatering	15	#	S

Disc-stack	Centrifugation	Remaining Dewatering	1.83E-05	kWh/g afdw-algae	[64]
Recovery Efficiency	Centrifugation	Remaining Dewatering	95	%	[65]
Water returned to culture	Centrifugation	Remaining Dewatering	0.0109	L/g afdw-algae	C
Algae returned to culture	Centrifugation	Remaining Dewatering	0.0526	g/g-out	C
Retention efficiency	DAF	Remaining Dewatering	90	#	[188]
Return lost algae to culture	DAF	Remaining Dewatering	yes	#	S
Initial solids	DAF	Remaining Dewatering	10	g afdw-algae/L	S
Output solids	DAF	Remaining Dewatering	60	g afdw-algae/L	S
Power consumption	DAF	Remaining Dewatering	0.000133	kWh/g afdw-algae	[64, 188]
Chitosan Efficiency	Flocculants / Coagulants	Remaining Dewatering	90	#	[189]
Chitosan Dose	Flocculants / Coagulants	Remaining Dewatering	40	mg/L	[187]
Chitosan	Flocculants / Coagulants	Remaining Dewatering	0.004	g/g afdw-algae	C
Water returned to culture	DAF	Remaining Dewatering	0.0944	L/g afdw-algae	C
Algae returned to culture	DAF	Remaining Dewatering	0.111	g/g-out	C
Total Rdewatering water recycle to culture	Process Summary	Remaining Dewatering	0.105	L/g afdw-algae	C
Total Rdewatering algae recycle to culture	Process Summary	Remaining Dewatering	0.164	kWh/g afdw-algae	C
Total Rdewatering electricity	Process Summary	Remaining Dewatering	1.67E-04	kWh/g afdw-algae	C
Ash content of algae	Dry Mass Composition	Extraction	V	g ash/g algae	C
Non ash content	Dry Mass Composition	Extraction	V	g afdw-algae/g algae	C
N, wt% of oil	HTL Oil description	Extraction	V	%	C
P, wt% of feed lost to solids and oil	HTL Oil description	Extraction	V	%	C
Nitrogen loss	HTL Oil description	Extraction	V	%	C
Solid content, on ash free basis	Reaction Parameters	Extraction	15	%	S
Solid content, including ash	Reaction Parameters	Extraction	V	%	C
Temperature	Reaction Parameters	Extraction	20	C	S
Reaction Temperature	Reaction Parameters	Extraction	350	C	S
Reaction Pressure	Reaction Parameters	Extraction	16530000	Pa	S
Density	Reaction Parameters	Extraction	1000	kg/m3	A
cp for slurry	Reaction Parameters	Extraction	4.12	kJ/kg/K	S
cp of ACP	Reaction Parameters	Extraction	4.12	kJ/kg/K	S
cp of water at 320C	Reaction Parameters	Extraction	5.89	kJ/kg/K	C
cp of water at T exiting HX	Reaction Parameters	Extraction	5.913	kJ/kg/K	C
Efficiency, combined pump and motor	High pressure pump	Extraction	75	%	[158]
Pumping energy cost	High pressure pump	Extraction	22040	J/kg slurry	C
Pumping energy	High pressure pump	Extraction	V	J/g biocrude	C
HTL oil	Products	Extraction	V	g biocrude/g afdw-algae	C
Solids	Products	Extraction	V	g solids/g afdw-algae	C
Solid residue	Products	Extraction	V	g/g afdw-algae	C
Total Gas	Products	Extraction	V	g/g afdw-algae	C

CO_2 fraction of gas	Products	Extraction	1	wt% of total gas	[158]
Organic C in aqueous	Products	Extraction	V	g C/g afdw-algae	C
Mass into HX	Products	Extraction	1	%	C
Effectiveness	Heat exchanger	Extraction	85	%	[158, 190]
Temperature of hot stream	Heat exchanger	Extraction	20	C	S
Temperature of cold stream	Heat exchanger	Extraction	350	C	S
Temperature of hot stream after HX	Heat exchanger	Extraction	69.5	C	C
Temperature of cold stream after HX	Heat exchanger	Extraction	300.5	C	C
Total pumping energy	Energy input	Extraction	V	J/g biocrude	C
HTL heat demand	Energy input	Extraction	292.7	kJ/kg slurry	C
Temperature rise entering CHG	Energy input	Extraction	30	C	[158]
CHG heat demand	Energy input	Extraction	V	kJ/kg slurry	C
Heat demand	Energy input	Extraction	V	J/g biocrude	C
Carbon in algae	Carbon balance	Extraction	V	%	C
Carbon recovered in gas	Carbon balance	Extraction	1	%	[158]
Carbon recovered in solid	Carbon balance	Extraction	6	%	[158]
Carbon wt% of oil	Carbon balance	Extraction	V	%	C
Carbon recovered in oil	Carbon balance	Extraction	V	Fraction	C
Sum of partitions	Carbon balance	Extraction	1	Fraction	C
Biogas yield on organic	CHG Integrated with HTL	Extraction	0.99	Fraction	[45, 165]
CH4 fraction	CHG Integrated with HTL	Extraction	0.6	Fraction	[45, 165]
Electricity demand	CHG Integrated with HTL	Extraction	0.00031	kWh/g-CH4	[158]
ACP yield on biocrude	CHG Integrated with HTL	Extraction	V	g organic-C/g biocrude	C
Methane yield	CHG Integrated with HTL	Extraction	V	g CH4/g organic-C	C
CO_2 yield	CHG Integrated with HTL	Extraction	V	g CO_2/g organic-C	C
Electricity demand	CHG Integrated with HTL	Extraction	V	kWh/g organic-C	C
Pond Liner Thickness	Design Selections	Infrastructure	40	mm	[161]
Pond Concrete	Material Lifetimes	Infrastructure	30	years	[161]
HDPE Liners	Material Lifetimes	Infrastructure	30	years	[161]
LDPE Covers	Material Lifetimes	Infrastructure	30	years	[161]
Paddlewheel Steel	Material Lifetimes	Infrastructure	30	years	[161]
Settling Tank Concrete	Material Lifetimes	Infrastructure	30	years	[161]
DAF Concrete	Material Lifetimes	Infrastructure	30	years	[161]
Centrifuge Steel	Material Lifetimes	Infrastructure	30	years	[161]
Centrifuge Concrete	Material Lifetimes	Infrastructure	30	years	[161]
Homogenizer Steel	Material Lifetimes	Infrastructure	30	years	[161]
Homogenizer Concrete	Material Lifetimes	Infrastructure	30	years	[161]
Column Steel	Material Lifetimes	Infrastructure	30	years	[161]
Column Concrete	Material Lifetimes	Infrastructure	30	years	[161]

