The physical, environmental, and evolutionary determinants of biological architecture

by

Christopher Andrew Poling Kempes

Submitted to the Department of Earth, Atmospheric and Planetary Sciences, Interdepartmental Program of Physical Biology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physical Biology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2013

Author .....
Department of Earth, Atmospheric and Planetary Sciences,
Interdepartmental Program of Physical Biology
December 13, 2012

Certified by .....
Michael J. Follows
Senior Research Scientist
Thesis Supervisor

Accepted by .....
Robert D. van der Hilst
Schlumberger Professor of Earth and Planetary Sciences
Head, Department of Earth, Atmospheric and Planetary Sciences

The physical, environmental, and evolutionary determinants of biological architecture

by

Christopher Andrew Poling Kempes

Submitted to the Department of Earth, Atmospheric and Planetary Sciences,
Interdepartmental Program of Physical Biology
on December 13, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physical Biology

Abstract

The relationship between structure and function is a longstanding and central topic in biology, evolution, and ecology. The importance of morphology is clearly visible in the diverse forms taken by innumerable organisms in order to perform a myriad of functions. Examining the great variety of morphological characteristics it would seem that the overall principle of evolution is the only way to generalize the observed diversity: given differences in environments and random biological variation a great multitude of body plans have been invented as adaptations to many dynamic habitats given specific evolutionary histories. In this thesis I will show how focusing on diverse organisms makes it possible to identify common first-order laws of evolutionary organization. More specifically I will show how these common laws derive from a connection between organism structure, physical limitations, environmental constraints, and basic metabolic, biochemical, or energetic principles. Furthermore, I will show how this top level of biological organization holds significant predictive power for regional ecology and for interpreting the general trends of evolutionary history.

In Chapter 2 we begin by deriving a model for the growth of single cells and populations of cells. This model is based on the partitioning of metabolic resources and the scaling relationship between metabolism and body size. We show that the growth of diverse classes of organisms is connected by common unit energetics. However there exist striking differences in the broad trends between growth rate and body size across these different classes and we show that this is a consequence of major evolutionary transitions which adjust the partitioning of metabolic resources. We interpret major evolutionary transitions to occur in response to energetic limitations.

We also find that multicellular living for unicellular organisms provides a metabolic and reproductive advantage. In Chapters 3 and 4 we further investigate these features in microbial biofilms which exhibit rich spatial patterning. Using a mathematical model and experimentation we find that the tall vertical structures produced by these biofilms have optimal geometry for resource uptake and the growth efficiency of the entire colony. Our model allows us to predict the observed changes in feature ge-

ometry given alterations to the environmental conditions that the biofilms are grown in. Furthermore, we are able to show that the morphology of these structures is dependent on single cell physiology. For example, single genetic knockouts of flagellar motility radically alter the temporal dynamics of feature spacing. Our work highlights morphology as a central property in multicellular organisms which mediates the interaction between environmental conditions and physiology.

In Chapter 5 we highlight the importance of morphology in complex multicellular life where we develop a general model of tree architecture which we link to physiological success within a given environment. Although this model is general, uses only tree size as a governing parameter, and does not consider speciation we are able use local resource availability to predict broad regional patterns in plant traits such as maximum tree height.

Each of these chapters highlights the importance of structure and morphology at multiple biological scales. In Chapter 6 we show how the importance of structure extends to the genetic level where the specific encoding of a gene can have implicit information and functionality beyond the basic translation of codons. We investigate the observed implicit function of dramatic and frequent changes in the mutation rate of an organism given the structure of the mutL gene. We show mathematically that altering mutation rates is an evolutionarily advantageous strategy, and we show bioinformatically that the specific genetic structure that gives rise to this trait is under positive evolutionary selection.

Thesis Supervisor: Michael J. Follows
Title: Senior Research Scientist

Acknowledgments

First, I am grateful to my committee for extending a broad range of intellectual freedom to carry out this interdepartmental degree program. I especially thank my adviser, Mick Follows, for innumerable stimulating conversations, many brilliant ideas, and for very balanced guidance. I would like to thank Richard Murray for hosting me at Caltech for over two years where I found his group and conversations enormously stimulating. The collaboration that I have engaged in with Lars Dietrich, which began at MIT, has been enormously beneficial. The work throughout this thesis was supported by a NSF Graduate Research Fellowship, the MIT Society of Presidential Fellows, and the Gordon and Betty Moore Foundation. While at MIT I had countless stimulating conversations with Andrew Barton, Alexa Price-Whelan, Eric Downes, Raffaele Ferrari, Malte Jansen, John Marshall, Matt Rigby, and Maureen Coleman amongst many others. I am grateful to my Caltech officemate Dionysios Barmpoutis for endless conversations on diverse topics which greatly enriched my academic work and stimulated many new ideas. I am grateful to Suzanne Kern for constant comment, useful critique, insight, and encouragement regarding my work.

For Chapter 2 we thank two anonymous reviewers whose constructive critiques and comments provided inspiration and led to significant improvements in this study. We thank Mario Feingold for supplying data from refs. [231, 232], Francisco Feijó Delgado and the Manalis group for supplying data from ref. [110], and Sallie Chisholm, Robert Olson, and David Soll for discussing the methods and data presented in refs. [201] and ref. [119] respectively. We thank Raffaele Ferrari, Suzanne Kern, and Maureen Coleman for stimulating discussions. We thank the online BioNumbers database which we found useful for finding references containing many relevant energetic calculations. This work was supported by the Gordon and Betty Moore Foundation (C.P.K., M.J.F. and S.D.), NASA (M.J.F.), and the NSF (M.J.F.).

The work in Chapters 3 and 4 was supported the Gordon and Betty Moore Foundation (C.P.K. and M.J.F.), NASA (M.J.F.), and the NSF (M.J.F.). We would like to thank Roman Stocker for early guidance during a term project for his course "Physical

Ecology at the Microscale".

For Chapter 5 I would like to thank B. Whitten for many stimulating discussions and academic support, S.E. Kern and M.J. Follows for suggestions and comments on earlier versions of the manuscript, and J. Kattge for contributing data from the TRY database.

We thank Martin Nowak for early guidance as Chapter 6 began in an evolutionary dynamics course taught at Harvard. The author's are very grateful to Paul D. Sniegowski for providing a valuable empirical perspective as well as helpful suggestions.

This doctoral thesis has been examined by the Committee of the Interdepartmental Program in Physical Biology as follows:

Sallie W. Chisholm
Lee and Geraldine Martin Professor of Environmental Studies, Department of Civil and Environmental Engineering, and Department of Biology, MIT

Richard M. Murray
Thomas E. and Doris Everhart Professor of Control and Dynamical Systems and Bioengineering, California Institute of Technology

Roman Stocker
Associate Professor of Civil and Environmental Engineering, MIT

Bruce Tidor
Professor of Biological Engineering and Computer Science, MIT

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Abstract

The relationship between structure and function is a longstanding and central topic in biology, evolution, and ecology. The importance of morphology is clearly visible in the diverse forms taken by innumerable organisms in order to perform a myriad of functions. Examining the great variety of morphological characteristics it would seem that the overall principle of evolution is the only way to generalize the observed diversity: given differences in environments and random biological variation a great multitude of body plans have been invented as adaptations to many dynamic habitats given specific evolutionary histories. In this thesis we show how focusing on diverse organisms makes it possible to identify common first-principle laws of evolutionary organization. More specifically we show how these common laws derive from a connection between organism structure, physical limitations, environmental constraints, and basic metabolic, biochemical, or energetic principles. Furthermore, we show how this top level of biological organization holds significant predictive power for regional ecology and for interpreting the general trends of evolutionary history.

Since the time of the greek Alcmaeon we have been concerned with the connection between biological form and function [240]. Aristotle was the first to consider a unity of body plan [240] articulated later by Goethe as "Nature seems to operate always according to an original and general plan, from which she departs with regret and whose traces we come across everywhere" (quoted by ref. [240]). More recently following the work of Wright 1932 we have come to view biological diversity as a stochastic random walk through a dynamic fitness landscape which is defined by a connection between the current environment and genetically determined organism traits [74, 302, 54]. Out of these perspectives comes an important question: is the amazing amount of observed biological diversity governed only by highly localized environments and unique evolutionary trajectories or are there also, in the language of physics, first-order or dominant constraints which are constant across all of life and which manifest as obvious commonalities between organisms?

It has been recognized that in addition to the laws of chemistry, which constrain internal physiological processes, the physical laws are important for regulating numerous processes such as diffusion via body size and structure. Physical laws impose hard limits and constraints on evolutionary possibilities [249]. For example it has been noted that many classes of organisms share a fundamental similarity in terms of common biochemical pathways (with some variation) and thus the major differences between species are those of the specific chemical interaction networks along with body structure and size [45]. The scale and structure of organisms alter which physical processes are important and the degree to which these laws have an effect on physiology [45]. Thus there has been a great emphasis historically and recently on using body size as a central property governing organism physiology and function [249, 45, 43]. Much of this work has had great success in describing the broad trends of physiology and form across a vast diversity of organisms [43]. Examples include a power-law relationship between body size and metabolic or growth rates in numerous classes of organisms, the connection between average plant size and the density of trees within forests, and the scaling between body size and genome length in prokaryotes [43, 82, 83, 80, 69].

It is quite challenging to connect the fundamental laws of physics with the highly complex and diverse world of biology. Schrödinger stated that in biology we would likely find "other laws" of physics [250, 54]. There is much debate over the meaning of this statement [54], but I interpret it as follows: biology, although entirely constrained by the fundamental laws of physics, combines those laws in such a complex fashion that the simplest descriptions are novel macroscopic "laws" derived from reasonable and physically informed principles. This is similar to work in non-linear dynamics where features such as chaos are derived from a set of assumptions which cannot necessarily be traced back to something as fundamental as quantum mechanics but which provide a foundation for mechanistically describing behavior. Thus in biology we must look for new descriptions, we must identify the core principles from which to derive our theory. For example, much recent work has shown that whole organism metabolic rate can largely be derived from considerations of network architecture and the challenges of resource distribution [292, 291, 14, 16, 246]. The rules of network architecture and axioms of graph theory then become a fundamental constraint in biology. Perhaps these principles are also alluded to in Goethes concept of a "general plan".

In Dobzhanskys classic work "Genetics and the Origin of Species", which is argued to have inspired the modern evolutionary synthesis (e.g. [54]), he proposes that evolutionary biology must connect the causal perspective, which he considers to be genetic and chemical, with the "historical" perspective which he defines as the comparative study of organismal morphology [74]. Similarly, today I think the challenge is to connect morphology to the environment on one end and to the underlying biochemistry, chemical energetics, and genetics at the other, and to do this while still maintaining a general perspective on the broad constraints which organize biology.

In this thesis we use several case studies to investigate the connection between morphology and either environmental success or fundamental energetics and genetics.

In Chapter 2 we examine the connection between fundamental energetics, metabolism, and the growth and reproductive success of microorganisms. Here we show how focusing on size and fundamental biophysical energetics leads to a remarkably powerful

and general model which can unify the growth of single cells with the energetic pressures for major evolutionary transitions. We show that body size is important only up to a point: shifts in the general whole organism structure dramatically alters their internal metabolic processes along with the basic limitations they face. The fundamental feature of growth and internal metabolic energetics is related to the packaging of metabolic function. For example the transition from prokaryotes to eukaryotes is most striking in terms of the dramatic shift in internal organization (e.g. the evolution of mitochondria), and the evolution of multicellular eukaryotes involves a further evolution of hierarchical structure with dramatic changes in the possibilities for growth and metabolism [69, 131]. We find that the major evolutionary transitions in the history of life are accompanied by shifts in both energetics and overall body organization (e.g. the addition of mitochondria). Because metabolism and internal structure are closely linked, the energetic limitations faced by organisms drives evolutionary changes in body structure to overcome these limitations.

In our work on microbial energetics we found that the evolutionary transition to multicellular life provides significant energetic and reproductive advantages to organisms. We show that evolving more complex biological structure is advantageous at given biological scales but this does not explain the process by which these transitions are able to occur. For complex structure to evolve each incremental step must have some immediate benefit. What are these intermediate benefits and how does something as complex as multicellularity evolve? In chapter 3 we investigate microbial biofilms as rudimentary multicellular organisms. We find that basic cooperation within these communities leads to a benefit for the entire population while also producing complex spatial morphology. Most striking is that the overall success of the community is strongly connected with this morphology. The biofilm patterning is the product of complicated collective cellular behavior and we are able to determine that the resulting morphology of the biofilm is controlled in such a way that the interaction with the external environment is optimized for resource acquisition and the growth of the colony. Furthermore, some of this morphology is not just the simple emergent pattern from the feedback between resource availability and growth. Morphology is

the product of both cellular physiology and response to the environmental situation. The behavior has likely been selected for such that the emergent patterns resulting from the feedback with the environment are optimal for the colony. In a sense this is a rudimentary model for complex multicellular organisms, and provides a plausible step along the trajectory towards complex multicellular life.

The biofilm itself may represent an evolutionary transition in which cells went from planktonic and competitive living to the formation of structured communities with at least minor cooperation. Understanding the mechanisms behind this transition require us to determine how various cellular physiologies and responses combine to produce emergent patterns and how these might be selected for in time to modify community structure. In Chapter 4 we investigate the direct relationship between physiology and biofilm morphology. Here we examine how an array of single genetic knockouts leads to a spectrum of spatial patterning. We show that the basic temporal dynamics of biofilm ridge spacing can be explained by the scales of diffusion and resource supply. Our analysis of several genetic knockouts, particularly one that decreases cellular motility, shows that physiology can radically alter the temporal dynamics or patterning. These genetic knockouts demonstrate that a wide variety of cellular processes combine to control patterning and form a complex "program" of sorts for emergent morphology. In order to better uncover the key components of this "program" we use graph theoretic metrics to analyze the entire spatial pattern and to find key differences between genetic knockouts. We find that connectivity and topology can vary significantly between genetic mutants and that colonies can be partially categorized based on their similarity to a either a regular lattice or a graph with random connections. Comparing the colony networks to an observed distribution network (the power grid) we show that the colonies are much more similar to a lattice. The regular spacing of features within a lattice suggests a dominant spatial scale and this supports our finding that the processes of diffusion and resource management are setting the specifics of spatial patterning. This also suggests that the emergent networks are likely not serving any direct function, such as resource distribution, beyond increasing the uptake of oxygen as described in Chapter 3. This

work provides the foundations for a model system which can address long-standing questions regarding how much of morphology is controlled by genetics opposed to emergent dynamics.

For biofilms the organization of individual cells into morphologically rich communities can confer significant advantages and it is likely that these communities eventually evolved into complex multicellular life. Once a community of cells is encapsulated into a single complex organism the overall morphological structure then becomes a central aspect of its evolutionary trajectory. This is because morphology continues to be the key mediator between the environment and overall fitness. In Chapter 5 we show this explicitly by developing a model which links the general architecture of trees to metabolic success within a given environment. We show that body size, generalized architecture, and energy budgets can be simply combined to interpret regional environmental limitations and predict key plant traits such as maximum allowable tree height. Our model is relatively simple yet predicts features with complicated dependencies highlighting that size and structure coupled with physical processes are the dominant factors governing physiology.

Throughout this thesis we show how structure is important at various scales from the internal architecture of microbes (Chapter 2) to the overall morphology of communities and complex multicellular organisms. A concept that has been proposed in the metabolic theory of ecology is a hierarchy of structures and processes which regulate function at a continuum of physical scales [295, 69]. As highlighted by Delong et al. [69] metabolism at different scales is likely controlled by the constraints of different architectures: network driven resource distribution for complex multicellular life opposed to the overall genome size for prokaryotes. Much of this thesis is focused on intermediate or large-scale morphologies which modulate metabolism via resource acquisition or supply. It is interesting to consider the constraints of structure at the most fundamental biological scales. In Chapter 6 we turn our attention to considerations of the importance of structure at the scale of genome architecture. We examine how the structure of the encoding for a gene can have embedded within it implicit information beyond the basic translation of codons [52]. We investigate a

specific implicit function which directly controls the rate of evolution and fitness of an organism. Specifically, a repeated codon module in a gene controlling mutation rate (mutL) leads to a situation where during genome replication the gene can easily gain or lose a copy of the module. This process leads to a situation where the gene is intermittently disabled and the corresponding mutation rate of the organism switches between one of two rates separated by a factor of roughly 10^2. We show that this mechanism is evolutionarily advantageous as it balances the ability for an organism to search for novel physiologies with increased fitness (high mutation rate) while also maintaining a stable population with the current fitness (low mutation rate). This mechanism is conserved across many strains and bacterial species and we show that it is likely under positive evolutionary selection.

In general, we show in this thesis that the continuing evolution of biological structure can be described by considering the dominant physical processes which connect structure to function. We find that biological diversity is the product of both the dominant physical laws and evolutionary diversification: for a given class of organisms there exists a set of dominant physical constraints which lead to fundamental commonalities, however these constraints shift across different biological scales and environments and, as such, evolutionary innovations lead to diverse groups of organisms with distinct shared commonalities.

Chapter 2

Growth, metabolic partitioning, and the size of microorganisms

Preface

In this chapter we show how connecting body size to basic principles of energetics allows us to derive a general description of the growth of single cells and populations and anticipate, and interpret the implications of, major evolutionary life-history transitions. Our work highlights how simple physical principles can be coupled with structure to predict a wide variety of biological properties from the scale of single cells to evolutionary occurrences across vastly different biological scales.