Column Packing Lifetime	Material Lifetimes	Infrastructure	30	years	[161]
AD Concrete	Material Lifetimes	Infrastructure	30	years	[161]
AD Steel	Material Lifetimes	Infrastructure	30	years	[161]
Makeup Water PVC	Material Lifetimes	Infrastructure	30	years	[161]
Makeup Water Concrete	Material Lifetimes	Infrastructure	30	years	[161]
Cast Iron Pump	Material Lifetimes	Infrastructure	30	years	[161]
Gas Distribution Concrete	Material Lifetimes	Infrastructure	30	years	[161]
Gas Blowers	Material Lifetimes	Infrastructure	30	years	[161]
Operating Days Per Year	Material Lifetimes	Infrastructure	330	days	[161]
Pond Width	Large Growth Ponds	Infrastructure	60	m	[161]
Pond Length	Large Growth Ponds	Infrastructure	690	m	[161]
Number of Ponds	Large Growth Ponds	Infrastructure	1013	ponds	[161]
Berm Height	Large Growth Ponds	Infrastructure	0.9	m	[161]
Concrete Volume for Sump	Large Growth Ponds	Infrastructure	53	m^3	[161]
Unlined Pond Plastic Surface Area	Large Growth Ponds	Infrastructure	9420	m^2	[161]
Lined Pond Plastic Surface Area	Large Growth Ponds	Infrastructure	49400	m^2	[161]
Internal Surface Area	Large Growth Ponds	Infrastructure	40627.4	m^2	[161]
Volume of Excavated Material	Large Growth Ponds	Infrastructure	V	m^3	[161]
Area Factor for Production Operations	Culture Parameter	Infrastructure	2.43E-08	$1/m^2$	[161]
Total Facility Area for Distribution	Culture Parameter	Infrastructure	8093712	m^2	[161]
Area Factor for Distribution	Culture Parameter	Infrastructure	1.24E-07	$1/m^2$	[161]
Scaled Productivity 1	Culture Parameter	Infrastructure	V	#	C
Scaled Productivity 2	Culture Parameter	Infrastructure	V	#	C
Excavation Volume	Large Growth Pond Construction	Infrastructure	9.09E-05	$m^3/m^2/day$	[161]
Excavation Volume	Inoculum Pond Construction	Infrastructure	0.00001	$m3/m^2/day$	[161]
Excavation Volume	Settling Tank Construction	Infrastructure	V	$m^3/m^2/day$	[161]
Excavation Volume	DAF Construction	Infrastructure	V	$m^3/m^2/day$	[161]
Excavation Volume	AD Construction	Infrastructure	V	$m^3/m^2/day$	[161]
Excavation Volume	Makeup Water Piping	Infrastructure	V	$m^3/m^2/day$	[161]
HDPE Liner	Large Growth Pond Construction	Infrastructure	1.21E-04	$kg/m^2/day$	[161]
Polypropylene	Large Growth Pond Construction	Infrastructure	2.49E-05	$kg/m^2/day$	[161]
PVC	Makeup Water Piping	Infrastructure	V	$kg/m^2/day$	[161]
LDPE Plastic Cover	Inoculum Pond Construction	Infrastructure	1.40E-07	$kg/m^2/day$	[161]
HDPE Liner	Inoculum Pond Construction	Infrastructure	1.33E-05	$kg/m^2/day$	[161]
Polypropylene	Inoculum Pond Construction	Infrastructure	2.74E-06	$kg/m^2/day$	[161]
Concrete	Large Growth Pond Construction	Infrastructure	2.77E-04	$kg/m^2/day$	[161]
Concrete	CO_2 Distribution System	Infrastructure	1.92E-04	$kg/m^2/day$	[161]
Concrete	Inoculum Pond Construction	Infrastructure	3.04E-05	$kg/m^2/day$	[161]
Concrete	Settling Tank Construction	Infrastructure	V	$kg/m^2/day$	[161]

Concrete	DAF Construction	Infrastructure	V	kg/m^2/day	[161]
Concrete	Extraction Centrifugation Construction	Infrastructure	V	kg/m^2/day	[161]
Concrete	Dewatering Centrifugation Construction	Infrastructure	V	kg/m^2/day	[161]
Concrete	AD Construction	Infrastructure	V	kg/m^2/day	[161]
Concrete	Makeup Water Piping	Infrastructure	V	kg/m^2/day	[161]
Steel	Dewatering Centrifugation Construction	Infrastructure	V	kg/m^2/day	[161]
Steel	Extraction Centrifugation Construction	Infrastructure	V	kg/m^2/day	[161]
Steel	Homogenizer Construction	Infrastructure	V	kg/m^2/day	[161]
Steel	Liquid/Liquid Extraction Column	Infrastructure	V	kg/m^2/day	[161]
Steel	Stripping Column	Infrastructure	V	kg/m^2/day	[161]
Steel	AD Construction	Infrastructure	V	kg/m^2/day	[161]
Steel	CO_2 Distribution System	Infrastructure	8.32E-07	kg/m^2/day	[161]
Cast Iron	Makeup Water Piping	Infrastructure	V	kg/m^2/day	[161]

References

[1] Emily Harvey, Jeffrey Heys, and Tomáš Gedeon. Quantifying the effects of the division of labor in metabolic pathways. *Journal of theoretical biology*, 360:222–242, 2014.

[2] Sara Mitri and Kevin Richard Foster. The genotypic view of social interactions in microbial communities. *Annual review of genetics*, 47:247–273, 2013.

[3] Alan R Pacheco and Daniel Segrè. A multidimensional perspective on microbial interactions. *FEMS microbiology letters*, 366(11):fnz125, 2019.

[4] Laurent Philippot, Jos M Raaijmakers, Philippe Lemanceau, and Wim H Van Der Putten. Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology*, 11(11):789–799, 2013.

[5] P. Bhattacharyya and Dhruva K Jha. Plant growth-promoting rhizobacteria (pgpr): emergence in agriculture. *World Journal of Microbiology and Biotechnology*, 28(4):1327–1350, 2012.

[6] EI Newman. The rhizosphere: carbon sources and microbial populations. *Ecological interactions in soil: plants, microbes and animals*, pages 107–121, 1985.

[7] MS Girvan, CD Campbell, K Killham, James Ivor Prosser, and Lesley Anne Glover. Bacterial diversity promotes community stability and functional resilience after perturbation. *Environmental microbiology*, 7(3):301–313, 2005.

[8] Patrick D Schloss and Jo Handelsman. The last word: books as a statistical metaphor for microbial communities. *Annual review of microbiology*, 61, 2007.

[9] Bruce E Rittmann, Martina Hausner, Frank Loffler, Nancy G Love, Gerard Muyzer, Satoshi Okabe, Daniel B Oerther, Jordan Peccia, Lutgarde Raskin, and Michael Wagner. A vista for microbial ecology and environmental biotechnology, 2006.

[10] J Shine and L Dalgarno. The 3'-terminal sequence of escherichia coli 16s ribosomal rna: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences*, 71(4):1342–1346, 1974.

[11] Reshma P Shetty, Drew Endy, and Thomas F Knight. Engineering biobrick vectors from biobrick parts. *Journal of biological engineering*, 2(1):5, 2008.

[12] Yili Qian, Hsin-Ho Huang, José I Jiménez, and Domitilla Del Vecchio. Resource competition shapes the response of genetic circuits. *ACS synthetic biology*, 6(7):1263–1272, 2017.

[13] Michael B Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[14] Melissa B Miller and Bonnie L Bassler. Quorum sensing in bacteria. *Annual Reviews in Microbiology*, 55(1):165–199, 2001.

[15] Jordi Garcia-Ojalvo, Michael B Elowitz, and Steven H Strogatz. Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proceedings of the National Academy of Sciences*, 101(30):10955–10960, 2004.

[16] Frederick K Balagaddé, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You. A synthetic escherichia coli predator–prey ecosystem. *Molecular systems biology*, 4(1), 2008.

[17] Pazhamalai Anbarasan, Zachary C Baer, Sanil Sreekumar, Elad Gross, Joseph B Binder, Harvey W Blanch, Douglas S Clark, and F Dean Toste. Integration of chemical catalysis with extractive fermentation to produce fuels. *Nature*, 491(7423):235, 2012.

[18] P Biller and AB Ross. Potential yields and properties of oil from the hydrothermal liquefaction of microalgae with different biochemical content. *Bioresource technology*, 102(1):215–225, 2011.

[19] Chris J Paddon and Jay D Keasling. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology*, 12(5):355–367, 2014.

[20] Hans C Bernstein, Steven D Paulson, and Ross P Carlson. Synthetic escherichia coli consortia engineered for syntrophy demonstrate enhanced biomass productivity. *Journal of biotechnology*, 157(1):159–166, 2012.

[21] Yusuf Chisti. Biodiesel from microalgae. *Biotechnology advances*, 25(3):294–306, 2007.

[22] Laura Garcia Alba, Cristian Torri, Chiara Samorì, Jaapjan van der Spek, Daniele Fabbri, Sascha RA Kersten, and Derk WF Brilman. Hydrothermal treatment (htt) of microalgae: evaluation of the process as conversion method in an algae biorefinery concept. *Energy & fuels*, 26(1):642–657, 2011.

[23] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[24] Alissa Kerner, Jihyang Park, Audra Williams, and Xiaoxia Nina Lin. A programmable escherichia coli consortium via tunable symbiosis. *PLoS One*, 7(3), 2012.

[25] David N Carruthers, Chang Kyu Byun, Bradley J Cardinale, and Xiaoxia Nina Lin. Demonstration of transgressive overyielding of algal mixed cultures in microdroplets. *Integrative Biology*, 9(8):687–694, 2017.

[26] Corine M Jackman, Kyle W Deans, Larry J Forney, and Xiaoxia Nina Lin. Microdroplet co-cultivation and interaction characterization of human vaginal bacteria. *Integrative Biology*, 11(3):69–78, 2019.

[27] Tatyana E Saleski, Alissa R Kerner, Meng Ting Chung, Corine M Jackman, Azzaya Khasbaatar, Katsuo Kurabayashi, and Xiaoxia Nina Lin. Syntrophic co-culture amplification of production phenotype for high-throughput screening of microbial strain libraries. *Metabolic engineering*, 54:232–243, 2019.

[28] Kenneth A Johnson and Roger S Goody. The original michaelis constant: translation of the 1913 michaelis–menten paper. *Biochemistry*, 50(39):8264–8269, 2011.

[29] Jacques Monod. The growth of bacterial cultures. *Annual review of microbiology*, 3(1):371–394, 1949.

[30] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.

[31] Gad Yagil and Ezra Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophysical journal*, 11(1):11, 1971.

[32] Tom Bradley, Daniel Maga, and Sara Antón. Unified approach to life cycle assessment between three unique algae biofuel facilities. *Applied Energy*, 154:1052–1061, 2015.

[33] Jeremy J Minty, Marc E Singer, Scott A Scholz, Chang-Hoon Bae, Jung-Ho Ahn, Clifton E Foster, James C Liao, and Xiaoxia Nina Lin. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences*, 110(36):14592–14597, 2013.

[34] Ismail Rawat, R Ranjith Kumar, Taurai Mutanda, and Faizal Bux. Biodiesel from microalgae: a critical evaluation from laboratory to large scale production. *Applied energy*, 103:444–467, 2013.

[35] Tara Shirvani, Xiaoyu Yan, Oliver R Inderwildi, Peter P Edwards, and David A King. Life cycle energy and greenhouse gas analysis for algae-derived biodiesel. *Energy & Environmental Science*, 4(10):3773–3778, 2011.

[36] David Conchouso, Ehab Al Rawashdeh, Arpys Arevalo, D Castro, and IG Foulds. Simulation of a 3d flow-focusing capillary-based droplet generator. In *COMSOL Conference, Rotterdam, Netherlands*.

[37] Casey M Godwin, David C Hietala, Aubrey R Lashaway, Anita Narwani, Phillip E Savage, and Bradley J Cardinale. Algal polycultures enhance coproduct recycling from hydrothermal liquefaction. *Bioresource technology*, 224:630–638, 2017.

[38] Paul J Crutzen, Arvin R Mosier, Keith A Smith, and Wilfried Winiwarter. *N2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels*, pages 227–238. Springer, 2016.

[39] Pierre Collet, Arnaud Hélias, Laurent Lardon, Monique Ras, Romy-Alice Goy, and Jean-Philippe Steyer. Life-cycle assessment of microalgae culture coupled to biogas production. *Bioresource technology*, 102(1):207–214, 2011.

[40] Nico Schneider, Tara J Fortin, Roland Span, and Mandy Gerber. Thermophysical properties of the marine microalgae nannochloropsis salina. *Fuel Processing Technology*, 152:390–398, 2016.

[41] Edgard Gnansounou and Jegannathan Kenthorai Raman. Life cycle assessment of algae biodiesel and its co-products. *Applied Energy*, 161:300–308, 2016.

[42] Ray Dixon and Daniel Kahn. Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology*, 2(8):621–631, 2004.

[43] Charles AS Hall, Stephen Balogh, and David JR Murphy. What is the minimum eroi that a sustainable society must have? *Energies*, 2(1):25–47, 2009.

[44] National Research Council. *Sustainable development of algal biofuels in the United States*. National Academies Press, 2013.

[45] D C Elliott, G G Neuenschwander, and T R Hart. Combined hydrothermal liquefaction and catalytic hydrothermal gasification system and process for conversion of biomass feedstocks, 2017. US Patent 9,758,728.

[46] Juan Cesar Federico Ortiz-Marquez, Mauro Do Nascimento, and Leonardo Curatti. Metabolic engineering of ammonium release for nitrogen-fixing multispecies microbial cell-factories. *Metabolic engineering*, 23:154–164, 2014.

[47] Evelyne A Groen, Eddie AM Bokkers, Reinout Heijungs, and Imke JM de Boer. Methods for global sensitivity analysis in life cycle assessment. *The International Journal of Life Cycle Assessment*, 22(7):1125–1137, 2017.

[48] David Tilman, Peter B Reich, Johannes Knops, David Wedin, Troy Mielke, and Clarence Lehman. Diversity and productivity in a long-term grassland experiment. *Science*, 294(5543):843–845, 2001.

[49] Jie Pan, Anna L Stephenson, Elena Kazamia, Wilhelm TS Huck, John S Dennis, Alison G Smith, and Chris Abell. Quantitative tracking of the growth of individual algal cells in microdroplet compartments. *Integrative Biology*, 3(10):1043–1051, 2011.