This chapter is a reproduction of the article "Growth, metabolic partitioning, and the size of microorganisms" published by C.P. Kempes, S. Dutkiewicz, and M.J. Follows in the Proceedings of the National Academy of Sciences [131] where I am the corresponding author.

For this work I conceived the theoretical framework, compiled all of the data, and carried out the analysis and writing. Mick Follows and Stephanie Dutkiewicz contributed by providing guidance, discussion, and editing portions of the published text.

Abstract

Population growth rate is a fundamental ecological and evolutionary characteristic of living organisms but individuals must balance the metabolism devoted to biosynthesis and reproduction against the maintenance of existing structure, and other functionality. Here we present a mathematical model which relates metabolic partitioning to the form of growth. The model captures the observed growth trajectory of single cells and individuals for a variety of species and taxa spanning prokaryotes, unicellular eukaryotes, and small multicellular eukaryotes. We find significant variability in the relationship between metabolism and body size both within a species and between species of the same taxon. In contrast, our analysis suggests that the per unit costs of biosynthesis and maintenance are conserved across prokaryotes and eukaryotes. However, the relative metabolic expenditure on growth and maintenance of whole organisms clearly differentiates taxa: prokaryotes spend an increasing fraction of their entire metabolism on growth with increasing cell size, while eukaryotes devote a diminishing fraction. These differences allow us to predict the minimum and maximum size for each taxonomic group, anticipating observed evolutionary life-history transitions. The framework provides energetic insights into taxonomic tradeoffs related to growth and metabolism and constrains traits which are important for size-structured modeling of microbial communities and their ecological and biogoechemical effects.

2.1 Introduction

Understanding the fundamental principles which underpin the rates of growth and reproduction of organisms is of central ecological importance, ultimately affecting long term evolutionary trajectories of populations and communities. The growth of an individual, including single cells, depends on multiple metabolic, biochemical, and physiological processes (e.g. [171, 282, 256, 279, 299, 229]). Microbes exhibit a diversity of biochemical and metabolic strategies, making it difficult to evaluate and synthesize the associated fitness tradeoffs between species. How do different organisms allocate and manage internal metabolic resources in order to govern the complicated process of reproduction? Here we combine basic cellular bioenergetics with macro-ecological perspectives to produce a framework which is useful for examining and interpreting the growth trajectory of a single cell, and also provides insight into major ecological patterns in population growth rates. It helps us to interpret major evolutionary life-history transitions between micro-organisms from a bioenergetic perspective. The model is general and can be applied across a broad spectrum of

species spanning two of the three domains (Eukarya and Bacteria) and four kingdoms (Bacteria, Fungi, Plantae, and Animalia).

2.2 Metabolic Perspectives

In general the growth and metabolism of organisms have been investigated at multiple levels of organization. At one end of this spectrum, experimental techniques are expanding our ability to track single cells through a division cycle revealing the growth trajectories of individual organisms even at the microbial scale (Figure 2-1) [110, 231, 232, 201, 119, 281]. At the population scale, single-species culture studies have been used to understand the relationship between resource use and growth rate [211, 92, 245, 254, 145, 33]. At the largest scale ecologists have looked across many species and taxa of organisms to characterize general power law relationships between body size and growth rate, metabolism or other traits (see for example [193, 140, 44]). Here we present a model which combines these three perspectives in order to understand the form for growth of single cells and populations of individuals for diverse organisms.

2.2.1 Community metabolism and energetics

We begin at the population scale where numerous continuous culture studies have been used to characterize microbial energetics. Typically, they reveal a linear relationship between biomass weighted resource consumption and growth (dilution) rate [211, 92, 245, 254, 145, 33]. The Pirt model [211] interprets this relationship in terms of a partitioning between growth and maintenance:

$$Q = \frac{\mu}{Y} + P \tag{2.1}$$

where Q is a consumption rate per unit mass of a limiting resource (mol resource $\cdot$ s^{-1} $\cdot$ g cells $^{-1}$), μ is the specific growth rate (s^{-1}), Y is a yield coefficient (g cells $\cdot$ mol resource $^{-1}$), and P is a maintenance term (mol resource $\cdot$ s^{-1} $\cdot$ g cells $^{-1}$) (for a list of symbols see Table A.5). Maintenance metabolism is defined as the consumption rate

at zero growth, or the minimal requirement for survival. The total consumption rate, Q, can be measured by oxygen use [92, 245, 254], light absorption [145], prey ingestion [33], or any other resource consumption rate which is assumed to be proportional to the total metabolic rate of the organism.

This perspective of cellular energetics has broad ranging applications, from interpreting the metabolic drivers of food or biofuel production efficiency, to understanding the energetics of human pathogens [48]. Here we will use the Pirt model to phenomenologically motivate a relationship between single cell growth and metabolic allocation, drawing analogies with population studies.

2.2.2 Interspecific Metabolism

In a complementary view power law relationships between body size and numerous organisms traits, including total metabolism, have been identified for organisms ranging from microbes to mammals [193, 295, 140]. Recent theoretical work suggests mechanisms which underpin these scalings for multicellular organisms (e.g. [44, 140, 15]). Total metabolic rate, B (W), is typically measured over several orders of magnitude and is expressed as

$$B = B_0 m^{\alpha}, \tag{2.2}$$

where B_0 (W $g^{-\alpha}$) is a parameter reflecting the size normalized metabolism and accounts for metabolic differences between organisms that are not related to size, including temperature dependence (see Appendix A.2). The exponent α describes how quickly total metabolism changes with body mass. There has been much debate over the value of the exponent of these scalings and its interpretation (e.g. [297, 179]). Recent work shows that different broad groups of organisms follow characteristically different power relationships [69, 179] with α ranging from 1.96 for prokaryotes to .79 for metazoans [69].

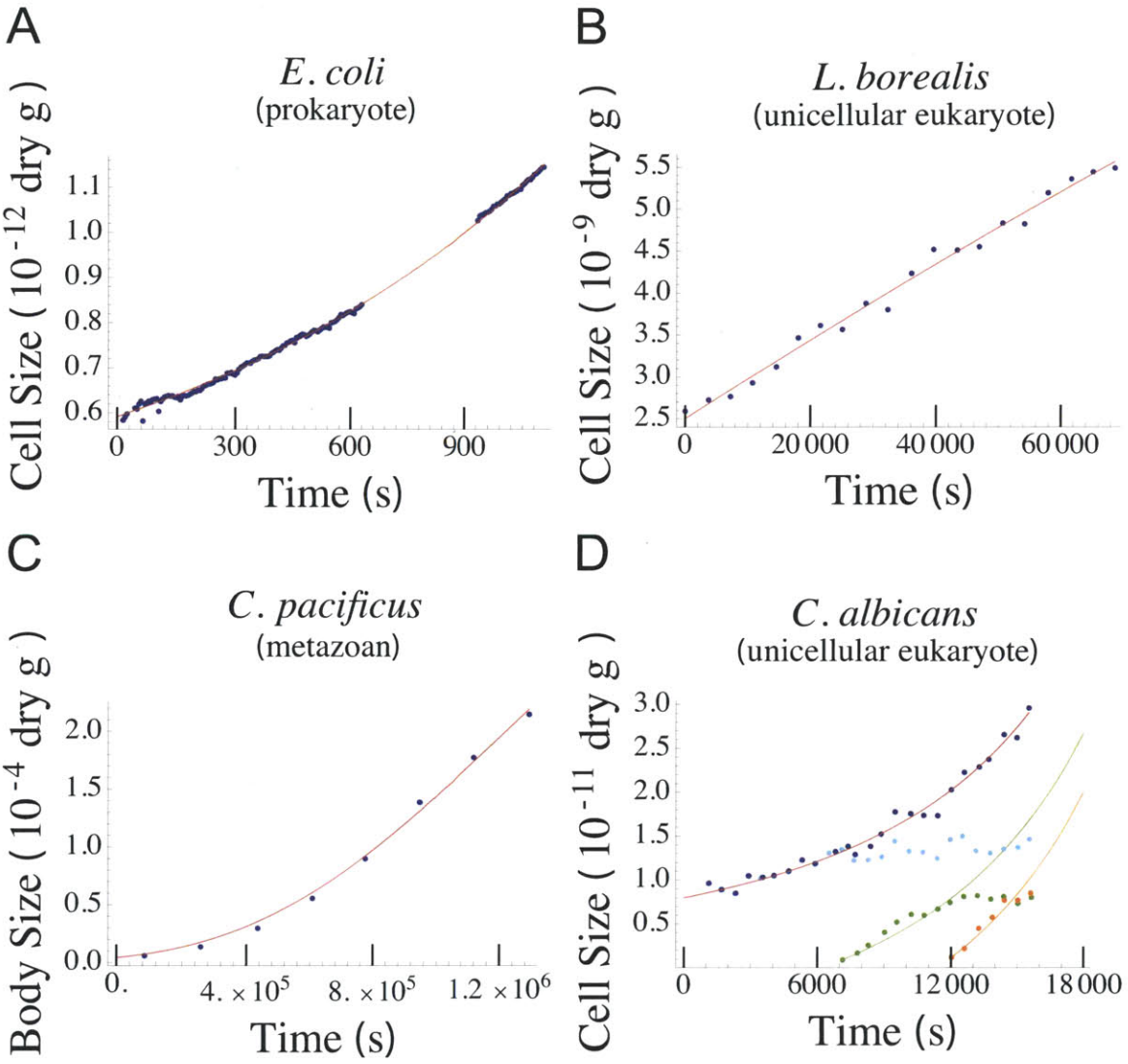


Figure 2-1: Growth curves for (A) *E. coli* data from refs. [231, 232] (B) *L. borealis* from ref. [201] (C) *C. pacificus* from ref. [281] (D) *C. albicans* from ref. [119]. Dots indicate the cell mass as a function of time from the experiments in refs. [231, 232, 201, 281, 119]. The red lines are the best fit of Eq. 2.8. The values of α, γ_0 and b that provide this best fit are given in Table A.4. For *C. albicans* (D) the red curve tracks the fitted growth of the total budding complex of yeast (blue points), while the green and orange points and curves represent the growth of individual daughter buds. The orange and green curves are predictions assuming that all growth energy from the entire complex is devoted to a bud as it forms (see Appendix A.7 for details). The cyan points represent the growth of the mother cell which is identical to the total complex until the formation of daughter buds. Data and model fits for additional individual cells are provided in the Appendix (Figures A-4 through A-10).

2.2.3 Growth trajectories of single cells and individuals

Models of cellular division have traditionally focused on the interconnected fluxes of available resources into the cell and between different internal physiological processes or partitions (e.g. [256, 299, 229]) or on the kinetics of regulatory proteins (e.g. [279]). We consider a growth model for unicellular organisms that incorporates the basic energetic partitioning of the Pirt model along with metabolic scaling principles. The mass specific consumption rate (Eq. 3.1) from the Pirt model of an entire population of cells can be related to the metabolic rate of an average single cell by

$$\bar{B} = Q\bar{m}N = \frac{N}{Y}\frac{d\bar{m}}{dt} + PN\bar{m} \tag{2.3}$$

where N is a conversion between resource use and energy production (J mol resource $^{-1}$). $\bar{m}$ and $\bar{B}$ are respectively the population average, per-cell mass and metabolic rate. Eq. 2.3 implies that maintenance costs scale linearly with mass and that the unit cost of biomass production is constant (consistent with refs. [293, 179, 106, 121]). Let us also assume that the total metabolism of a single cell follows this relationship and is divided between the growth of new biomass and the maintenance of existing biomass leading to the following single-cell model of metabolic partitioning

$$B = E_m\frac{dm}{dt} + B_m m. \tag{2.4}$$

Here B_m (W g^{-1}) is the metabolic expenditure to support an existing unit of mass, E_m (J g^{-1}) is the metabolic energy needed to synthesize a new unit of biomass, m (g) is the current mass of the organism, and dm/dt (g s^{-1}) is the growth rate of an individual. When B follows a power law with mass (Eq. 2.2) and is combined with Eq. 2.4, this set of assumptions is termed the Ontogenetic Growth Model (e.g.[293, 179, 106, 121]) which has previously been applied to multicellular organisms assuming a fixed value of α. Here we will treat α as a free parameter which might differ between species.

Normalizing by B, this single cell budget can be expressed as the fraction of total

metabolism being used for growth and maintenance

$$1 = \frac{E_m}{B}\frac{dm}{dt} + \frac{B_m}{B}m \tag{2.5}$$

Substituting B from Eq. 2.2, we can define the time dependent fraction of metabolism devoted to growth and maintenance respectively as

$$\gamma = \frac{E_m}{B_0 m^\alpha}\frac{dm}{dt} \tag{2.6}$$

and

$$\rho = \frac{B_m}{B_0}m^{1-\alpha}. \tag{2.7}$$

Rearranging for dm/dt and integrating, we can describe the growth trajectory of an individual cell, with initial mass m_0, by

$$m(t) = \left[1 - \gamma_0 e^{-b(1-\alpha)t}\right]^{1/(1-\alpha)} \left(\frac{1}{1-\gamma_0}\right)^{1/(1-\alpha)} m_0, \tag{2.8}$$

(e.g. [24]) where $b = B_m/E_m$ (s^{-1}) is the ratio of the maintenance metabolic rate to biosynthetic cost, and $\gamma_0 \equiv \gamma(t=0) = 1 - \frac{B_m}{B_0}m_0^{1-\alpha}$ is the fraction of metabolism devoted to growth for a cell at this initial mass (see Appendix A.3 for details). Thus, given an initial mass, the growth trajectory $m(t)$ depends on only three parameters γ_0, b, and α. These parameters simply combine the fundamental biological quantities of E_m, B_m, B_0 and α.

This solution for $m(t)$ suggests a variety of possible forms for the growth trajectory of an individual, as shown schematically in Figure 2-2 and for individual growth data in Figure 2-1. The bioenergetic parameters α, b, and γ_0 exert a strong influence on the growth of an individual to division. A shorter generation time (faster division rate) is the result of increasing the metabolic scaling exponent α, increasing the initial percentage of metabolism devoted to growth γ_0, or increasing b, the ratio of the unit maintenance metabolism to the unit biosynthetic cost (illustrated in Figure 2-2 A and B). Increasing α yields a greater total metabolism and increasing the growth

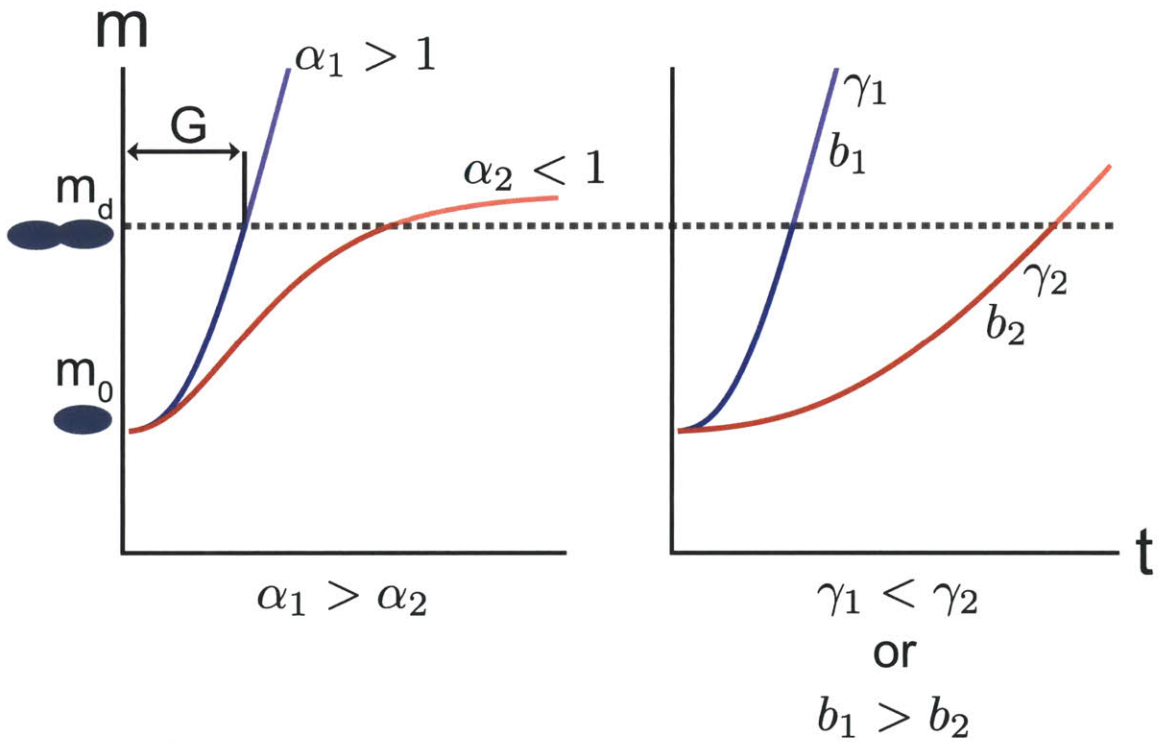


Figure 2-2: Schematic of the growth trajectories from our model. The growth trajectories predicted by Eq. 2.8; the mass of an individual is plotted against time. Division occurs when the mass of the organism crosses the division mass, m_d. The relative contribution of the three metabolic parameters α, b, and γ_0 to division time are illustrated. In each panel species 1 (blue) is compared to a species (red) which is identical except for a change to one of the three parameters. Organisms will divide more quickly if the metabolic scaling exponent, α, or the ratio of the unit maintenance metabolism to unit biosynthetic cost, b, are increased or if the initial percentage maintenance, γ_0 is decreased. For $\alpha < 1$ organisms grow towards an asymptotic mass, while for $\alpha > 1$ organisms do not.

fraction γ_0 implies a greater fraction of metabolism devoted to biosynthesis. Similarly, increasing b alone requires decreasing the biosynthetic cost E_m (increasing B_m will simultaneously decrease γ_0) and this implies more efficient biosynthesis.