[50] Bradley J Cardinale, Justin P Wright, Marc W Cadotte, Ian T Carroll, Andy Hector, Diane S Srivastava, Michel Loreau, and Jerome J Weis. Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the national academy of sciences*, 104(46):18123–18128, 2007.

[51] Garrett Hardin. The competitive exclusion principle. *science*, 131(3409):1292–1297, 1960.

[52] Man Kee Lam and Keat Teong Lee. Microalgae biofuels: a critical review of issues, problems and the way forward. *Biotechnology advances*, 30(3):673–690, 2012.

[53] Douglas C Elliott. Review of recent reports on process technology for thermochemical conversion of whole algae to liquid fuels. *Algal Research*, 13:255–263, 2016.

[54] Justin Hoffman, Ronald C Pate, Thomas Drennen, and Jason C Quinn. Techno-economic assessment of open microalgae production systems. *Algal Research*, 23:51–57, 2017.

[55] Kirill A Datsenko and Barry L Wanner. One-step inactivation of chromosomal genes in escherichia coli k-12 using pcr products. *Proceedings of the National Academy of Sciences*, 97(12):6640–6645, 2000.

[56] Thomas E Kuhlman and Edward C Cox. Site-specific chromosomal integration of large synthetic constructs. *Nucleic acids research*, 38(6):e92–e92, 2010.

[57] Colin T Archer, Jihyun F Kim, Haeyoung Jeong, Jin Hwan Park, Claudia E Vickers, Sang Yup Lee, and Lars K Nielsen. The genome sequence of e. coli w (atcc 9637): comparative genome analysis and an improved genome-scale reconstruction of e. coli. *BMC genomics*, 12(1):9, 2011.

[58] Michele Bruschi, Simon J Boyes, Haryadi Sugiarto, Lars K Nielsen, and Claudia E Vickers. A transferable sucrose utilization approach for non-sucrose-utilizing escherichia coli strains. *Biotechnology advances*, 30(5):1001–1010, 2012.

[59] Matthias Finkbeiner, Atsushi Inaba, Reginald Tan, Kim Christiansen, and Hans-Jürgen Klüppel. The new international standards for life cycle assessment: Iso 14040 and iso 14044. *The international journal of life cycle assessment*, 11(2):80–85, 2006.

[60] Yongli Zhang, Xiaowei Liu, Mark A White, and Lisa M Colosi. Economic evaluation of algae biodiesel based on meta-analyses. *International Journal of Sustainable Energy*, 36(7):682–694, 2017.

[61] David C Hietala, Cristina K Koss, Anita Narwani, Aubrey R Lashaway, Casey M Godwin, Bradley J Cardinale, and Phillip E Savage. Influence of biodiversity, biochemical composition, and species identity on the quality of biomass and biocrude oil produced via hydrothermal liquefaction. *Algal Research*, 26:203–214, 2017.

[62] Anita Narwani, Aubrey R Lashaway, David C Hietala, Phillip E Savage, and Bradley J Cardinale. Power of plankton: Effects of algal biodiversity on biocrude production and stability. *Environmental science & technology*, 50(23):13142–13150, 2016.

[63] M Wang, Ye Wu, and Amgad Elgowainy. Operating manual for greet: Version 1.7. *Center for Transportation Research, Energy Systems Division, Argonne National Laboratory, Iowa*, page 119, 2007.

[64] Ryan E Davis, Daniel B Fishman, Edward D Frank, Michael C Johnson, Susanne B Jones, Christopher M Kinchin, Richard L Skaggs, Erik R Venteris, and Mark S Wigmosta. Integrated evaluation of cost, emissions, and resource potential for algal biofuels at the national scale. *Environmental science & technology*, 48(10):6035–6042, 2014.

[65] Ryan Davis, Daniel Fishman, Edward D Frank, Mark S Wigmosta, Andy Aden, Andre M Coleman, Philip T Pienkos, Richard J Skaggs, Erik R Venteris, and Michael Q Wang. Renewable diesel from algal lipids: an integrated baseline for cost, emissions, and resource potential from a harmonized model. Report, National Renewable Energy Laboratory (NREL), Golden, CO., 2012.

[66] Aaron M Appel, John E Bercaw, Andrew B Bocarsly, Holger Dobbek, Daniel L DuBois, Michel Dupuis, James G Ferry, Etsuko Fujita, Russ Hille, Paul JA Kenis, et al. Fron-

tiers, opportunities, and challenges in biochemical and chemical catalysis of co2 fixation. *Chemical reviews*, 113(8):6621–6658, 2013.

[67] Gary G Mittelbach and Brian J McGill. *Community ecology*. Oxford University Press, Sunderland, Mass., 2019.

[68] Suhas G Jadhav, Prakash D Vaidya, Bhalchandra M Bhanage, and Jyeshtharaj B Joshi. Catalytic carbon dioxide hydrogenation to methanol: A review of recent studies. *Chemical Engineering Research and Design*, 92(11):2557–2567, 2014.

[69] Daniel C Ducat, J Abraham Avelar-Rivas, Jeffrey C Way, and Pamela A Silver. Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol.*, 78(8):2660–2668, 2012.

[70] Edgardo Ortiz-Reyes and Robert P Anex. Life cycle environmental impacts of non-cellulosic fermentable carbohydrates for the production of biofuels and chemicals. *The International Journal of Life Cycle Assessment*, 25(3):548–563, 2020.

[71] João C Setubal, Patricia Dos Santos, Barry S Goldman, Helga Ertesvåg, Guadelupe Espin, Luis M Rubio, Svein Valla, Nalvo F Almeida, Divya Balasubramanian, and Lindsey Cromes. Genome sequence of azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *Journal of bacteriology*, 191(14):4534–4545, 2009.

[72] Ray Dixon. The oxygen-responsive nifl-nifa complex: a novel two-component regulatory system controlling nitrogenase synthesis in γ-proteobacteria. *Archives of Microbiology*, 169(5):371–380, 1998.

[73] A Bali, Gonzalo Blanco, Susan Hill, and Christina Kennedy. Excretion of ammonium by a nifl mutant of azotobacter vinelandii fixing nitrogen. *Applied and Environmental Microbiology*, 58(5):1711–1718, 1992.

[74] Brett M Barney, Lauren J Eberhart, Janet M Ohlert, Carolann M Knutson, and Mary H Plunkett. Gene deletions resulting in increased nitrogen release by azotobacter vinelandii: application of a novel nitrogen biosensor. *Applied and environmental microbiology*, 81(13):4316–4328, 2015.

[75] Juan Cesar Federico Ortiz-Marquez, Mauro Do Nascimento, Maria de los Angeles Dublan, and Leonardo Curatti. Association with an ammonium-excreting bacterium allows diazotrophic culture of oil-rich eukaryotic microalgae. *Applied and environmental microbiology*, 78(7):2345–2352, 2012.

[76] Rafael Ambrosio, Juan Cesar Federico Ortiz-Marquez, and Leonardo Curatti. Metabolic engineering of a diazotrophic bacterium improves ammonium release and biofertilization of plants and microalgae. *Metabolic engineering*, 40:59–68, 2017.