Previously, growth trajectories of single cells have been considered to follow either an exponential or linear relationship (e.g. [110, 231, 232]). The model described in Eq. 2.8 indicates additional forms for growth where an exponential is a special case of Eq. 2.4 with $\alpha = 1$. For $\alpha > 1$ the growth rate of an individual, and the fraction of metabolism devoted to biosynthesis, continues to increase with mass until division. For $\alpha < 1$ an organism will grow towards an asymptotic mass (Figure 2-2 A) with growth rate initially increasing for young cells but decreasing as they approach division, and with the biosynthesis fraction decreasing over the entire life cycle.

2.2.4 Growth of populations

A common measure of fitness in an ecosystem is the population growth rate, μ, of a species. The single cell (or single individual) model of Eq. 2.8 can be evaluated to provide the generation time G [293], the time for one organism to reach reproductive mass, $M_d \equiv \epsilon m_0$,

$$G = \frac{1}{b\left(1-\alpha\right)} \ln\left[\frac{\gamma_0}{1-\epsilon^{(1-\alpha)}\left(1-\gamma_0\right)}\right] \tag{2.9}$$

where for simple fission $\epsilon \approx 2$. Following ref. [247] the population growth rate, μ is given by this generation time along with the average fecundity, f, and percentage of the population to reach the age of reproduction, L:

$$\mu = \ln\left(Lf\right)/G. \tag{2.10}$$

Considering simple fission one could assume that $\mu_{max} \approx \ln\left(2\right)/G$.

2.3 Bioenergetics of species and taxa

Here we fit the framework outlined above to observed growth trajectories of single cells (or single individuals) and to data relating population growth rate to body size.

We compare the underlying bioenergetics of different species, as represented by the parameters of our framework, using these fits along with separate estimates obtained from growth-resource consumption data for populations of cells.

We compiled published measurements of the size of single cells as they grow through a division cycle for five species including the heterotrophic bacteria *E. coli* [231, 232] and *B. subtilis* [110], two photo-autotrophic marine diatoms (*Thalassiosira weissflogii* and *Lauderia borealis*) [201], as well as budding yeast *C. albicans* [119]. In addition, we used growth data for individuals from two small (sub-millimeter), multicellular marine copepods (*Calanus pacificus* and *Pseudocalanus sp.*) [281]. Fitting model Eq. 2.8 to the time-changing size of each individual, we infer the values for b, γ_0 and α (Table A.4). We used a least squares analogy and the Nelder-Mead simplex method (please see Appendix A.9 for details and a discussion on curve fitting).

We found that Eq. 2.8 was able to accurately describe the growth trajectories of unicellular and small multicellular individuals from two of the three domains (Eukarya and Bacteria) and four kingdoms (Bacteria, Fungi, Plantae, and Animalia) (Figure 2-1, A-4 through A-10).

We were able to estimate a single interspecific value of α and the bioenergetic parameters by fitting Eq. 2.10 to compiled measurements [69] of population growth rate and body size. We also compiled continuous culture experiments in order to estimate the bioenergetics of a single species at the population scale. Combining Eqs. 3.1, 2.3, and 2.4, the population average value for b can be evaluated from continuous culture data as $\bar{b} = PY$, summarizing all of the information of a standard "Pirt plot". Similarly, the average fraction of metabolism devoted to growth is given by

$$\bar{\gamma} = \frac{1}{1 + \frac{PY}{\mu}} = \frac{1}{1 + \frac{\bar{b}}{\mu}}. \tag{2.11}$$

Here the fraction of metabolism devoted to growth has a simple meaning in terms of the dimensionless parameter $b/\mu = \frac{B_m}{E_m \mu}$, representing the ratio of maintenance and biosynthesis rates per unit mass.

2.3.1 Variation in α the metabolic exponent

Metabolic considerations of ecology often emphasize the scaling exponent α, with macro-ecological compilations revealing relatively constant values for α over large ranges of body size. Here, our evaluations of α from individuals reveal considerable variation at both the species and single cell level (Figure 2-3 C, Table A.4). For individual *E. coli* cells the exponent varies between 0.57 and ≈ 3, though all have have a similar form to their growth trajectory (Figure 2-1 A and A-4) due to the short timescale for division.

The cell-to-cell variation in the exponent could be the result of a variety of physiological effects at the individual scale including natural phenotypic variability, fine-scale differences in experimental conditions, or the prior history of each cell. As well as variation within a species we also find that the average exponent for a species often deviates significantly from the interspecific value for each of the three major taxonomic groups. For example, for the prokaryote *B. subtilis* $\alpha = 0.98 \pm .30$, significantly different than the interspecific value of $1.96 \pm .18$ based on a range of prokaryotes found in ref. [69]. We thus illustrate that the general constraints which are hypothesized to organize allometric scalings at the taxonomic level can be substantially violated at the single species level. This is perhaps least surprising for prokaryotic organisms (see Discussion) but it should be noted that deviations of this magnitude may be significant when attempting to apply uniform scalings and parameterizations in ecological models.

We also obtain an interspecific estimate of the metabolic exponent, α, by fitting our population model (Eq. 2.10) to the data for growth rate against body size for many species (see Appendix A.6). We find that that the fitted values of α agree with the compiled data [69] for metabolic rates (we find 1.66 for prokaryotes and and .80 for eukaryotes). The fitted curves accurately capture the interspecific form for growth (Figure 2-3 A) which is normally fit by a power law (also drawn in Figure 2-3 A). The model fitted here is more complicated than a simple power law but it brings to bear additional information in the form of fundamental bioenergetic constraints (i.e.

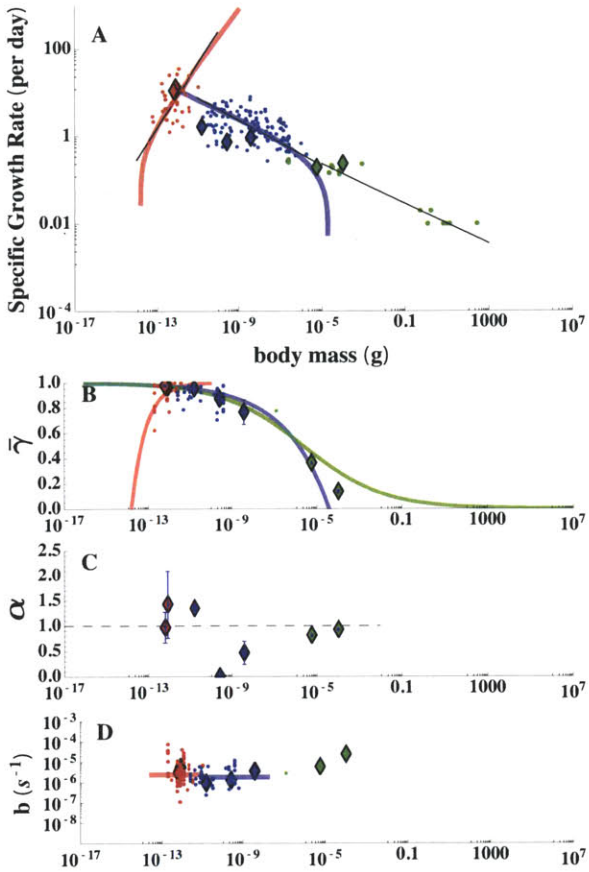


Figure 2-3: (A) The interspecific plot of specific growth rate (day^{-1}) against size along with the metabolic constants (B-D) inferred from fits of our model to individual growth trajectories all plotted against organism mass. The prokaryotes are colored red, the eukaryotes blue, and the small metazoans green. In each panel diamonds represent the results for single individuals while the points are estimates from compiled population studies (ref. [69] for A, and our own compilation for C and D). In (A) the colored curves represent the best fit of the framework to interspecific growth using Eqs. 2.9 and 2.10 while the black curves are the best fit power law relationships [69]. The asymptotes illustrate the size limitations of prokaryotes and unicellular eukaryotes. (B) The average percentage of metabolism devoted to growth, $\bar{\gamma}$, illustrates the differences in metabolic partition across the three major taxonomic groups (for prokaryotes there is a dense clustering of points near 1 that is not visible). The colored curves are predictions based on fits from (A). (C) the metabolic scaling exponent, α, shows variation at the species and taxonomic levels where the dashed line represents $\alpha = 1$. (D) The maintenance to biosynthesis cost ratio, b, is on average constant across species and major evolutionary transitions and the colored curves represent the mean value for each taxon. The error bars represent the standard deviation from the mean of each parameter.

unit biosynthetic costs), discussed in detail below. The model approximates a simple power-law over much of the size-range of the organisms under examination.

2.3.2 Common energetic costs

From the fits to individual cells we find that b, the ratio of the unit maintenance metabolism to biosynthetic cost, ranges between $\approx 10^{-6}$ and $\approx 10^{-5}$ s^{-1} but exhibits no trend with body size (Figure 2-3 D, Table A.4) nor any clear distinction between the three major evolutionary life-history groups (prokaryotes, single-cell eukaryotes, metazoans). Compiled population-based estimates, $\bar{b}$, covering a variety of prokaryotic and eukaryotic species also show no trend with size or taxonomic grouping (Figure 2-3 D) but have a greater variance (more than an order of magnitude) owing to the range of culture conditions and growth on different substrates (Fig. A-1, Table A.2). Though noisy, the relative constancy of b across cell size and between taxa suggests a fundamental bioenergetic commonality (see Discussion).

2.3.3 Variations in metabolic partitioning

Given Eq. 2.7 the metabolic partitioning should follow a relationship which depends on the size of an organism, the unit bioenergetic costs (represented by b), and the value of α. Since b is on average constant this implies that we should see changes in the partitioning across taxa based on changes in the metabolic scaling exponent α. Using Eq. 2.11 and the interspecific fit for μ (Figure 2-3 A) it is possible to predict the interspecific dependence of $\bar{\gamma}$ on mass for each species and taxa. In Figure 2-3 B we provide the predicted curves for the metabolic partitioning of prokaryotes, unicellular eukaryotes, and small multicellular organisms. For prokaryotes the percentage of metabolism devoted to growth is constantly increasing with increasing body size while for eukaryotes it is constantly decreasing. This is associated with the change in the interspecific value of α which is larger than one for prokaryotes and less than one for eukaryotes. The fits to single cell growth trajectories provide an estimate for the average metabolic partitioning, $\bar{\gamma}$. These species-specific points generally

follow the predicted curve (Figure 2-3 B). For population-based estimates we used the compilation of continuous culture data to estimate $\bar{\gamma}$ for each species; these also fall generally along the predicted curve.

This analysis reveals that large prokaryotes and small eukaryotes devote the highest fraction of total metabolism to growth and there is a size range where the two classes have similar metabolic partitioning. Here we infer a tradeoff between devoting resources to maintenance purposes and dividing quickly which is most easily seen by normalizing each growth trajectory (see Appendix A.4). Figure A-2 illustrates that as organisms live over longer normalized time-scales the fraction of metabolism devoted to growth decreases and this corresponds to moving across the three major life-history transitions.

2.3.4 Bioenergetic limitations and differing reproductive strategies

Given a set of bioenergetic parameters, the model presented here predicts some general limitations and constraints facing different classes of organism. First we note that a viable organism must allocate metabolic resources to growth at its initial size, that is, $\gamma_0 > 0$. This implies a limitation on the initial mass dependent on the metabolic scaling: for $\alpha > 1$,

$$m_0 > \left(\frac{B_m}{B_0}\right)^{1/(\alpha-1)} \tag{2.12}$$

while for $\alpha < 1$

$$m_0 < \left(\frac{B_m}{B_0}\right)^{1/(\alpha-1)}. \tag{2.13}$$

This gives a lower bound on size if $\alpha > 1$ and an upper bound on size if $\alpha < 1$. These limitations appear as the asymptotes in Figure 2-3 A for our fitted population growth equation. For prokaryotes this model predicts a lower bound on size of 1.59×10^{-14} g, which is close to the mass of the smallest observed prokaryotes, of the genus Mycoplasma, which are between $1.2-4\times10^{-14}$ g (the smaller estimated from ref. [180] and the larger given by [69]). For unicellular eukaryotes we predict an upper bound

on size at 3.89×10^{-5} which corresponds to the general scale at which there is a major evolutionary life-history transition to multicellularity. Thus the model presented here can anticipate evolutionary shifts based on underlying energetic considerations.

As prokaryotes grow larger, and as eukaryotes grow smaller, they move away from the limit where all energy is devoted to maintenance. In doing so the fractional allocation of metabolism devoted to growth approaches unity for large prokaryotes and small eukaryotes (Figure 2-3 B) and this imposes a different limitation: at some point the increasing biosynthetic rates cannot be sustained by the fundamental processes of the cell (see Discussion). It should be noted that there is a size range where both eukaryotes and prokaryotes devote nearly all of their resources to biosynthesis, and it is within this range that the evolutionary transition between the two occurs. The transition from prokaryotes to eukaryotes shifts the metabolic exponent such that the unit biosynthesis rates begin to decrease with increasing body size allowing eukaryotes to grow larger.

In summary, the minimum size for prokaryotes is bounded by the limit where all energy is devoted to maintenance while the maximum size is bounded by the demand for unsustainable biosynthesis rates. The transition to unicellular eukaryotes is accompanied by shifts in physiology and metabolic scaling which allow them to devote an increasing fraction of total metabolism to maintenance with increasing size. Hence their lower size limit is bounded by unachievable biosynthesis rates, while the largest unicellular eukaryotes face a limit where they cannot allocate any metabolic energy to biosynthesis, setting the stage for the transition to multicellularity. Here there is not a dramatic shift in the metabolic scaling, rather the more advanced body plans of multicellular organisms afford more complicated reproductive strategies such as internal gestation and individuals hatched from an egg, resulting in various alterations to the generation time as represented by Eq. 2.9.

The importance of reproductive strategy in the eukaryotes is illustrated by the budding yeast *C. albicans* which reproduces in a manner that differs from the standard fission of unicellular eukaryotes. For a complex of budding yeast, as new buds form other parts of the complex reach an asymptotic size (Figure 2-1 D) which would

suggest that $\alpha < 1$. Yet fits of the model (Eq. 2.8) to the entire complex yield $\alpha = 1.14 \pm 0.07$. The asymptotic size reached by each bud is not the result of a metabolic limit but is rather due to the entire complex sharing metabolic resources. The growth of the buds can be predicted by assuming that the bud is using all of the growth energy of the entire complex (see Appendix A.7). Buds grow more rapidly due to the assistance of the entire complex than if growing in isolation. We also find that the resting phases of the autotrophic diatoms (e.g. *T. weissflogii*) alter the apparent energetics found within our framework (see Appendix A.8) because biosynthetic activities are not linked to changes in size [171]. Such reproductive strategies explain how the upper bounds proposed by our model can be overcome and how they could lead to the transition to multicellular life because of the corresponding growth advantage.

2.4 Discussion

Population growth rate is a fundamental characteristic, critical in our understanding and interpretation of cellular processes, from basic physiology to the ecology of microbial communities. Simplified models of cellular growth and metabolism are widely applied in laboratory and ecological contexts. Macro-ecological descriptions generally capture broad trends across many species and highlight major patterns in biology but often lack the ability to interpret differences between species or describe detailed cellular processes. At the other end of the spectrum, explicit models of cell division attempt to capture the interconnected dynamics of internal biochemistry, external resources, and/or physiological regulation (e.g. [256, 279]). Recent advances in genomics enable detailed metabolic reconstructions which can predict the growth dynamics of a cell (e.g. [77, 229]). Here we have developed and applied a relatively simple framework which combines macro-ecological and energetic perspectives. It provides insight into single cell growth dynamics and is also useful for considering interspecific and taxonomic trends in population growth. We have used the model to show how organisms in different taxonomic groups are subject to different constraints

on metabolic expenditure at the individual scale, allowing the model to anticipate and interpret three major evolutionary life history transitions from a metabolic partitioning perspective.

Until recently, macro-ecological studies have found that the scaling exponent for metabolism with body size, α, is approximately constant (the often used 2/3 or 3/4 power law) over many orders of magnitude in body size [44, 297]. A recent empirical analysis has revealed that power law relationships exist for micro-organisms with different values of α based on major taxonomic groupings[69].

In this study it was critical to allow α to be a free parameter in order to accurately interpret the bioenergetics of individual microbial species (see Appendix A.9.1). We find that α not only varies between taxa [69] but also between cells of an individual species and between species of the same taxonomic group. The averages over many species and individuals (i.e. the interspecific values) of α reveal the broad physiological organization of different taxa. Variation around these values illustrates that individual species can modify their physiology away from the general constraints facing a class of organisms.