[77] Juan A Villa, Erin E Ray, and Brett M Barney. Azotobacter vinelandii siderophore can provide nitrogen to support the culture of the green algae neochloris oleoabundans and scenedesmus sp. ba032. *FEMS microbiology letters*, 351(1):70–77, 2014.

[78] Stephanie G Hays and Daniel C Ducat. Engineering cyanobacteria as photosynthetic feedstock factories. *Photosynthesis research*, 123(3):285–295, 2015.

[79] Henrike Niederholtmeyer, Bernd T Wolfstädter, David F Savage, Pamela A Silver, and Jeffrey C Way. Engineering cyanobacteria to synthesize and export hydrophilic products. *Applied and environmental microbiology*, 76(11):3462–3466, 2010.

[80] Daniel C Ducat, J Abraham Avelar-Rivas, Jeffrey C Way, and Pamela A Silver. Rerouting carbon flux to enhance photosynthetic productivity. *Applied and environmental microbiology*, 78(8):2660–2668, 2012.

[81] Bradley W Abramson, Benjamin Kachel, David M Kramer, and Daniel C Ducat. Increased photochemical efficiency in cyanobacteria via an engineered sucrose sink. *Plant and Cell Physiology*, 57(12):2451–2460, 2016.

[82] Stephanie G Hays, Leo LW Yan, Pamela A Silver, and Daniel C Ducat. Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. *Journal of biological engineering*, 11(1):4, 2017.

[83] Taylor L Weiss, Eric J Young, and Daniel C Ducat. A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production. *Metabolic engineering*, 44:236–245, 2017.

[84] Derek T Fedeson, Pia Saake, Patricia Calero, Pablo Iván Nikel, and Daniel C Ducat. Biotransformation of 2, 4-dinitrotoluene in a phototrophic co-culture of engineered synechococcus elongatus and pseudomonas putida. *Microbial biotechnology*, 2020.

[85] Matthew J Smith and Matthew B Francis. A designed a. vinelandii–s. elongatus coculture for chemical photoproduction from air, water, phosphate, and trace metals. *ACS synthetic biology*, 5(9):955–961, 2016.

[86] Patricia C Dos Santos. *Molecular biology and genetic engineering in nitrogen fixation*, pages 81–92. Springer, 2011.

[87] Simon Unthan, Alexander Grünberger, Jan van Ooyen, Jochem Gätgens, Johanna Heinrich, Nicole Paczia, Wolfgang Wiechert, Dietrich Kohlheyer, and Stephan Noack. Beyond growth rate 0.6: What drives corynebacterium glutamicum to higher growth rates in defined medium. *Biotechnology and bioengineering*, 111(2):359–371, 2014.

[88] Suriana Sabri, Lars K Nielsen, and Claudia E Vickers. Molecular control of sucrose utilization in escherichia coli w, an efficient sucrose-utilizing strain. *Appl. Environ. Microbiol.*, 79(2):478–487, 2013.

[89] Jörn Kalinowski, Brigitte Bathe, Daniela Bartels, Nicole Bischoff, Michael Bott, Andreas Burkovski, Nicole Dusch, Lothar Eggeling, Bernhard J Eikmanns, and Lars Gaigalat. The complete corynebacterium glutamicum atcc 13032 genome sequence

and its impact on the production of l-aspartate-derived amino acids and vitamins. *Journal of biotechnology*, 104(1-3):5–25, 2003.

[90] Lothar Eggeling and Michael Bott. *Handbook of Corynebacterium glutamicum*. CRC press, 2005.

[91] Paul Vos, George Garrity, Dorothy Jones, Noel R Krieg, Wolfgang Ludwig, Fred A Rainey, Karl-Heinz Schleifer, and William B Whitman. *Bergey's manual of systematic bacteriology: Volume 3: The Firmicutes*, volume 3. Springer science & Business Media, 2011.

[92] David C White, Susan D Sutton, and David B Ringelberg. The genus sphingomonas: physiology and ecology. *Current Opinion in Biotechnology*, 7(3):301–306, 1996.

[93] Mi-Sung Yim, Yvonne CW Yau, Anne Matlow, Jae-Seong So, Jitao Zou, Cecily A Flemming, Heidi Schraft, and Kam Tin Leung. A novel selective growth medium-pcr assay to isolate and detect sphingomonas in environmental samples. *Journal of microbiological methods*, 82(1):19–27, 2010.

[94] Amé María Valeria, Echenique Jose Ricardo, Pflugmacher Stephan, and Wunderlin Daniel Alberto. Degradation of microcystin-rr by sphingomonas sp. cba4 isolated from san roque reservoir (córdoba–argentina). *Biodegradation*, 17(5):447–455, 2006.

[95] Michael Gatheru Waigi, Kai Sun, and Yanzheng Gao. Sphingomonads in microbe-assisted phytoremediation: tackling soil pollution. *Trends in biotechnology*, 35(9):883–899, 2017.

[96] H Scott Fogler. *Essentials of Chemical Reaction Engineering: Essenti Chemica Reactio Engi*. Pearson Education, 2010.

[97] HL Smith. Bacterial growth. *Arizona State University*, page 14, 2006.

[98] Karin Kovárová-Kovar and Thomas Egli. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. *Microbiol. Mol. Biol. Rev.*, 62(3):646–666, 1998.

[99] Manfred Zinn, Bernard Witholt, and Thomas Egli. Dual nutrient limited growth: models, experimental observations, and applications. *Journal of biotechnology*, 113(1-3):263–279, 2004.

[100] Urs Lendenmann and Thomas Egli. Kinetic models for the growth of escherichia coli with mixtures of sugars under carbon-limited conditions. *Biotechnology and bioengineering*, 59(1):99–107, 1998.

[101] Hannes Löwe, Karina Hobmeier, Manuel Moos, Andreas Kremling, and Katharina Pflüger-Grau. Photoautotrophic production of polyhydroxyalkanoates in a synthetic mixed culture of synechococcus elongatus cscb and pseudomonas putida cscab. *Biotechnology for biofuels*, 10(1):190, 2017.

[102] Ratul Saha, Nabaneeta Saha, Robert S Donofrio, and Lorelle L Bestervelt. Microbial siderophores: a mini review. *Journal of basic microbiology*, 53(4):303–317, 2013.

[103] Lance C Seefeldt, Brian M Hoffman, and Dennis R Dean. Mechanism of mo-dependent nitrogenase. *Annual review of biochemistry*, 78:701–722, 2009.

[104] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[105] Dana Boyd, David S Weiss, Joseph C Chen, and Jon Beckwith. Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda inch, a simple escherichia coli plasmid-chromosome shuttle system. *Journal of Bacteriology*, 182(3):842–847, 2000.

[106] Jérôme Wong Ng, Didier Chatenay, Jérôme Robert, and Michael Guy Poirier. Plasmid copy number noise in monoclonal populations of bacteria. *Physical Review E*, 81(1):011909, 2010.

[107] Suriana Sabri, Jennifer A Steen, Mareike Bongers, Lars K Nielsen, and Claudia E Vickers. Knock-in/knock-out (kiko) vectors for rapid integration of large dna sequences, including whole metabolic pathways, onto the escherichia coli chromosome at well-characterised loci. *Microbial cell factories*, 12(1):60, 2013.