In contrast to the taxonomic variations in α, we find that the average value of b (the ratio of unit maintenance metabolism to biosynthesis cost) is unchanging across major evolutionary transitions and has no significant trend with body mass within or between taxa. This suggests that the basic, per unit mass costs of biosynthesis and maintenance are common across broad taxonomic divides; these energetic costs are likely related to highly conserved, common metabolic machinery and function [86, 69]. This is consistent with the success of thermodynamic interpretations of the yield, Y, which is a predictable function of the substrate in heterotrophic microbes [284, 118]. Yet b also requires an understanding of the maintenance metabolism P whose fundamental meaning still requires future investigation.

The framework developed here is broad and flexible and can capture the interspecific relationship between growth rate and body mass in micro-organisms using the interspecific value of α for each taxon and a single average value of b. Since b is, on average, a constant value the interspecific trend in metabolic partitioning (rep-

resented by γ) is also dictated by body size and α (e.g. Eq. 2.11). By compiling laboratory data within this framework, we have a vehicle to interpret and hypothesize the taxon specific energetic limitations to population growth and its relationship to body size (Figure 2-3 A, Eq. 2.9). We can anticipate the minimum and maximum size of prokaryotes and eukaryotes, and suggest the body size where shifts between evolutionary life histories occur.

Prokaryotes are bounded by a minimum size below which they are unable to conduct biosynthesis since all metabolism must be devoted to maintenance. As they become larger, growth rate continually increases up to a maximum size, where higher biosynthesis rates are not achievable.

The smallest known prokaryotes are similar in size to the predicted lower bound and also possess the smallest observed genomes, which are assumed to be close to the minimum required to live alone in culture [180]. Our framework therefore suggests that the minimal genome is connected to the limit where biosynthesis is just viable and should consist of the required maintenance genome plus the smallest set of additional genes needed to conduct essential biosynthesis. We hypothesize that larger prokaryotic genomes allocate an increasing fraction of genes to biosynthesis corresponding to an increase in the fraction of metabolism devoted to growth. This is consistent with the argument that the interspecific scaling of prokaryotic metabolism is due to a relationship between body size, genome length, the overall metabolic complexity of prokaryotes and their ability to encode for a diverse set of enzymes [69]. Yet genome length does not change over the life cycle of an individual cell and thus it is unclear what sets the scaling of metabolism at the single cell scale. The potential link between genome length and the species specific value of the metabolic exponent could be directly tested by pairing numerous existing prokaryote sequences with a larger number of future estimates for α obtained from single cell growth trajectories or population studies which track average cell size.

Many other factors may contribute to metabolic scaling at the single species level. New techniques to track the transcriptional composition of a population have revealed that prokaryotic cells can maximize growth rate by adjusting the partitioning of

transcriptional resources between producing basic components for protein synthesis (amino acids) or building more ribosomes to construct them [308]. Faster growing cells are observed to allocate a greater proportion of promotor activity to ribosomes than to metabolism [308]. Our finding that population growth rate increases as the fraction of total metabolic resources devoted by individual to growth increases is consistent with this observation (Eqs. 2.6 and 2.11). Further experiments could be designed to seek correlations between transcriptional partitioning or the number of ribosomes and variations in metabolic scaling, bioenergetic parameters, and metabolic partitioning of particular prokaryotic species. Such a link would be important in underpinning the fundamental drivers of metabolic scaling and the key differences between species and taxa.

We hypothesize that the maximum size for prokaryotes is determined by a point where the overall biosynthesis rates cannot be met by basic cellular processes. For example prokaryotic genome length increases with body size following a power law [69], while the generation time decreases with body size according to Eq. 2.9, and this predicts that at some point the time to replicate the genome will be slower than the generation time (e.g. the maximum nucleotide copying rate is insufficient). Another possibility is that since biosynthesis rates scale more quickly than the overall volume of a bacterium and at some point the number of ribosomes required for protein synthesis cannot fit within the cell.

The framework developed here suggests that eukaryotes, in contrast to prokaryotes, exhibit a decrease in metabolism and population growth rate with body size, also consistent with [69]. Smaller eukaryotes allocate all of their metabolism to growth (similar to the larger prokaryotes) and cannot be any smaller and still achieve the required rate of biosynthesis (the genetic and synthesis arguments for prokaryotes moving towards larger size should also apply for eukaryotes moving toward smaller sizes). As cell-size increases, growth rate decreases and expenditure on maintenance is enhanced. Eventually, the single-celled eukaryotes reach a size at which they are devoting all metabolism to maintenance and, with no energy for biosynthesis, they are no longer viable. Beyond this upper size limit, other strategies come into play,

including multi-cellularity.

It appears that the transition between prokaryotes and eukaryotes is accompanied by a fundamental shift in strategies for metabolic partitioning, consistent with significant differences in their basic physiology. For example, the average number of proteins made per mRNA is an order of magnitude higher in eukaryotes than in prokaryotes [160] implying that eukaryotes are able to produce far more proteins for the same amount of transcriptional resources. Eukaryotes are also able to expend an order of magnitude greater metabolic power on each gene, regardless of function, compared with prokaryotes [143] which may alleviate genome length or transcription related constraints. The increased metabolic power for eukaryotes has been hypothesized to be the result of the presence of mitochondria, which create a greater internal surface area for ATP production [143]. Similarly, the metabolic exponent for unicellular eukaryotes is hypothesized to be the result of the linear scaling between mitochondrial volume and overall cell volume [69]. These hypotheses could potentially be tested by correlating the value of α estimated from single cell or continuous culture growth data with concurrent measurements of the total mitochondrial volume in a cell.

Our framework suggests that the bioenergetic limitations of binary fission faced by the simplest unicellular eukaryotes may be overcome by the development of more complicated reproductive strategies. This is consistent with the transition to more elaborate body plans and the increasing presence of internal distribution networks which are argued to underpin the metabolic scaling (i.e. value of α) observed across a vast range of multicellular body sizes [69]. Although the basic metabolic framework outlined here can be expanded to account for altered reproductive strategies (e.g budding yeast) certain reproductive strategies, such as resting or dormant phases, highlight the limitations of this framework when trying to estimate metabolic scalings from the growth trajectories of individuals alone (e.g. *T. weissflogii*). To better constrain α and to test hypotheses concerning the role of multicellularity and other strategies, simultaneous experimental determinations of the time evolution of body size and resource consumption could provide critical information.

2.5 Conclusion

We have developed a framework to characterize the growth and division of individual cells. We have used it to interpret the metabolic scaling, bulk bioenergetics, and metabolic partitioning of diverse species using the growth trajectory of individuals and laboratory populations. We find that the relative per unit metabolic costs of biosynthesis and maintenance are the same for prokaryotes, unicellular eukaryotes, and small metazoans, cutting across major evolutionary life history transitions. In contrast, prokaryotes and eukaryotes have fundamentally different strategies and limitations in the allocation of their metabolic resources which coincide with the observed evolutionary transitions. The allocation of all metabolic resources to maintenance purposes limits the size of the smallest prokaryotes and largest unicellular eukaryotes, while an inability to meet the ever increasing biosynthesis rates limits the largest prokaryotes and smallest unicellular eukaryotes. Metabolic constraints for larger eukaryotes are relieved by alternative reproductive strategies and multicellularity.

The metabolic theory has traditionally focused on average values and general principles across many species, here we illustrate strong variation of the metabolic size-scaling exponent at the individual species level. This framework provides avenues for connecting metabolic scaling to cellular characterizations such as genome length, or the partitioning of transcriptional resources, which may help to further elucidate the fundamental factors driving the evolutionary shifts discussed in this paper. This framework connects ecological fitness with basic bioenergetics and it may provide a mechanistic strategy for incorporating taxonomic differences into size-based approaches for modeling microbial systems (e.g. [12, 155]) in order to better understand complex microbial communities and their role in biogeochemical cycles.

Chapter 3

Biofilm features: environmental response and optimization

Preface

In Chapter 2 we saw how major evolutionary transitions are accompanied by energetic implications, including advantages which allow organisms to evolve greater size and complexity. We found that single-celled organisms living in a communal context gain significant energetic advantages by sharing metabolic resources. We also found that complex multicellular life allows organisms to escape the limitation faced by single-celled eukaryotes where all resources must be devoted to maintenance purposes. Nevertheless, there is still much to understand in terms of the processes by which complex multicellular life evolved, along with what specific metabolic advantages are conferred by this transition. In this next chapter we study biofilms which exhibit complex spatial patterning and which represent the possible beginnings of complex multicellularity. We find that aspects of spatial patterning can largely be explained by the metabolic dynamics within colony features and the effect that geometry has on resource acquisition.

This Chapter is in the form of a mature draft of a manuscript "Biofilm features: environmental response and optimization" by C.P. Kempes, C. Okegbe, Z. Mears-Clarke, M.J. Follows, and L.Dietrich. I will be the co-corresponding author for this

paper.

The work on biofilms stems from a close collaboration between theoretical developments and modeling, which I developed, and experimentation carried out by LarsDietrich, Chinweike Okegbe, and Zwoisy Mears-Clarke. Our collaborative relationship has evolved into the continual feedback between experiments which inspire theory and modeling predictions which suggest certain experiments or specific data analyses. For example the video measurements of the colony show interesting temporal dynamics which inspired me to create a computer interface for manually extracting the position of colony features. For this work I have derived many theoretical models and numerical simulations and conducted a variety of novel data extractions and analyses.

Lars Dietrich and I have had many long discussions regarding the various processes occurring within the biofilm system and the implications for complex multicellular evolution. Many of the ideas in this thesis stem from those conversations.

Mick Follows provided discussion and feedback on the theoretical, modeling, and ecological aspects of the work. Some of the early development of this work (along with Chapter 4) occurred during a term project that I conducted for Roman Stocker's course on microscale physical ecology. This course guided many of our concepts of cellular response and diffusive processes.

Abstract

A longstanding goal in biology is to understand the relationship between form and function and to determine how multicellularity morphologies evolved and conferred metabolic and fitness advantages to organisms. Here we show that bacterial biofilms produce structures which maximize cellular growth and the potential reproductive success of the colony. We develop a mathematical model of resource availability and metabolic response which accurately predicts the measured oxygen distribution within colony features. Using this model we demonstrate that the observed geometry of the colony is optimal with respect to whole colony growth rate. Since our model is based on resource dynamics we are also able to anticipate shifts in feature geometry based on external oxygen availability.

3.1 Introduction

The growth of cells as part of a community is a common occurrence in natural systems ranging from bacterial biofilms in the lungs of cystic fibrosis patients [301] to cyanobacterial colonies and mats in marshes, lakes, and oceans [264, 124]. Even complex multicellular life, such as plants and animals, is in essence a diverse community of cells living within a highly structured and regulated environment where survival is based on selection of the whole community and its properties. Communal living at multiple scales has been shown to provide various advantages. From a resource perspective recent work has shown that simple cooperation within aqueous microbial biofilms allow groups of genetically similar cells to grow tall mushroom-like structures which reach beyond local depletion zones into areas of fresh resources (e.g. [304, 305]). In a metabolic context other work has shown that basic colonial growth (e.g. budding yeast) and the evolution of complex multicellular life is accompanied by enhanced growth efficiency and several energetic advantages [131, 143, 295]. A central question in each of these contexts is once community organization is established how does the structure of and processes within that community evolve in time?

A key aspect of multicellularity is the organizational structure of and relationship between a community of cells. For example, in mammals the metabolic rate of individual cells living *in vivo* is regulated by the size of the whole organism which confers greater efficiency, while those same cells will have dramatically elevated

metabolic rates living as individuals in culture [295]. This regulation is considered to be the natural consequence of resource supply within hierarchical vascular networks [295, 292, 291, 14, 16, 246] highlighting the importance of morphology and structure in dictating cellular behavior for multicellular systems.

For complex multicellular life variations in morphological structure provide a wealth of information about speciation and environment given the inherent tradeoffs related to network structure. For the leaves of plants it has been shown that the variation of venation patterns across species is fundamentally related to tradeoffs associated with carbon and transpiration rates, water supply rates, and overall mass investment [28]. The variation in network structure is important because of the way in which resources are supplied to individual cells. In general the gradients of resources, particularly oxygen, are similarly important in modulating the dynamic development of embryos, lungs, and tumors (e.g. [258, 23]).

Thus one of the dominant effects of multicellularity is to alter the environment experienced by the individual cell. For example, the formation of gradients of resources is an obstacle to any multicellular organism. This points to broader biological questions, namely what is the interconnected relationship between morphology, metabolism, and environmental conditions and how do these features govern the fundamental benefits and tradeoffs of multicellularity? Understanding these connections may also help us to understand the general trajectory by which simple and complex multicellular life evolved and allow us to better identify, quantify, and interpret potential paleobiological structures such as those discussed in Albani et. al [4].

Here we use microbial biofilms which exhibit spatial patterning [73, 223, 228] and environmental sensitivity as a model system for exploring morphology as a metabolic adaptation. In our biofilm system we can modify the environmental and biological conditions and examine the resulting shifts in structure and metabolism. This allows for a deeper understanding of how morphology organizes metabolism and could be selected for and optimized.

The majority of mechanistic studies addressing biofilm development are performed in a flow-cell system. Bacteria are grown on a glass slide and exposed to a continuous

flow of nutrient-rich water. In this aqueous regime bacteria often forms mushroom-like structures in response to external nutrient gradients (e.g. [68, 304, 209].) The Gram-negative bacterium *Pseudomonas aeuroginosa* forms structurally diverse biofilms in different environments. We have studied *P. aeruginosa* biofilms by spotting 10 μl of a cell suspension onto an air-exposed agar plate and following the development of the colony over multiple days. This assay has proven fruitful for mechanistic studies, as genetic manipulation or changes in the environment result in morphological changes of the colony that can be easily observed by eye. We recently showed that a class of redox-active signaling molecules, so called phenazines, have dramatic effects on community behavior. While a wild-type colony remains smooth for the first two days of growth, a phenazine-null mutant (Δphz) undergoes a severe morphotypic transformation [73]. At day 2 Δphz begins to spread over the surface of the agar plate and forms wrinkles in the center of colony and spokes that radiate towards the edge of the colony (both features are vertical structures; see Figure 3-1). While a wild type colony also forms wrinkles in the center after three days of growth, a phenazine-over producer remains smooth when monitored for six days, highlighting an inverse correlation between an increased surface of the colony and phenazine production [73].

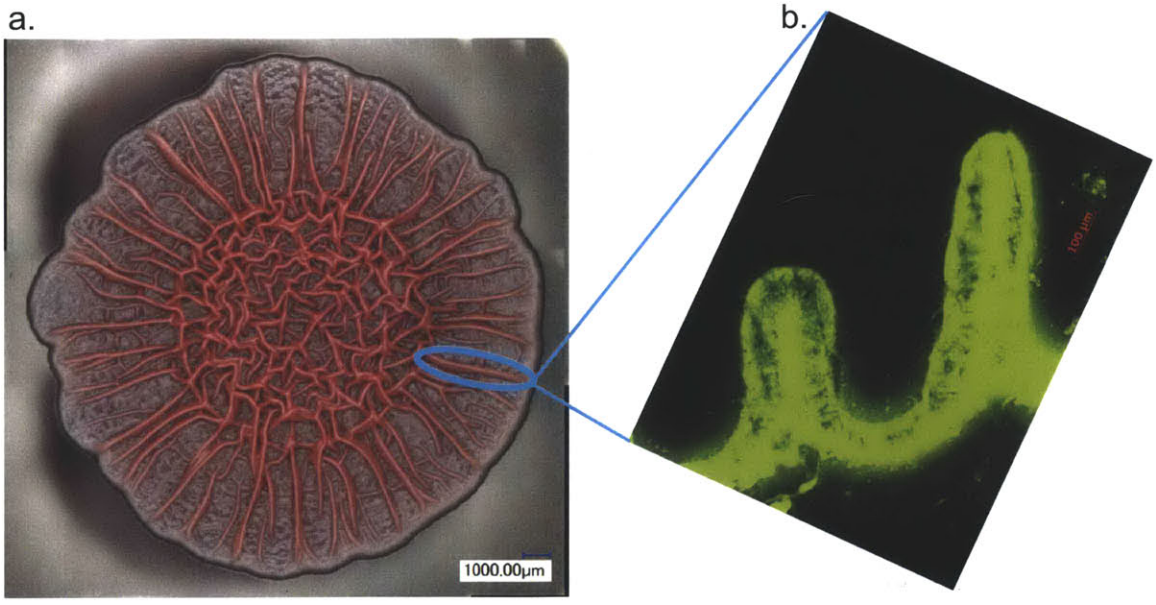


Figure 3-1: The horizontal and vertical structure of the *P. aeruginosa* biofilm. The colony is living on agar and exposed to the open atmosphere.

Here we demonstrate that the morphological change is a strategy of the community to increase oxygen uptake into the colony. Phenazines attenuate this mechanism, presumably because these redox-active compounds function as electron shuttles that allow cells in the anoxic parts of to balance their internal redox status by reducing phenazines. We propose that reduced cellular redox status triggers a morphological change on the community level. In favor of this model, addition of the electron acceptor nitrate to the agar, or increasing oxygen concentrations (to 40%) prevents spreading of the Δphz colony and discourages wrinkle and spoke formation. Decrease of external oxygen (to 15%) increase wrinkle and spoke formation as well as spreading of the colony.

Oxygen profiling of the colonies revealed that the Δphz mutant is well oxygenated within wrinkles and spokes, suggesting that its morphotype is an adaptation for optimal oxygen uptake. The sensitivity of colony morphology to oxygen availability and/or the oxidative mediation of phenazines suggests a significant role for oxidative capacity - specifically oxygen supply for Δphz types - in modulating morphology. Here we develop a simple model, rooted in established physiological and physical laws, which demonstrates and confirms this significance. Our work pairs quantitative experiments with our model allowing us to concretely verify our methodology and to make predictions and interpretations that would be infeasible using any single aspect. We demonstrate that the geometry of the features represent an optimal structure in terms of maximizing cellular growth within the biofilm.

3.2 Modeling biofilm metabolism and oxygen dynamics

To investigate the processes within the biofilm we develop a predictive mathematical model which is based on the simplest hypothesis for feature regulation: the dynamic feedback between metabolism, resource availability, and physical diffusion. We rely on some of the the most well-established, reproducible, and general cellular physiology:

1.) the Pirt model of community metabolism [211] which is applicable to a broad range of prokaryotic and eukaryotic species (see for example [211, 92, 245, 254, 145, 33] and ref. [118] for a review of multiple unicellular species); 2.) the Monod model for the dependence of population growth rate on a limiting resource [172] which has been widely applied.

For organisms living in a continuous culture it has been demonstrated that many species follow a linear relationship between their total metabolism and growth rate [211, 92, 245, 254, 145, 33]. The classic Pirt model [211] describes this relationship in terms of a metabolic partitioning between maintenance and growth:

$$Q = \frac{\mu}{Y} + P \tag{3.1}$$

where Q is a consumption rate per unit mass of a limiting resource which provides energy and/or structural material to the community of cells (mol resource $\cdot$ s^{-1} $\cdot$ g cells $^{-1}$), μ is the specific growth rate (s^{-1}), Y is a yield coefficient (g cells $\cdot$ mol resource $^{-1}$), and P is a maintenance term (mol resource $\cdot$ s^{-1} $\cdot$ g cells $^{-1}$). Maintenance metabolism is defined as the consumption rate at zero growth, or the minimal requirement for survival.

The metabolic activity of a population is dependent on the availability of oxygen. We add this process to equation 3.1 by considering the dependence of growth rate on the local oxygen concentration. For *P. aeruginosa* and many other species growth is observed to saturate at increasing concentrations of many potentially limiting substrates including oxygen (e.g. [3]). This sensitivity is often described using Monod kinetics [172]:

$$\mu = \mu_{max} \frac{[O_2]}{k_s + [O_2]} \tag{3.2}$$

where μ_{max} is the maximum growth rate approached as the substrate is taken to infinity, and k_s is the half-saturation constant. Substituting equation 3.2 into 3.1 gives oxygen consumption as a saturating function of available oxygen:

$$Q = \frac{\mu_{max}}{Y} \frac{[O_2]}{k_s + [O_2]} + P. \tag{3.3}$$

Depletion of oxygen within the biofilm creates a gradient with the atmosphere and drives a diffusive flux of oxygen into the colony. Thus the time dependent spatial concentration of oxygen can be described by

$$\frac{\partial [O_2]}{\partial t} = D\nabla^2 [O_2] - \left(\frac{\mu_{max}}{Y} \frac{[O_2]}{k_s + [O_2]} + P \right) a \tag{3.4}$$

where a is the density of cells in the colony (g cells $\cdot$ m^{-3}). At steady state, the diffusive supply of oxygen is maintained by the gradient in oxygen concentration, and must balance the consumption associated with growth at each point within the colony. It should also be noted that at zero oxygen concentration the cells either die or become dormant and thus we take the maintenance term P to be zero for low oxygen concentration. Equation 3.4 can be conveniently nondimensionalized:

$$\frac{\partial [O_2]^\star}{\partial t^\star} = \nabla^2 [O_2]^\star - \left(\frac{[O_2]^\star}{1 + [O_2]^\star} + g \right). \tag{3.5}$$

where temporal and the spatial scales have respectively been normalized by the factors

$$t_{fac} = \frac{a\mu_{max}}{k_s Y}, \qquad x_{fac} = \left(\frac{a\mu_{max}}{k_s Y D} \right)^{1/2}, \tag{3.6}$$

oxygen concentrations have have been divided by k_s, and the nondimensional maintenance term is $g = YP/\mu_{max}$. For steady state solutions our model only relies on the concentration of oxygen at the colony surface, which is determined by the atmospheric mixing ratio and pressure, and the two parameters x_{fac} and k_s. We determined the value for both parameters by first compiling published measurements and then constructing the mean of every combination within this compilation (see Appendix B.1).

3.3 Observations of internal oxygen distribution

In order to investigate how these mechanisms regulate morphology we study a single feature or "ridge" (Figures 3-1 and 3-2) and the surrounding geometry.

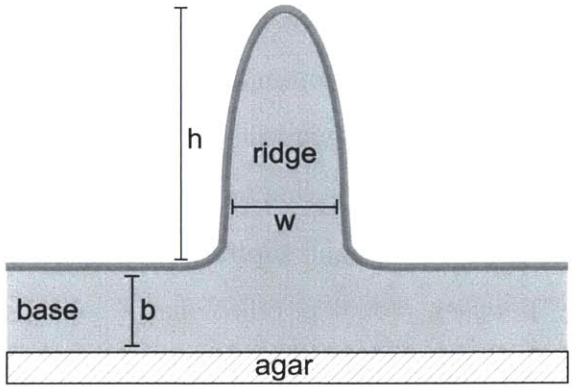


Figure 3-2: Schematic of the geometry of ridge and the surrounding area of the colony. The height of the "base" is represented by b, the "ridge" width is given by w, and the height of the ridge relative to the base is given by h.

The fundamental measurements that we make are oxygen profiles of the base (b in Figure 3-2) of the colony. Here the geometry is simple and fairly uniform. We can use the observed oxygen profiles to calibrate our model and examine our literature-based estimates for the parameters of the model.

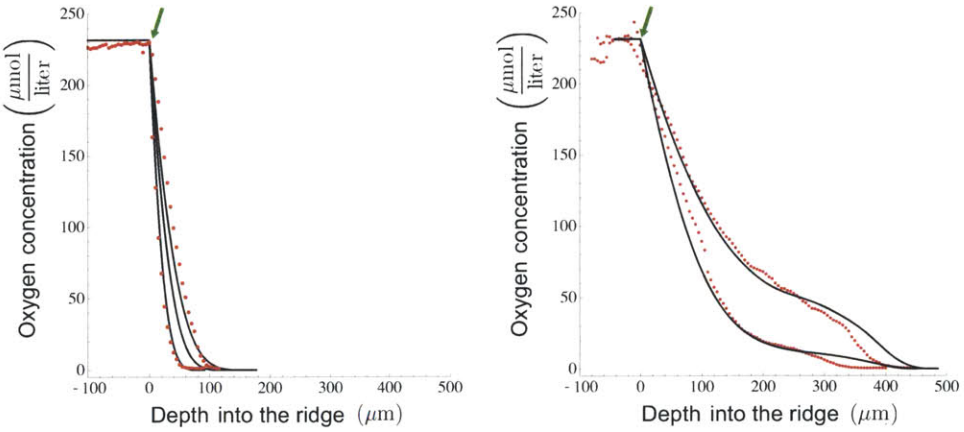
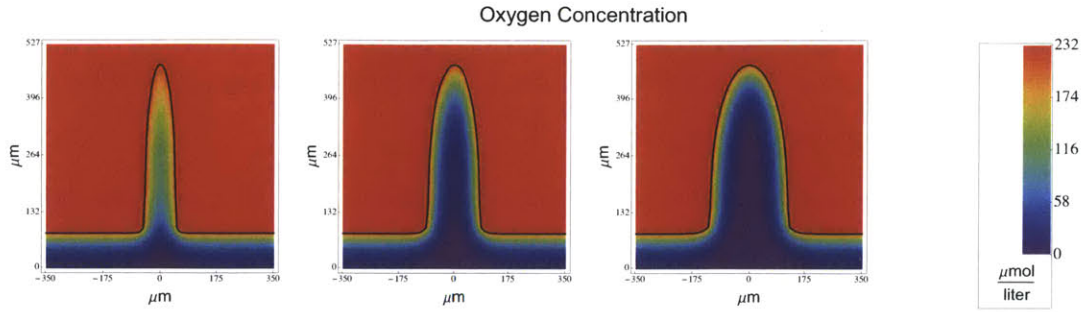


Figure 3-3: Depth profiles of oxygen within the biofilm. (a) the base region and (b) the ridge. Measured values are given by points and simulation are the solid curves. We took many depth profile measurements and the two sets in each plot represent the variation within the observed profiles.

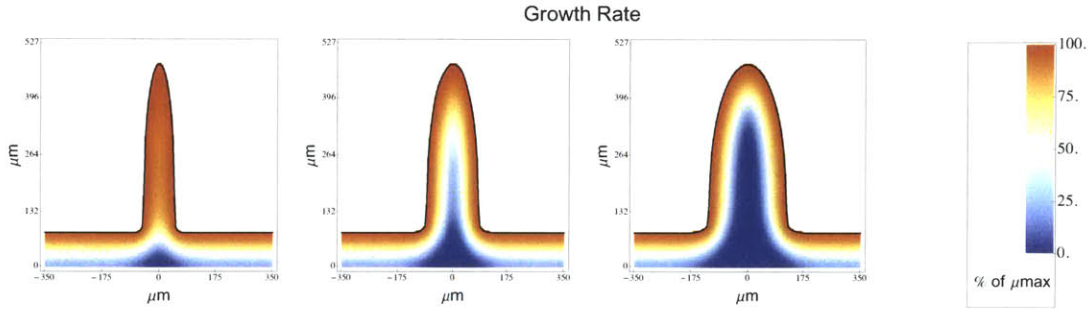
Plots of oxygen concentration against depth into the colony exhibit a decay of oxygen availability with increasing depth (Figure 3-3). There is some variation in the decay rate and the depth at which oxygen can no longer be detected. Most of this

observed variation can be accounted for by rescaling the depth axis by a simple factor (see Appendix B.1). This highlights common overall biological behavior consistent with the model but with variation in parameter values between colonies. Given our definition of x_{fac} this variation could be due to differences in physiological parameters such as maximum growth rate, though empirical evidence suggests that this is not likely to differ greatly between clonal populations of cells (see refs. [231, 232] for an example of growth rate variation), in which case variations in the oxygen decay rate would be due to differences in the physical parameters of the colony. The colony matrix may exhibit local and regional differences in the specific structure of the matrix related to density, wetness, and overall diffusivity of oxygen. Either way we fit the observed variation in vertical one dimensional profiles by varying x_{fac} and solving for the steady state oxygen concentration in our model for an isolated base geometry while keeping k_s fixed. Figure 3-3(b) shows that steady state solutions of our model in the base captures the range of observed oxygen profiles. This range of profiles also allows us to calibrate the range of x_{fac} to between 5.1×10^4 and 1.0×10^5. We also show a simulated oxygen profile which represents the median between the two profiles for a parameter value of 6.7×10^4 which compares well with our estimate from the literature of 6.8×10^4 (see Appendix B.1).

Experimentally, we find that oxygen penetrates much deeper in the ridge compared to the base (Figure 3-3). This is likely the result of increased surface area for a reduced cellular mass. To interpret this phenomena we use the calibrated model to simulate the steady-state distribution of oxygen availability within an entire region of the colony, including the ridge, and we do this for various geometries (Figure 3-4(a)). Our model also captures the general decay rate and shape of these ridge profiles using the mean value of the calibration and finding the best width of each ridge (it is experimentally challenging to measure the width and oxygen profile simultaneously) (Figure 3-3(b)). Our ability to anticipate the general shape and range of ridge profiles gives us confidence in our calibration and the model's ability to make predictions. Once these calibrations have been made all other results in this paper follow directly as predictions.



(a)



(b)

Figure 3-4: The modeled internal distribution of oxygen and growth rate within the colony. (a) The modeled internal availability of oxygen along with (b) the spatial distribution of growth rate within a biofilm for features of different size.

3.4 The impact of colony and feature geometry

The increased depth of oxygen penetration in the ridge region (Figure 3-3(b)) illustrates the importance of geometry in enhancing oxygen availability within the colony. The question then becomes how changes in this geometry affects the oxygen availability within the colony. Or why the geometry happens to be the one that is observed. To explore these questions we simulate the simplest geometric variation where we preserve the overall form of the ridge but vary the width of the ridge for a fixed height. Figure 3-4 shows the distribution of oxygen within ridges of different widths. From the internal oxygen distribution and equation 3.2 we can calculate the local growth rate and we find that ridges generally have an enhanced growth rate relative to the flatter regions (Figure 3-4(b)). This is due to the fact that thin tall features increase the total surface area for a relatively small amount of biomass which enhances the oxygen flux per total consumption as can be seen in the enhanced oxygen penetration.

However, we also see in Figure 3-4 that as ridges grow wider the effect of enhanced growth is diminished and the layer of growing cells within the ridge region becomes very similar to that of the base regions.

This effect implies a tradeoff for community metabolism: for thin features oxygen is able to penetrate the entire feature and cells grow incredibly quickly, but there are not many cells growing due to the small size of the feature; for thick features there are many more cells but the region of enhanced oxygen penetration and growth is diminished. This tradeoff can be summarized as a return on mass investment for the colony. We quantify this as the added growth per added mass given the the changing ridge width: $E = dG_{tot}/dM_{tot}$ where G_{tot} is the total growth of the ridge and M_{tot} is the mass of the region of interest.

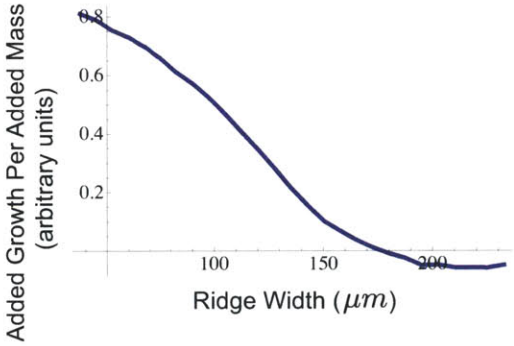


Figure 3-5: The relationship between the width of a ridge and the simulated efficiency, E, the added growth per added mass of the colony. This plot illustrates that there is a diminishing return on investment as widths grow larger. At a given width the total growth of the colony is maximized and any further increasing in ridge width yields a negative return on mass investment. This can be interpreted as the width which optimizes the reproductive success of the colony.

We calculate the efficiency, E, for an ensemble of simulated features each with different widths but fixed feature height (Figure 3-5). We find that E goes negative for ridges wider than a specific value. This critical width represents the optimal for efficient offspring production within the colony. More importantly the optimal width found in our model accurately predicts the observed average width of colony features along with the observed variation in widths (Figure 3-6). This prediction is made by solving for the optimal width using the minimum, maximum and average parameter

values from our calibration to base profiles.

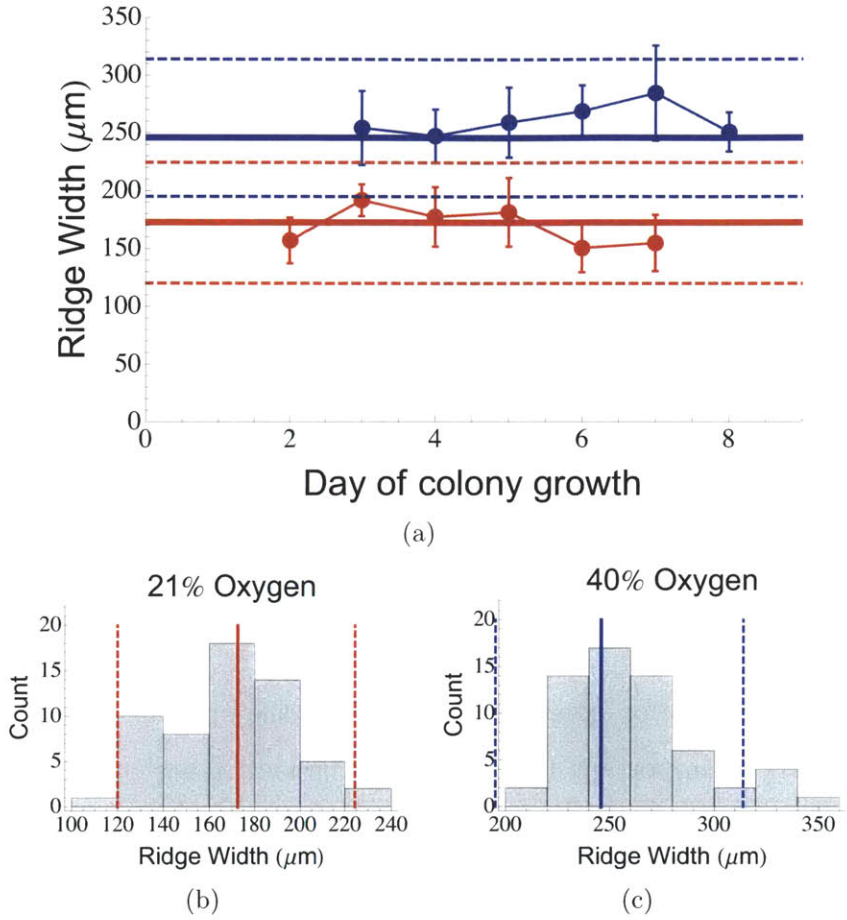


Figure 3-6: Predicted and observed ridge widths. (a) The average width of colony ridges as a function of time. The red points are for colonies grown in 21% oxygen while the blue points are for growth in 40% oxygen. The solid blue and red lines represent our predictions for optimal ridge width and the dashed lines are our predictions for the bounds of variation in ridge width given the calibrated experimental and biological variation in parameters. (b) and (c) represent histograms of all of the measured ridge widths in 21% and 40% oxygen respectively. The solid and dashed lines have the same meaning as in (a).