[108] François St-Pierre, Lun Cui, David G Priest, Drew Endy, Ian B Dodd, and Keith E Shearwin. One-step cloning and chromosomal integration of dna. *ACS synthetic biology*, 2(9):537–541, 2013.

[109] Marcelo C Bassalo, Andrew D Garst, Andrea L Halweg-Edwards, William C Grau, Dylan W Domaille, Vivek K Mutalik, Adam P Arkin, and Ryan T Gill. Rapid and efficient one-step metabolic pathway integration in e. coli. *ACS synthetic biology*, 5(7):561–568, 2016.

[110] Yu Jiang, Biao Chen, Chunlan Duan, Bingbing Sun, Junjie Yang, and Sheng Yang. Multigene editing in the escherichia coli genome via the crispr-cas9 system. *Appl. Environ. Microbiol.*, 81(7):2506–2514, 2015.

[111] Michael E Pyne, Murray Moo-Young, Duane A Chung, and C Perry Chou. Coupling the crispr/cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in escherichia coli. *Appl. Environ. Microbiol.*, 81(15):5103–5114, 2015.

[112] Jun Fu, Silke C Wenzel, Olena Perlova, Junping Wang, Frank Gross, Zhiru Tang, Yulong Yin, A Francis Stewart, Rolf Müller, and Youming Zhang. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic acids research*, 36(17):e113–e113, 2008.

[113] Patricia Nevers and Heinz Saedler. Transposable genetic elements as agents of gene instability and chromosomal rearrangements. *Nature*, 268(5616):109–115, 1977.

[114] Kyoung-Hee Choi and Herbert P Schweizer. mini-tn 7 insertion in bacteria with single att tn 7 sites: example pseudomonas aeruginosa. *Nature protocols*, 1(1):153–161, 2006.

[115] Igor Y Goryshin, Jerry Jendrisak, Les M Hoffman, Ronald Meis, and William S Reznikoff. Insertional transposon mutagenesis by electroporation of released tn5 transposition complexes. *Nature biotechnology*, 18(1):97–100, 2000.

[116] Jack A Bryant, Laura E Sellars, Stephen JW Busby, and David J Lee. Chromosome position effects on gene expression in escherichia coli k-12. *Nucleic acids research*, 42(18):11383–11392, 2014.

[117] Agustino Martínez-Antonio, Alejandra Medina-Rivera, and Julio Collado-Vides. Structural and functional map of a bacterial nucleoid. *Genome biology*, 10(12):247, 2009.

[118] Scott A Scholz, Rucheng Diao, Michael B Wolfe, Elayne M Fivenson, Xiaoxia Nina Lin, and Peter L Freddolino. High-resolution mapping of the escherichia coli chromosome reveals positions of high and low transcription. *Cell systems*, 8(3):212–225. e9, 2019.

[119] Colin T Archer, Jihyun F Kim, Haeyoung Jeong, Jin Hwan Park, Claudia E Vickers, Sang Yup Lee, and Lars K Nielsen. The genome sequence of e. coli w (atcc 9637): comparative genome analysis and an improved genome-scale reconstruction of e. coli. *BMC genomics*, 12(1):9, 2011.

[120] Knut Jahreis, Lars Bentler, Jürgen Bockmann, Stephan Hans, Astrid Meyer, Jörg Siepelmeyer, and Joseph W Lengeler. Adaptation of sucrose metabolism in the escherichia coli wild-type strain ec3132. *Journal of bacteriology*, 184(19):5307–5316, 2002.

[121] H Tsunekawa, S Azuma, M Okabe, R Okamoto, and S Aiba. Acquisition of a sucrose utilization system in escherichia coli k-12 derivatives and its application to industry. *Appl. Environ. Microbiol.*, 58(6):2081–2088, 1992.

[122] Elsayed T Mohamed, Hemanshu Mundhada, Jenny Landberg, Isaac Cann, Roderick I Mackie, Alex Toftgaard Nielsen, Markus J Herrgård, and Adam M Feist. Generation of an e. coli platform strain for improved sucrose utilization using adaptive laboratory evolution. *Microbial cell factories*, 18(1):116, 2019.

[123] Ryan A LaCroix, Troy E Sandberg, Edward J O'Brien, Jose Utrilla, Ali Ebrahim, Gabriela I Guzman, Richard Szubin, Bernhard O Palsson, and Adam M Feist. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of escherichia coli k-12 mg1655 on glucose minimal medium. *Appl. Environ. Microbiol.*, 81(1):17–30, 2015.

[124] Christopher P Long, Jacqueline E Gonzalez, Adam M Feist, Bernhard O Palsson, and Maciek R Antoniewicz. Fast growth phenotype of e. coli k-12 from adaptive laboratory evolution does not require intracellular flux rewiring. *Metabolic engineering*, 44:100–107, 2017.

[125] Yalun Arifin, Suriana Sabri, Haryadi Sugiarto, Jens O Krömer, Claudia E Vickers, and Lars K Nielsen. Deletion of cscr in escherichia coli w improves growth and poly-3-hydroxybutyrate (phb) production from sucrose in fed batch culture. *Journal of biotechnology*, 156(4):275–278, 2011.

[126] T. E. Saleski, M. T. Chung, D. N. Carruthers, A. Khasbaatar, K. Kurabayashi, and X. N. Lin. Stable and optimized heterologous pathway expression from the escherichia coli chromosome via high-throughput screening of random site integration libraries. *(Manuscript in preparation)*, 2020.

[127] UniProt Consortium. Uniprot: a worldwide hub of protein knowledge. *Nucleic acids research*, 47(D1):D506–D515, 2019.

[128] Ingrid M Keseler, Amanda Mackie, Alberto Santos-Zavaleta, Richard Billington, César Bonavides-Martínez, Ron Caspi, Carol Fulcher, Socorro Gama-Castro, Anamika Kothari, and Markus Krummenacker. The ecocyc database: reflecting new knowledge about escherichia coli k-12. *Nucleic acids research*, 45(D1):D543–D550, 2017.

[129] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Molecular systems biology*, 2(1), 2006.

[130] Antonino Baez, Kwang-Myung Cho, and James C Liao. High-flux isobutanol production using engineered escherichia coli: a bioreactor study with in situ product removal. *Applied microbiology and biotechnology*, 90(5):1681–1690, 2011.

[131] Xin Wang, James F Preston, and Tony Romeo. The pgaabcd locus of escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *Journal of bacteriology*, 186(9):2724–2734, 2004.

[132] KYOJI Yamaguchi and Masayori Inouye. Lipoprotein 28, an inner membrane protein of escherichia coli encoded by nlpa, is not essential for growth. *Journal of bacteriology*, 170(8):3747–3749, 1988.

[133] Carmen Schwechheimer and Meta J Kuehn. Synthetic effect between envelope stress and lack of outer membrane vesicle production in escherichia coli. *Journal of bacteriology*, 195(18):4161–4173, 2013.

[134] Zhongge Zhang, Jérôme N Feige, Abraham B Chang, Iain J Anderson, Vadim M Brodianski, Alexei G Vitreschak, Mikhail S Gelfand, and Milton H Saier. A transporter

of escherichia coli specific for l-and d-methionine is the prototype for a new family within the abc superfamily. *Archives of Microbiology*, 180(2):88–100, 2003.

[135] Troy E Sandberg, Christopher P Long, Jacqueline E Gonzalez, Adam M Feist, Maciek R Antoniewicz, and Bernhard O Palsson. Evolution of e. coli on [u-13c] glucose reveals a negligible isotopic influence on metabolism and physiology. *PLoS One*, 11(3), 2016.

[136] Douglas McCloskey, Sibei Xu, Troy E Sandberg, Elizabeth Brunk, Ying Hefner, Richard Szubin, Adam M Feist, and Bernhard O Palsson. Evolution of gene knockout strains of e. coli reveal regulatory architectures governed by metabolism. *Nature communications*, 9(1):1–15, 2018.

[137] Erol Kavvas, Adam Feist, Maciek R Antoniewicz, Bernhard Palsson, Jonathan Monk, Yang Ding, and Chris Long. Laboratory evolution of multiple e. coli strains reveals unifying principles of adaptation but diversity in driving genotypes. *bioRxiv*, 2020.

[138] Shota Atsumi, Taizo Hanai, and James C Liao. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *nature*, 451(7174):86, 2008.

[139] Jennifer A Steen, Nina Bohlke, Claudia E Vickers, and Lars K Nielsen. The trehalose phosphotransferase system (pts) in e. coli w can transport low levels of sucrose that are sufficient to facilitate induction of the csc sucrose catabolism operon. *PloS one*, 9(2), 2014.

[140] Kang Zhou, Lihan Zhou, Qing'En Lim, Ruiyang Zou, Gregory Stephanopoulos, and Heng-Phon Too. Novel reference genes for quantifying transcriptional responses of escherichia coli to protein overexpression by quantitative pcr. *BMC molecular biology*, 12(1):18, 2011.

[141] Jennifer B Dunn, Felix Adom, Norm Sather, Jeongwoo Han, Seth Snyder, Chang He, Jian Gong, Dajun Yue, and Fengqi You. Life-cycle analysis of bioproducts and their

conventional counterparts in greet. Report, Argonne National Lab.(ANL), Argonne, IL (United States), 2015.

[142] Jeong Wook Lee, Sol Choi, Jin Hwan Park, Claudia E Vickers, Lars K Nielsen, and Sang Yup Lee. Development of sucrose-utilizing escherichia coli k-12 strain by cloning β-fructofuranosidases and its application for l-threonine production. *Applied microbiology and biotechnology*, 88(4):905–913, 2010.

[143] D Ryan Georgianna and Stephen P Mayfield. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature*, 488(7411):329–335, 2012.

[144] David Tilman and John A Downing. *Biodiversity and stability in grasslands*, pages 3–7. Springer, 1996.

[145] Robert C McBride, Salvador Lopez, Chris Meenach, Mike Burnett, Philip A Lee, Fiona Nohilly, and Craig Behnke. Contamination management in low cost open algae ponds for biofuels production. *Industrial Biotechnology*, 10(3):221–227, 2014.

[146] Pierre Collet, Arnaud Hélias, Laurent Lardon, Jean-Philippe Steyer, and Olivier Bernard. Recommendations for life cycle assessment of algal fuels. *Applied Energy*, 154:1089–1102, 2015.

[147] Asha Parmar, Niraj Kumar Singh, Ashok Pandey, Edgard Gnansounou, and Datta Madamwar. Cyanobacteria and microalgae: A positive prospect for biofuels. *Bioresource Technology*, 102(22):10163–10172, 2011.

[148] Jonathan B Shurin, Rachel L Abbott, Michael S Deal, Garfield T Kwan, Elena Litchman, Robert C McBride, Shovon Mandal, and Val H Smith. Industrial-strength ecology: trade-offs and opportunities in algal biofuel production. *Ecology letters*, 16(11):1393–1404, 2013.

[149] Steven W Van Ginkel, Micah Bidwell, Thomas Igou, Ruben Gijon-Felix, Emille Jeane Novais Ribeiro Salvi, Silvia Helena Rego De Oliveira, Luis Henrique Kalil Duarte, Djessica Steiner, Zixuan Hu, and Rachel Johnston. The prevention of saltwater algal pond contamination using the electron transport chain disruptor, rotenone. *Algal Research*, 18:209–212, 2016.

[150] Casey M Godwin, David C Hietala, Aubrey R Lashaway, Anita Narwani, Phillip E Savage, and Bradley J Cardinale. Ecological stoichiometry meets ecological engineering: Using polycultures to enhance the multifunctionality of algal biocrude systems. *Environmental science & technology*, 51(19):11450–11458, 2017.

[151] Deborah T Newby, Teresa J Mathews, Ron C Pate, Michael H Huesemann, Todd W Lane, Bradley D Wahlen, Shovon Mandal, Robert K Engler, Kevin P Feris, and Jon B Shurin. Assessing the potential of polyculture to accelerate algal biofuel production. *Algal Research*, 19:264–277, 2016.

[152] John Sheehan, Terri Dunahay, John Benemann, and Paul Roessler. Look back at the us department of energy's aquatic species program: biodiesel from algae; close-out report. Report, National Renewable Energy Lab., Golden, CO.(US), 1998.

[153] Deborah L Sills, Vidia Paramita, Michael J Franke, Michael C Johnson, Tal M Akabas, Charles H Greene, and Jefferson W Tester. Quantitative uncertainty analysis of life cycle assessment for algal biofuel production. *Environmental science & technology*, 47(2):687–694, 2012.

[154] Philip Kenny and Kevin J Flynn. In silico optimization for production of biomass and biofuel feedstocks from microalgae. *Journal of applied phycology*, 27(1):33–48, 2015.

[155] Casey M Godwin, Aubrey R Lashaway, David C Hietala, Phillip E Savage, and Bradley J Cardinale. Biodiversity improves the ecological design of sustainable biofuel systems. *GCB Bioenergy*, 2018.

[156] Diego López Barreiro, Wolter Prins, Frederik Ronsse, and Wim Brilman. Hydrothermal liquefaction (htl) of microalgae for biofuel production: state of the art review and future prospects. *Biomass and Bioenergy*, 53:113–127, 2013.

[157] SA Channiwala and PP Parikh. A unified correlation for estimating hhv of solid, liquid and gaseous fuels. *Fuel*, 81(8):1051–1063, 2002.

[158] Edward D Frank, Amgad Elgowainy, Jeongwoo Han, and Zhichao Wang. Life cycle comparison of hydrothermal liquefaction and lipid extraction pathways to renewable diesel from algae. *Mitigation and Adaptation Strategies for Global Change*, 18(1):137–158, 2013.

[159] Nolan D Orfield, Andrew J Fang, Peter J Valdez, Michael C Nelson, Phillip E Savage, Xiaoxia Nina Lin, and Gregory A Keoleian. Life cycle design of an algal biorefinery featuring hydrothermal liquefaction: effect of reaction conditions and an alternative pathway including microbial regrowth. *ACS Sustainable Chemistry and Engineering*, 2(4):867–874, 2014.