Similarly, for simulations of the base we find that E approaches zero when the height of the base reaches the maximum depth at which oxygen is able to penetrate. Thus, any additional height does not produce additional growth for the colony. Experimentally, we find that the average base thickness corresponds to the average depth at which oxygen is no longer detectable.

The observed colony features are such that the efficiency, E, is close to zero

but still positive which optimizes the total growth of the colony: If features grow wider (or taller in the case of the base) than this optimum the colony has invested a large amount of mass that will not grow quickly, if the features are smaller than this optimum the colony has invested a small amount of mass but there will less overall growth as a consequence. This raises the possibility that cellular behavior, including possible cooperation, has been selected for such that the emergent geometry of features maximally benefits the community.

It is surprising that in the top layer of cells growth seems to be arrested despite the rich amounts of oxygen and capacity for further growth. This could be the result of a rudimentary regulatory dynamic within the colony resulting in optimal community growth efficiency.

3.4.1 Morphological response to oxygen availability

We have shown evidence of oxygen control driving the optimization of colony form. Feature geometry is important because of the impact it has on the availability of oxygen within the colony. Given this interconnection between geometry and oxygen we would expect that changes in the external availability of oxygen would result in shifts in colony form. To test this hypothesis we exposed colonies to elevated concentrations of external oxygen (40% compared to the standard 21%).

In order to quantify these shifts we again calibrate the parameters of our model. Although it is infeasible to measure full oxygen profiles in the oxygen chambers it is possible measure the resulting base heights which we know correspond the depth at which oxygen is no longer detectable. This is sufficient for calibrating x_{fac} to the 40% oxygen conditions. We find that the bases height is much taller in the colonies exposed to 40% and that the range of x_{fac} values is between 4.5×10^4 and 7.7×10^4 with the expected value of 5.9×10^4. The fact that these values for x_{fac} are lower than for the 21% experiments implies a shift in the physiological processes or the physical properties and composition of the colony matrix in response to oxygen availability.

The observed widths in 40% are much wider than in 21%. Running our simulations with 40% external oxygen, and the calibrated range of parameter values, we again

successfully predict the mean, upper bound, and lower bound on observed ridge widths (Figure 3-6). Through our model we interpret the wider widths as the result of deeper oxygen penetration which extends the width at which a diminishing return on mass investment occurs.

Simple predictions for the dependence of feature size on oxygen availability

The change in the width of the ridges is related to the optimization of growth as regulated by the availability of oxygen. This is fundamentally related to how deep into the colony oxygen is able to penetrate. It has been shown in ref. [266] that an estimate for the penetration depth, d, of oxygen into the colony can be found by considering the steady state balance between diffusion and a constant consumption term for oxygen everywhere within the colony (instead of the reaction term used in equation 3.4 which depends on the local concentration of oxygen). A key feature of this estimate is that the penetration depth is related to the availability of oxygen according to $d \propto S_0^{1/2}$ where S_0 is the surface concentration of oxygen. This predicts that changing the surface concentration of oxygen from 21% to 40% will alter the penetration depth by a factor of $(40/21)^{1/2} = 1.38$. In our experimentation we find that the mean ridge width is 169.5 ± 3.5 μm (standard error of the mean) in 21% and in 40% the mean width is 261.1 ± 3.9 μm which is a factor of 1.54 larger which is comparable to what the simple estimate predicts. This suggests that features are regulated by the penetration depth of oxygen. In this context, ridge width would again be optimal given that any increases in width would cause cells to live within regions without oxygen where growth is impossible.

3.5 Discussion

The study of biofilms has far reaching applications: for example as model ecological systems to investigate the nature of cooperative and competitive behavior and dynamics [304], or to understand evolutionary life-history by interpreting geological structures which may represent microbial fossils [4, 208]. In each of these contexts it

is critical to understand the connection between community structure and resource availability.

One of the key findings of this paper is that the shape of features is such that total growth and return on mass investment is optimized based on how the feature interacts with the resource environment. As we have discussed the ridges maintain a constant and fairly uniform width even as they grow taller. Previous studies have shown that steady-state colony shape is the result of external resource limitations and occur when the environment can no longer supply resources for further growth (e.g. Refs. [304, 208, 209, 305]). Distinct from these previous findings our system is exposed to the open atmosphere where external resources gradients are effectively non-existent. Thus in our colonies the outermost layer of cells are not resource limited and do have the capacity for further growth as evidenced by the oxygen profiles and the observation that the height continues to grow. The ridges have the capacity for horizontal growth and any simple dynamic model driven solely by growth rate would predict increasing ridge width in time. Thus there must be some mechanism for maintaining constant widths which are advantageous to the overall colony. We have experimented with simple genetic knockouts and found that single knockouts of the pili motility leads to colonies with no ridges. This points to a complex interaction of multiple physiological features which manifest as the dynamic regulation of optimal feature formation. However, determining the precise system of interconnected mechanisms remains the subject of much future work.

Here we demonstrate that prokaryotic biofilms can exhibit complex spatial patterning at spatial scales which are comparable to possible microbial fossils. Previous interpretations have suggested that what could be microbial fossils are likely eukaryotic in nature [4], but our work shows that it could be hard to distinguish the difference between prokaryotic and eukaryotic fossils on the basis of spatial structure alone. Our model system, a wet substrate exposed to an oxygen rich atmosphere, corresponds to conditions that are possible for the early earth. Albani et. al also suggest that these biofilms could have lived in an oxygenated water column [4]. Our work shows that these same spatial scales could also be in situations where the colony is exposed

directly to the atmosphere in a moist environment.

Within our biofilm system we were able to directly test the effect of different atmospheric oxygen concentrations where we find that the width of features changes predictably. This opens up a variety of avenues for verifying microbial fossils and using them as proxies for the ancient environment.

The *P. aeruginosa* biofilm has important medical applications as *P. aeruginosa* is one of the most prevalent infections in patients who have the inherited disease cystic fibrosis which among other symptoms causes patients to build up mucus in the lungs. The nature of how *P. aeruginosa* is living in the lungs is not yet fully understood, but the thick mucus layer beneath an open atmosphere [301] has a strong correspondence with the *P. aeruginosa* biofilm system studied here. Here we find that biofilms produce optimal structures in order to enhance growth. The success of the colony depends on the ability to effectively build these optimal structures. We find that colonies of single genetic knockouts targeting motility lead to colonies without ridges. Targeting physiological aspects closely tied to colony morphology may open up new avenues for treatment of biofilm infections and even in eradicating biofilms from industrial and food systems.

Here we have shown that the geometry of these features is optimal for resource acquisition and reproductive success and that this geometry is responsive to external resource availability. Previous work has shown that simple cooperation in the form of extracellular matrix production can lead to beneficial feature formation [304]. Moving forward it will be important to understand the complex set of cellular behaviors that are responsible for the emergent regulation of these features.

3.6 Methods

3.6.1 Bacterial Strains and Growth Conditions

The phenazine mutant (Δphz) was used for all experiments. Bacterial cultures of Δphz were routinely grown at 37°C in Luria-Bertani (LB) broth overnight. For all

oxygen profiling experiments colony biofilms were grown on 1% tryptone/1% agar plates amended with 20 μg/mL Coomasie Blue (EMD) and 40 μg/mL Congo Red (EMD). Colony biofilms were grown at room temperature (22-25°C) and at high humidity (> 95%). Sixty milliliters of the medium was poured per 10-cm square plate (Simport, D210-16) and allowed to dry with closed lids at room temperature for 24 hours. Ten microliters of the overnight culture was spotted on these agar plates and colony biofilms were grown for 8 days.

Colony geometry measurements

We measured the width and height of spokes within the colony biofilm using a digital microscope (Keyence VHX-1000).

3.6.2 Oxygen depth profiles

Oxygen profiling was done using a miniaturized Clark-type oxygen sensor (unisense; 10 μm tip diameter) on motorized micromanipulator (unisense) for depth control. The electrode was connected to a picoampere amplifier multimeter (unisense) and polarized at -800mV. The sensor was calibrated using a two-point calibration system at atmospheric oxygen and zero oxygen. The atmospheric oxygen reading was obtained by placing the electrode in a calibration chamber (unisense) that contained well aerated deionized water. Complete aeration was achieved by constantly bubbling the water with air. The zero reading was obtained by bubbling water in the calibration chamber with ultra-high purity nitrogen gas (TechAir). All calibration readings and profile measurements were obtained using SensorTrace pro 2.0 software (unisense).

Chapter 4

Physical, genetic, and environmental controls of biofilm structure

Preface

In Chapter 3 we saw how features with a given geometry optimize the metabolism for individual biofilm features. Part of this regulation involves maintaining a fixed width as ridges grow taller. We still do not understand how this width is maintained by cellular physiological processes which is the key to understanding the transition to multicellularity. In this chapter we examine the connection between genetically determined physiology, environments, and the behavior and benefits of spatial patterning.

This chapter represents ongoing analyses and experimentation which we plan to submit for publication in the spring under the working title "Genetic and environmental control of the spatial organization of biofilm patterning" by C.P. Kempes, C. Okegbe, and L.Dietrich, where I will be the corresponding author. The colony mutant library along with image time-series was produced by Lars Dietrich and Chinweike Okegbe. I conducted all of the data extractions, analyses, and produced all of the

mathematical models and simulations.

4.1 Introduction

As we have discussed already, the morphology and architecture of multicellular systems is closely connected with function, physiology, and the ability of an organism to cope with a given environment. At all scales of life morphology is, at its most rudimentary, the dynamic emergence of spatial patterning. To understand the evolutionary history of morphology it is important to first determine how patterning evolved in unicellular communities and then how this may eventually have allowed for the transition to complex multicellular life. For patterning in unicellular communities there are two primary considerations: What are the mechanisms producing patterning? What are the benefits of the emergent pattern?

In a two-dimensional system horizontal patterns can emerge from the simple interplay and feedback of diffusion, reaction, and production of substrates and/or chemical signals as originally discussed by Turing [278]. Most previous efforts to model biofilms focus on the interplay between the response of motile bacteria to chemical gradients, and the production, consumption, diffusion and degradation of those chemicals [6, 84, 151] (see ref. [182] for a general treatment and ref. [276] for a review). Previous studies have shown that these principles lead to spatial patterns manifested in the direction in which cells are swimming in a two dimensional wet environment for organisms as diverse as the prokaryote *E. coli* to the slime mould *Dictyostelium discoideum* [6, 46, 248].

The simple physical treatments of these patterns, as originally conceived by Turing, have been criticized for lacking links to, or an understanding of, genetic control [129]. Similar criticisms are also emerging in the field of metabolic ecology. The metabolic theory of ecology begins with the observation that many properties of organisms scale with one another in a systematic manner across diverse species often following simple power laws (see for example [193, 140, 44]). Recently several mechanistic theories regarding the origin of these scalings laws have been developed [140, 292, 291]. One of the most widely applied of these focuses on the internal distribution (vascular) networks of plants and mammals [291, 292]. This model as-

sumes that over time evolution will optimize these networks in order to minimize the amount of energy expended to transport internal resources. In deriving the optimal distribution network the authors are able to show that the scalings of many organism traits can be explained using a single framework. However natural historians have commented that there is no mechanism for evolution in these models. At the foundation of this point is that these metabolic models do not have an explicit role for genetics and there is no explicit link between the network structure and its encoding within heredity or genetics. This poses a serious challenge for advancing the metabolic theory because the genetic regulation of morphological development is incredibly complicated in multicellular organisms.

More recent efforts in microbial colonies and embryonic development have linked basic pattern formation to specific genetic regulation pathways along with the processes of diffusion and reaction [248, 230, 5, 283, 276]. Yet much of this work is concerned with explaining the detailed origin of some very specific phenomena. For example embryos experience patterning as part of the development process and many models focus on explaining this. But the phenomena is singular, there is one pattern whose explanation is sought by means of increasingly accurate and sophisticated models. In order to understand the evolution of a specific pattern or morphology in response to a given set of selection pressures, say the spectrum of leaf venation patterns observed across a range of tree species, we need a simple model system where a variety of patterns can be created by systematic genetic knockouts and alterations to external environmental conditions.

As we saw in Chapter 3 our biofilm system responds to both external resource gradients and genetic manipulations. Given the rich spatial patterning exhibited by these biofilms, and the diversity of patterning that results from genetic knockouts, this system provides us with an ideal model system for exploring the evolution of developmental morphology in multicellular organisms.

Our ultimate interest is in understanding how the *P. aeruginosa* colonies evolved the capacity to form rich spatial patterning, what function (if any) this serves, and how this patterning is controlled by the evolved genetic and/or environmental regu-

lation.

4.2 Mechanisms controlling horizontal pattern formation

Our work on a single feature in Chapter 3 suggests that colonies regulate ridge width in order to optimize growth. Yet this top level analysis does not describe the underlying dynamics or physiological mechanisms which lead to maintaining a constant ridge thickness. Similarly, our analysis of a single feature suggests that ridges grow taller because they experience enhanced oxygen uptake and growth, but this does not explain why regularly spaced features form nor what spacing should be expected between ridges. In our biofilm system the colonies not only display rich horizontal patterning but also complex temporal dynamics where individual ridges are lost or added as the colony grows vertically and spreads horizontally. These temporal dynamics allow us to uncover the underlying mechanisms of patterning.

(a) (b) (c)

Figure 4-1: The temporal evolution of the biofilm.

Figure 4-1 displays a colony at three points in time. It can be seen that the ridges radiating outward from the center ring of the colony (referred to here as "spokes") change in number as the biofilm matures. Initially there are many ridges densely packed but in time many are lost. Over four days of growth roughly half of the spokes disappear. We quantify this temporal dynamic by looking at the the change in average spoke spacing as a function of time. The angular distance between two spokes is the best metric for making this quantification as we are looking at spacing

on a circle. We find that the mean of the angular distance between adjacent spokes continually grows in time following a fairly smooth curve (Figure 4-2(a).)

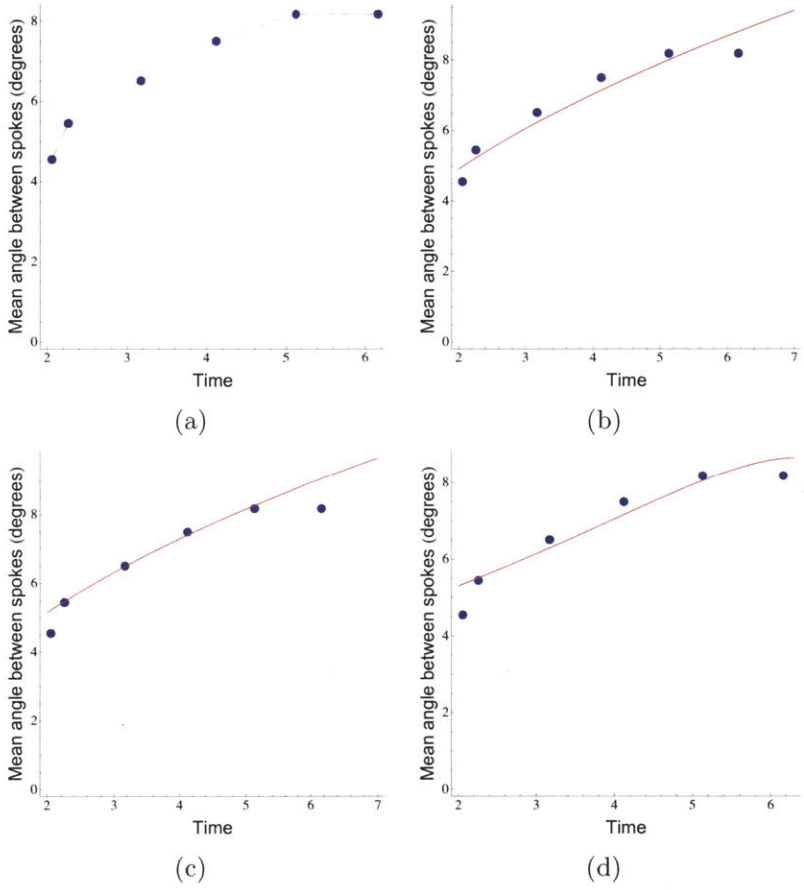


Figure 4-2: The measured temporal dynamics of mean angular distance between ridges. (a) The raw data for mean angular distance. (b) The best fit of a general power law to the data where the value of the exponent is .52 consistent with diffusive scaling. (c) The best fit of a general power law to the data excluding the first and last time-points.The exponent is .50 and the fit is free from any systematic offsets except for the the first and last time-points.(d) The predicted temporal dynamics of ridge spacing given that ridges maintain a constant average metabolic rate while increasing in size as described by equations 4.1-4.3. In each plot error bars representing the standard error of the mean are shown but fall within the plotted points demonstrating the regular spacing between neighboring ridges.