[160] Ian C Woertz, John R Benemann, Niu Du, Stefan Unnasch, Dominick Mendola, B Greg Mitchell, and Tryg J Lundquist. Life cycle ghg emissions from microalgal biodiesel–a ca-greet model. *Environmental science & technology*, 48(11):6060–6068, 2014.

[161] Christina E Canter, Ryan Davis, Meltem Urgun-Demirtas, and Edward D Frank. Infrastructure associated emissions for renewable diesel production from microalgae. *Algal Research*, 5:195–203, 2014.

[162] Philip Kenny and Kevin J Flynn. Coupling a simple irradiance description to a mechanistic growth model to predict algal production in industrial-scale solar-powered photobioreactors. *Journal of applied phycology*, 28(6):3203–3212, 2016.

[163] Philip Kenny and Kevin J Flynn. Physiology limits commercially viable photoautotrophic production of microalgal biofuels. *Journal of applied phycology*, 29(6):2713–2727, 2017.

[164] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

[165] ED Frank, J Han, I Palou-Rivera, A Elgowainy, and MQ Wang. Life-cycle analysis of algal lipid fuels with the greet model. *Center for Transportation Research, Energy Systems Division, Argonne National Laboratory, Oak Ridge*, pages 11–5, 2011.

[166] Rajendra K Pachauri, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, and Purnamita Dasgupta. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, 2014.

[167] Jeffrey W Moody, Christopher M McGinty, and Jason C Quinn. Global evaluation of biofuel potential from microalgae. *Proceedings of the National Academy of Sciences*, 111(23):8691–8696, 2014.

[168] Howard Passell, Harnoor Dhaliwal, Marissa Reno, Ben Wu, Ami Ben Amotz, Etai Ivry, Marcus Gay, Tom Czartoski, Lise Laurin, and Nathan Ayer. Algae biodiesel life cycle assessment using current commercial data. *Journal of environmental management*, 129:103–111, 2013.

[169] Rebecca L White and Rebecca A Ryan. Long-term cultivation of algae in open-raceway ponds: lessons from the field. *Industrial Biotechnology*, 11(4):213–220, 2015.

[170] Andy Hector and Robert Bagchi. Biodiversity and ecosystem multifunctionality. *Nature*, 448(7150):188, 2007.

[171] Jonathan S Lefcheck, Jarrett EK Byrnes, Forest Isbell, Lars Gamfeldt, John N Griffin, Nico Eisenhauer, Marc JS Hensel, Andy Hector, Bradley J Cardinale, and J Emmett Duffy. Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nature communications*, 6:6936, 2015.

[172] Jason C Quinn and Ryan Davis. The potentials and challenges of algae based biofuels: a review of the techno-economic, life cycle, and resource assessment modeling. *Bioresource technology*, 184:444–452, 2015.

[173] Susanne B Jones, Yunhua Zhu, Daniel B Anderson, Richard T Hallen, Douglas C Elliott, Andrew J Schmidt, Karl O Albrecht, Todd R Hart, Mark G Butcher, and Corinne Drennan. Process design and economics for the conversion of algal biomass to hydrocarbons: whole algae hydrothermal liquefaction and upgrading. Report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2014.

[174] Lindsay Soh, Mahdokht Montazeri, Berat Z Haznedaroglu, Cuchulain Kelly, Jordan Peccia, Matthew J Eckelman, and Julie B Zimmerman. Evaluating microalgal integrated biorefinery schemes: empirical controlled growth studies and life cycle assessment. *Bioresource technology*, 151:19–27, 2014.

[175] Qingshi Tu, Matthew Eckelman, and Julie Zimmerman. Meta-analysis and harmonization of life cycle assessment studies for algae biofuels. *Environmental science & technology*, 51(17):9419–9432, 2017.

[176] Avigad Vonshak. *Spirulina platensis arthrospira: physiology, cell-biology and biotechnology*. CRC Press, 2014.

[177] Casey M Godwin, Aubrey R Lashaway, David C Hietala, Phillip E Savage, and Bradley J Cardinale. Biodiversity improves the ecological design of sustainable biofuel systems. *GCB Bioenergy*, 10(10):752–765, 2018.

[178] David Tilman, Johannes Knops, David Wedin, Peter Reich, Mark Ritchie, and Evan Siemann. The influence of functional diversity and composition on ecosystem processes. *Science*, 277(5330):1300–1302, 1997.

[179] Bradley J Cardinale. Biodiversity improves water quality through niche partitioning. *Nature*, 472(7341):86–89, 2011.

[180] Abdulla Malek, Luca C Zullo, and Prodromos Daoutidis. Modeling and dynamic optimization of microalgae cultivation in outdoor open ponds. *Industrial&Engineering Chemistry Research*, 55(12):3327–3337, 2015.

[181] Po-Cheng Lin, Fuzhong Zhang, and Himadri B Pakrasi. Enhanced production of sucrose in the fast-growing cyanobacterium synechococcus elongatus utex 2973. *Scientific Reports*, 10(1):1–8, 2020.

[182] Cinthia Núñez, Renato León, Josefina Guzmán, Guadalupe Espín, and Gloria Soberón-Chávez. Role of azotobacter vinelandii muca andmucc gene products in alginate production. *Journal of bacteriology*, 182(23):6550–6556, 2000.

[183] Carlos Peña, Silvia López, Andrés García, Guadalupe Espín, Angel Romo-Uribe, and Daniel Segura. Biosynthesis of poly-β-hydroxybutyrate (phb) with a high molecular mass by a mutant strain of azotobacter vinelandii (opn). *Annals of microbiology*, 64(1):39–47, 2014.

[184] D Segura, J Guzmán, and G Espín. Azotobacter vinelandii mutants that overproduce poly-β-hydroxybutyrate or alginate. *Applied microbiology and biotechnology*, 63(2):159–163, 2003.

[185] Daniel Segura, Tania Cruz, and Guadalupe Espín. Encystment and alkylresorcinol production by azotobacter vinelandii strains impaired in poly-β-hydroxybutyrate synthesis. *Archives of microbiology*, 179(6):437–443, 2003.

[186] Terry L Marker. Opportunities for biorenewables in oil refineries. Report, UOP LLC, 2005.

[187] Mark S Wigmosta, André M Coleman, Richard J Skaggs, Michael H Huesemann, and Leonard J Lane. National microalgae biofuel production potential and resource demand. *Water Resources Research*, 47(3), 2011.

[188] Roy W Harris, M John Cullinane Jr, and Paul T Sun. Process design and cost estimating algorithms for the computer assisted procedure for design and evaluation of wastewater treatment systems (capdet). Report, ARMY ENGINEER WATERWAYS EXPERIMENT STATION VICKSBURG MS, 1982.

[189] Nyomi Uduman, Ying Qi, Michael K Danquah, Gareth M Forde, and Andrew Hoadley. Dewatering of microalgal cultures: a major bottleneck to algae-based fuels. *Journal of renewable and sustainable energy*, 2(1):012701, 2010.

[190] Ramesh K Shah and Dusan P Sekulic. *Fundamentals of heat exchanger design*. John Wiley & Sons, 2003.