The continuous nature of this curve implies some set of dynamics which regulate ridge spacing and smoothly evolve in time. Given previous work [278, 276, 6, 84, 151, 46, 248, 230, 5, 283] and our knowledge of the *P. aeuriginosa* biofilm (Chapter 3) the mechanisms for controlling ridge spacing could include temporally evolving resource

gradients, the dynamics of chemotaxis, or regulatory chemical production, diffusion, and response. This provides a wide search space for detecting a specific mechanism and so it becomes important to consider general analyses which might help us identify and classify the general category of mechanisms we are dealing with.

Many of the possible mechanisms (e.g. signaling or resource gradients) would be strongly connected to the details of diffusion within the biofilm. In general, diffusion is characterized by a known relationship of how fundamental distances scale with time. Diffusion, in any number of dimensions, is described by a distribution of random processes wherein the root-mean-square displacement scales like $d \propto t^{1/2}$. For any diffusive process the key length scales change with the square-root of time. The measured angular distances are proportional to the linear distances. We find the scaling between distances and time by fitting a general power, $d = d_0 t^{\alpha}$, to the data in figure 4-2(a). The best fit gives an exponent of $\alpha = .52$ which is very close to the expected value of $\alpha = .5$ for diffusive scaling. The best fit describes the data well although there is some amount of deviation (Figure 4-2(b)). This offset could be due to the fact that the resource supply, growth, and chemical consumption rates are not in steady state over the development of the colony. These changing features add second order temporal dynamics to the processes of diffusion. In addition, the offset is mainly the result of the first and last data points.

Excluding the first and last timepoints gives a much better fit with $\alpha = .50$ (Figure 4-2(c)). Excluding these points for the fit is justifiable because the late-stage time-points are likely in a situation where resources from the agar are starting to be diminished as evidenced by the suddenly decrease in vertical ridge growth. Early-stage time-points may be characterized by ridges which have not reached their mature width or by large resource gradient shifts. It should also be noted that there is significant variation in the mean angular distance for each timepoint highlighting the stochastic nature of the potential physiological-diffusive mechanisms regulating ridge spacing.

The temporal data suggests that horizontal patterning may follow the scaling of diffusive processes. If diffusion is involved in setting the temporal dynamics this

raises two key questions: 1.) How can we uncover what colony processes are behaving diffusively (e.g resource gradients or chemical signals) and how these are setting ridge spacing given a potentially large number of interacting physiological and genetic mechanisms? 2.) What purpose, if any, do the temporal dynamics for ridge spacing serve? We address question 1 in section 4.3.1.

We found in Chapter 3 that the geometry of an individual ridge is related to the optimization of resource acquisition and overall growth. Thus in order to address question 1 it makes sense to consider how multiple ridges interact via the resource environment. A longstanding concept for trees living in densely packed competitive forests is the "self-thinning" rule (e.g refs. [306, 296, 294, 83]) which states that as forests mature they become less dense. Recently a theory has been developed that quantitatively predicts that the distribution of tree sizes within a forest is associated with a specific average distance between trees [294, 83]. This result follows from the observation that metabolism and resource use increase with tree size and thus, given a fixed resource supply rate, larger average size implies that fewer trees can be supported. Similarly, in a mature forest where all the supplied resources are consumed the growth of any given tree implies the mortality of some other tree.

Our work in Chapter 3 suggests that the geometry of single features benefits entire regions of the colony opposed to individual cells. Cells within the feature have the capacity for horizontal ridge growth beyond the optimal width but this does not occur suggesting that there is a basic cooperative regulatory dynamic governing the colony. Given this observation we can think of each ridge as an "organ" or "individual" of sorts. Similar to forests each ridge is drawing resources from a common reservoir in competition with other ridges and this may have consequences for the spatial structure.

To understand these resource dynamics we use a zero-dimensional model where a common resource pool feeds a number of ridges and the total flux of resources is divided between each ridge. In our model ridges actively grow and, since we know that a constant width is maintained, all growth occurs vertically. The fixed width of ridges also ensures that the average availability of oxygen within a ridge is approximately

constant even as vertical growth occurs. Given that the oxygen concentration is fixed within the colony there is no need to explicitly model geometric changes. Instead we summarize growth by changes in the total mass which is proportional to height. In a zero-dimensional model this is equivalent to modeling the density of cells a. Following equation 3.4 This system can be modeled as follows:

$$\frac{\partial [N]}{\partial t} = \frac{1}{n_r} D \frac{\Delta [N]}{(\Delta x)^2} - n_r \left(\frac{\mu'_{max}}{Y_N} \frac{[N]}{k_N + [N]} + P_N \right) a \tag{4.1}$$

$$\frac{\partial a}{\partial t} = \left(\mu'_{max} \frac{[N]}{k_N + [N]} \right) a \tag{4.2}$$

where N is a limiting nutrient, k_N, Y_N, and P_N are respectively the half-saturation, yield coefficient, and maintenance consumption rate for that resource, μ'_{max} is the maximum growth rate for a given concentration of oxygen, and n_r is the number of ridges in the system, and a is the density of cells.

The first term in equation 4.1 represents the diffusive flux of resources from the common reservoir into each of the ridges. This is a "two box" model (reservoir and ridge) and thus $\Delta[N] = [N]_{reservoir} - [N]$ is the difference between the concentration of nutrients in the reservoir and the colony. The $1/n_r$ factor accounts for the even division of resources between ridges. The second term in equation 4.1 is the consumption of nutrients which consists of a growth and maintenance term. Equation 4.2 is the growth of the colony as represented by the density of cells a.

Numerically simulating the dynamics of equations 4.1 and 4.2 with a fixed number of ridges demonstrates that the ridges follow a standard sigmoidal growth curve and resources decrease in a similar fashion. These dynamics illustrate that each ridge is undergoing increasing resource stress over the course of its growth. In contrast, within a forest system the competition for resources causes the density of trees to decrease with increasing average tree size such that each tree experiences constant resource stress and is able to meet its metabolic needs. Applying this analogy to the colony we find the thinning rate of ridges which would allow each ridge to maintain a constant metabolic rate per unit mass. Metabolic rate in the ridge is proportional to

the nutrient consumption rate and a constant average metabolism is thus given by

$$\left(\frac{\mu'_{max}}{Y_N}\frac{[N]}{k_N+[N]}+P_N\right)n_r\Big|_t=\left(\frac{\mu'_{max}}{Y_N}\frac{[N]}{k_N+[N]}+P_N\right)n_r\Big|_{t+\Delta t} \tag{4.3}$$

We again numerically simulate the temporal dynamics from equations 4.1 and 4.2 and use this condition to solve for the number of ridges at each point in time. The number of ridges, n_r, must be adjusted as N declines to satisfy equation 4.3. This provides a thinning rate and thus the temporal dynamics for the angular distance between ridges which we find matches the general shape of the observations (Figure 4-2(d)). The measured thinning rate of ridges is consistent with the ideal for maintaining constant resource availability and average metabolism within each ridge.

Given that these dynamics are also consistent with a diffusive process, this observation provides for a couple of hypotheses about the dynamics of ridge spacing: 1.) ridges are sensitive to a critical resource supply rate. The competition for and diffusion of resources causes certain ridges to be outcompeted by others and this dynamically sets the spacing. 2.) The diffusive dynamics represent some chemical signaling and cellular response which has been selected for to dynamically inhibit some of the ridges thus allowing dominant features to maintain a constant average metabolism and growth rate.

4.3 Genetic and physiological mechanisms of pattern formation

As we have discussed, there could be numerous interacting physiological processes which produce the spatial patterning observed in these biofilms. These interacting mechanisms form a "program" for pattern formation and we would like to uncover what that program is in order to understand how it evolved and how it is dynamically modified to deal with a variety of environmental contexts (Figure 4-3). To do this we consider the patterns that result in colonies with various genetic knockouts of a single gene. The basic pattern formation studied in Chapter 3 is already the result

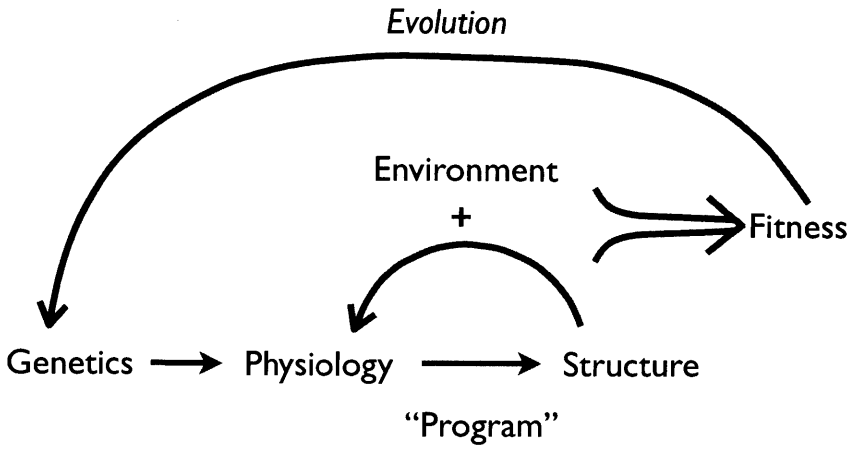
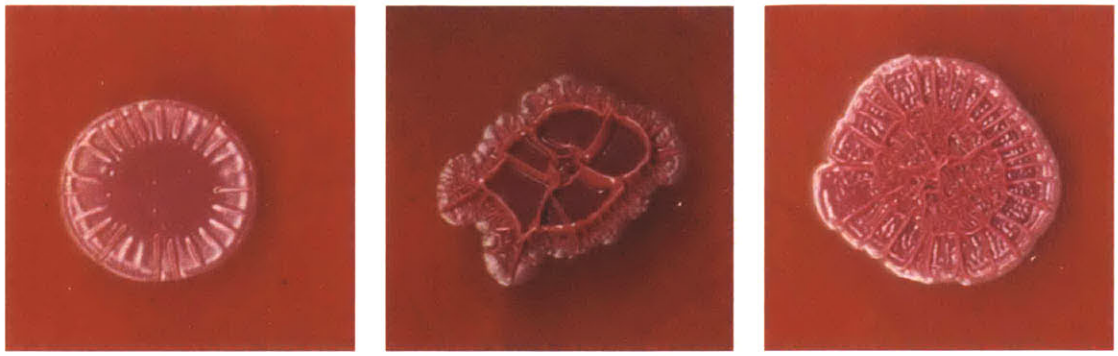


Figure 4-3: Schematic of the evolutionary process determining structure. Our concept of colony patterning is that genetics dictate cellular behaviors. Those behaviors combine through intricate interactions to produce colony structures. These structures serve as the primary mediator between the environment and the colony (forming a secondary or emergent physiology). As such, structure and environmental conditions determine the fitness and survival likelihood of the colony and thus feedback on the underlying genetics.

of a genetic knockout of the ability to produce phenazines. As we discussed this likely increases redox stress within the colony and thus the basic pattern formation is likely a response to environmental pressures. Many other genetic knockouts produce a variety of spatial patterning as illustrated in figure 4-4. By examining the effect of individual genes, and their physiological implications, on patterning we can begin to understand how various cellular processes combine to control spatial pattern in response to environmental stresses. In this section we develop various metrics that allow us to identify similarities and differences between various colony patterns.

4.3.1 Diffusive differences amongst colony mutants

Given our findings regarding diffusive control and resource competition as likely drivers of spatial patterning we first look at the time-series of patterning for a variety



(a) (b) (c)

Figure 4-4: Sample of colony patterns produced by single genetic knockouts. (a) is a knockout of an uncharacterized gene (b) is a knockout of fliC which is responsible for making flagellin a filamentous protein of the flagella [161] (c) is a knockout of the survival protein surE which is related to stress response [154]. The brightness and contrast of these images has been altered for clarity, this does not change the apparent structure of the colonies.

of mutants. Figure 4-4 depicts a sampling of mature colonies used for this analysis which represent a diverse set of morphologies. The mutants depicted in figure 4-4 are the single genetic knockouts of fliC, a filamentous protein involved in flagella formation [161], surE a protein related to survival in stationary phase (interpreted as stress response) [154], and an uncharacterized protein. Comparing the time-series of these mutants to Δphz we observe interesting similarities and differences as shown in figure 4-5. The ΔsurE knockout is most similar to the Δphz colony in terms of the scaling of angular distance with time. The exponent for ΔsurE is $\alpha = .62$ where the standard error is .29 and thus includes .5. This is again suggestive of a diffusively controlled process. The exponent for uncharacterized mutant is also $\alpha = .62$, but this is only for the early portion of the data. It can be seen in figure 4-5 that for later times this mutant deviates from this curve significantly and the best fit to the entire time-series is greater than linear. Similarly, the best fit of ΔsurE excludes the first data-point where superlinear scaling is also in play. This again suggests that there may be time-sensitive temporal dynamics which start to affect ridge-spacing in the late or early stages of colony development similar to what we observed for Δphz. It is interesting that these two mutants deviate from diffusive scaling at different stages of development (late vs. early) highlighting the importance of cellular physiology in

the specific temporal dynamics of development.

The most radical example of physiology altering temporal morphology is for ΔfliC where we find that ridges tend to fill open space, rather than thinning, as the colony matures. The exponent for ΔfliC is $\alpha = -.82$. Here the effects of altered motility overwhelm any diffusive dynamics. It is thus possible that the dynamics observed in many other colonies are the effect of motility and chemotaxis which often leads to diffusive behavior (e.g. refs. [63, 8]). This provides an alternative hypothesis to resource diffusion for the mechanism responsible for ridge spacing.

It is interesting that ridges thin more quickly and to a greater extent for the ΔsurE knockout compared to Δphz. This suggests that the elimination of this stress response mechanism produces additional pressures for the colony which requires that there be fewer ridges at any given resource supply rate.

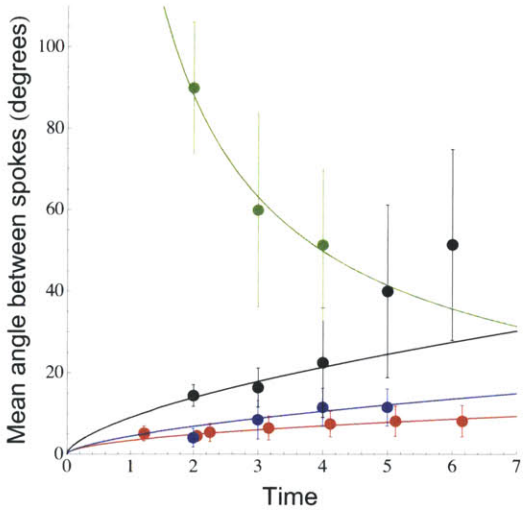


Figure 4-5: Differences in the relationship between time and the angular distance between adjacent ridges. The red curve is for the standard Δphz mutant already shown in figure 4-2, the green curve is for the ΔfliC knockout, the blue curve is the ΔsurE knockout, and the black curve is for an uncharacterized genetic knockout. The drawn curves are best fits of a general power law. For the blue and black curves this is fit to the last and first three time-points respectively. The exponents are $-.82$ for the green curve, and .62 for both the blue and black curves.

4.3.2 Network metrics for genetic categorization

It does seem that diffusive control, in the form of resource supply, the movement of organism, or chemical signaling, is important in most mutant colonies, however certain gene knockouts can radically shift these temporal dynamics highlighting that the "program" controlling spatial patterning is a complicated intersection of physiology and physics. In order to better uncover the various physiological dynamics at play we need additional metrics which take advantages of the rich patterning beyond the simple temporal change in spoke spacing. Here we consider that each of the spatial patterns can be treated as a network of interconnected ridges. This patterning most likely does not serve any direct function as a network (i.e. distribution), however it is a morphological trait which we can use for categorization or classification. The pattern of ridges does serve some function which perhaps we can uncover by using graph theoretic metrics as a general tool for identifying patterning. The broad aim is to develop a series of metrics which can be used to categorize mutants with similar impact on colony morphology (figure 4-6). Ultimately, we can look at the similarity of the genetic knockouts found within a given category and potentially gain insight into the major physiological processes responsible for patterning. Since the colony mutant library consists of over 1000 genetic knockouts this holds great potential for discovering the interplay of numerous physiological processes.

We selected a set of colony mutants which represent a spectrum of patterning and manually labelled the network of features observed in each of the mature patterns. Figure 4-7 shows the mutants that we analyzed along with an overlay of the graph which represents the spatial network. The mutants used are the basic Δphz mutant encoded for the whole colony and the center region alone (Figures 4-7(a) and 4-7(c)); an oxygen sensing mutant which we will refer to as ΔO_2 (Figure 4-7(b)); an oxygen and redox sensing PAS domain protein knockout which we will refer to as Δredox (Figure 4-7(d)); a cytochrome oxidase subunit knockout which we will refer to as Δoxidase (Figure 4-7(e)). It is interesting to note that these various knockouts produce visually distinct patterning even though all are involved in modulating redox

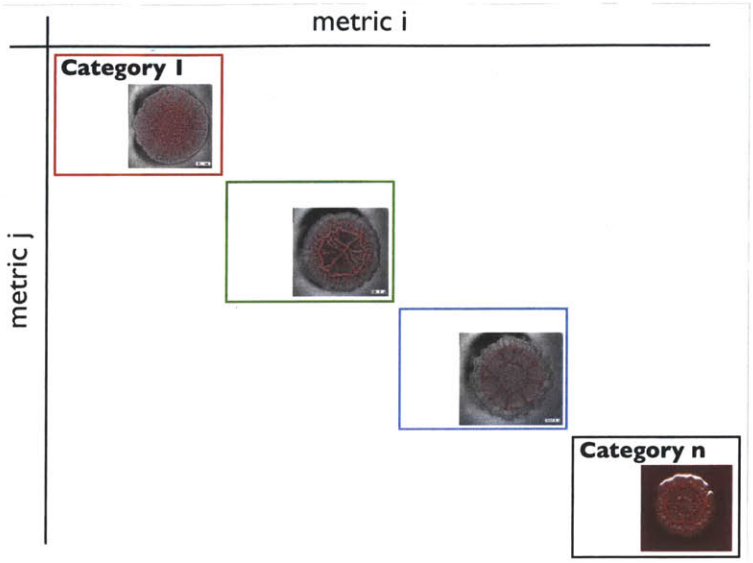


Figure 4-6: Schematic of graph theoretic metrics being used to classify colony patterns. Examples of possible key metrics are the average spacing between features, or the connectivity of the various features in the network.

stress within the colony. This highlights the complicated connection between morphology and redox stress and the complexity of the underlying "program" for spatial development.

For comparison we have also included an analysis of a regular square lattice graph (Figure 4-7(f)) which represents the highest level of regular structure and a graph with randomized connectivity (Figure 4-7(g)). We also include an analysis of other networks found in nature such as the electrical power grid of the United States [289] (Figure 4-7(h)).

We first consider the basic connectivity and topology of these graphs. The degree of a given node is defined as the number of other nodes that it is connected to and the entire graph can be summarized by considering the probability that a node has a given degree. Figure 4-8 shows this for the graphs analyzed on both linear and logarithmic scales. The linear scale is useful for analyzing the degrees with strong peaks for a given graph while the logarithmic scale is important because it shows the scaling of the degree distribution. For many diverse networks the degree distribution has been suggested to scale according to a power law with a negative exponent [17].

For the degree distribution a lattice has strong peaks at a small number of degrees

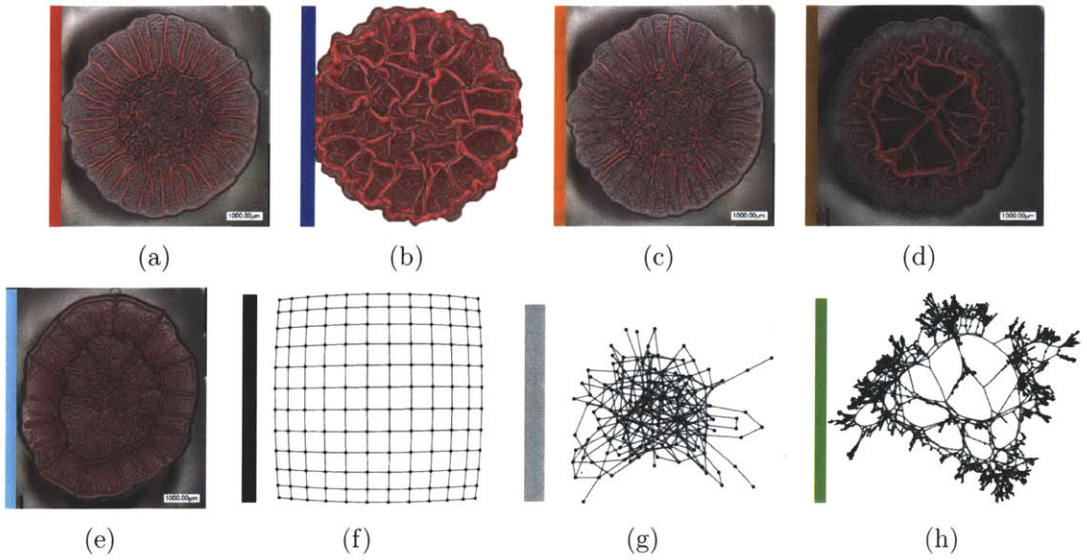


Figure 4-7: The comparison of colony mutant networks along with synthetic and other natural networks. (a) and (c) are the standard Δphz mutant where (a) represents only the center portion. (b) is an oxygen sensing mutant which we refer to as ΔO_2. (d) is an oxygen and redox sensing PAS domain protein knockout which we refer to as Δredox. (e) is a cytochrome oxidase subunit knockout which we refer to as Δoxidase. (f) is a regular lattice. (g) is a randomized graph. (h) is a network representing power grid of the United States [289]. The color bars on the left represent the legend for the plots in the following figures analyzing these colonies.

while random networks are more evenly distributed over many different degrees. We find that the center of the Δphz mutant is very similar to the ΔO_2 mutant, and both resemble a lattice with some amount of randomization. The ΔO_2 is more strongly peaked than the Δphz mutant which deviates more strongly from a lattice implying that there are more random connections. The entire Δphz colony is more evenly distributed and this highlights the differences introduced by the spoke pattern at the outside of the center pattern. Although the Δredox and Δoxidase mutants both have strong peaks at degree 3 they are distinguished from ΔO_2 by the strong peak at degree 1. The Δphz mutant is identified by the presence of a significant number of nodes with degree greater than 4.

The scaling of the degree distribution in figures 4-8(c) and 4-8(d) shows that generally the colony networks follow the trend of a lattice for small degree and a random graph for large degree. This highlights the random-lattice structure of these colonies. Each colony has a unique signature that is a combination of the small degree

and large degree composition. For example Δoxidase is characterized by a relatively large number of low degree nodes and few high degree nodes which corresponds to the loosely connected spoke topology of this colony. In contrast ΔO_2 has a relatively low percentage of both high and low degree nodes indicating the more lattice-like structure of this colony. These unique signatures are the first step in linking physiology and genetics to morphology.

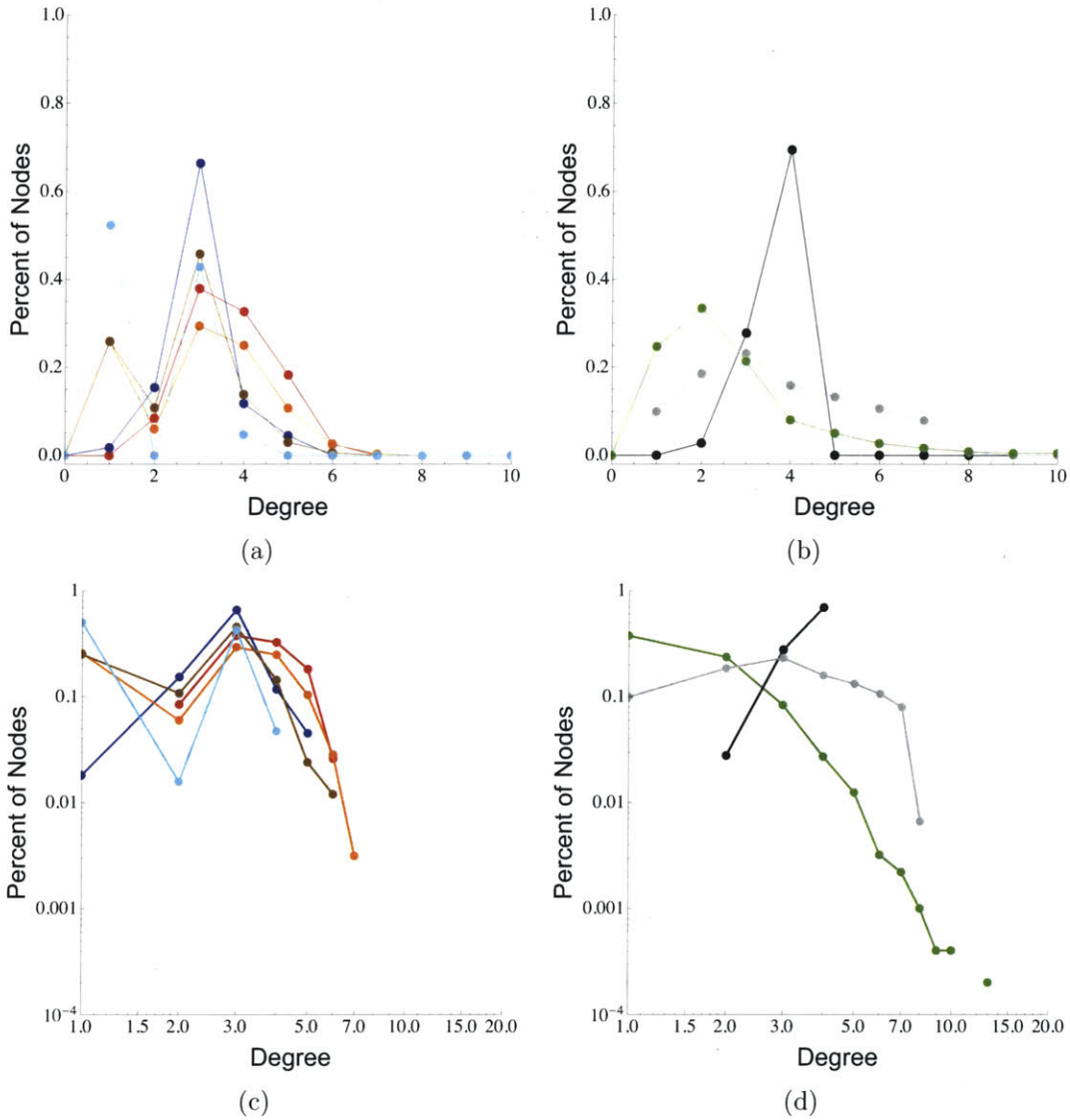


Figure 4-8: The degree distribution for various networks. The network can be identified by the color bars used in Figure 4-7. (a) and (b) use a linear scale while (c) and (d) use a logarithmic scale. Plots (a) and (c) show the results for the colony networks while plots (b) and (d) give the results for the power grid, random graph, and regular lattice.

Another metric for quantifying the overall connectivity of graphs is the betweenness centrality which is defined as the number of shortest paths (considering all node pairs) passing through a given node [186]. Distance is defined as the number of edges between nodes that are traversed. A large betweenness centrality for a given node represents that its hub-like nature while a low value corresponds to peripheral nodes. The distribution thus quantifies how many nodes are highly connected hubs compared to being peripheral.

Figure 4-9 gives the distribution of betweenness centrality for each of our networks. For this metric ΔO_2 and the center region of Δphz are nearly indistinguishable from each other and very similar to the regular square lattice. However the entire colony of Δphz is distinguishable from ΔO_2 due to the relatively high percentage of nodes with low betweenness centrality. Δphz, Δredox, and Δoxidase all have a larger number of nodes with low betweenness centrality owing to the presence of spokes. This is distinct from both the lattice and random graphs and provides a way to detect spokes in a colony. The distribution of betweenness centrality decomposes our colonies into two categories: lattice-like vs. hub and spoke, and moving forward this is an important metric for grouping physiologies.

The colony patterns are distinctly different from the power-grid network and are much more similar to either a regular lattice or random graph. The power grid is designed, or self-organized, to minimize the energy lost in distributing electricity between locations. We would thus expect this network to have very low average betweenness centrality, which it does, as this describes the resistance between one location and another. Low betweenness centrality should be the signature of a network designed for distribution. In contrast a lattice has much larger average betweenness centrality. The definition of a lattice is based on a dominant length scale which in real networks could be set by any number of processes including diffusion. Thus the observation that the colony networks look more similar to a lattice than a distribution network is consistent with our findings that diffusion is likely setting the spacing between features. The observation that the colonies are distinctly different from a distribution network implies that the observed "network" of ridges in the colony most

likely does not serve any direct function such as resource supply.

The function of the features in the colony are to gain access to oxygen as described in Chapter 3, the spacing of features within an emergent pattern is likely the product of diffusive dynamics and resource management. Thus the emergence of this network is likely the result of raised features seeking greater access to oxygen with a spatial organization related to resource competition and management.

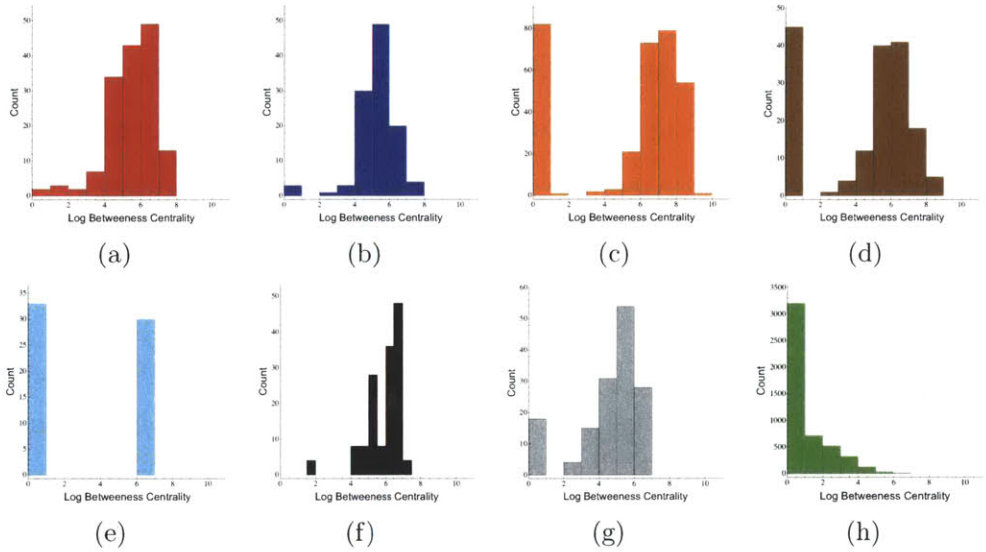


Figure 4-9: The distribution of betweenness centrality. The colors again match Figure 4-7.

4.4 Summary and Conclusion

We have found that the temporal dynamics of spatial pattern formation within the Δphz colonies follows a general diffusive process. This temporal dynamic is beneficial to individual features as it allows them to maintain a resource flux that supports a constant average metabolic rate. This complements our findings in Chapter 3 that show that colony features maintain a constant width to optimize oxygen acquisition and growth.

The underlying physiological structure of cells within the colony, as identified by genetic knockouts, can significantly alter colony morphology and development. The resulting colonies can have radically different temporal development, although, many

colonies display what is likely diffusively driven development over some portion of their growth. We find that disrupting motility eliminates the diffusive scaling of the temporal development implying that chemotaxis may be a key factor in the standard patterning and implied resource management.

In addition to motility we find that redox related functions are central to controlling patterning as determined by several knockouts that we studied. Using graph theoretic metics we found that several redox related knockouts give rise to distinct network structures. In these structures we are able to detect spokes and lattice motifs. We were successful in both grouping similar networks and distinguishing between subtly different networks. Moving forward it will be important to automate the application of these metrics to a large number of genetic knockouts to uncover the interplay of various physiological functions in producing patterning.

Once we uncover the key physiologies involved in this "program" we can then proceed with targeted experiments to measure the chemical and physiological state of the colony over the time-series of development. Two key techniques will be required for experimentally verifying proposed mechanisms: 1.) developing a variety of probes to measure the spatial and temporal state of key chemicals (other than oxygen) in the colony (e.g. resources such as nitrate and any candidate chemicals that might be used for signaling) 2.) tagging genes of interest with fluorescent proteins and monitoring their expression both spatially and temporally.

One of our original questions was whether the overall patterning of colonies served some complicated purpose such as distribution of resources through the network. Our analysis of the betweenness centrality of colony networks suggests that they are topologically different from distribution systems. The emergence of these networks is likely explained by the fact that individual features provide an advantage to the colony and the spatial arrangement of these features is the emergent feature of some complicated physiological-physical dynamics such as resource competition. However, whether the colony has evolved physiologies that lead to optimal management of resources between features remains an open question of much future interest.

In general, it is important to determine whether cellular traits have been selected

for to manage colony morphology, such as maintaining a constant ridge width. The connection between cellular physiology and the success of the colony can be determined by mixing together different genetic knockouts and simultaneously observing the competitive dynamics, effect on morphology, and regional health of the colony (measured in terms of resource availability and growth). Tracking the dynamic connection between morphology and competitive exclusion of various strains may help us to understand how physiology and morphology are co-selected by evolution and the general processes by which community organization and multicellularity evolve and are continually modified.

Chapter 5

Predicting Maximum Tree Heights and Other Traits from Allometric Scaling and Resource Limitations

Preface

In Chapters 4 we have investigated the evolution of organisms towards complex multicellular organization where structure and morphology have important implications for survival. Here we show how morphology continues to be of central importance in complex multicellular organisms as it mediates the interaction with local environments. We show that a focus on structure and basic physical laws makes it possible to predict regional trait variation from local resource availability.

This chapter is a reproduction of the article "Predicting maximum tree heights and other traits from allometric scaling and resource limitations" by C.P. Kempes, G.B. West, K. Crowell, and M. Girvan published in PLoS One [132]. I am the corresponding author for this study in which I conceived of and derived the theoretical framework and model calculations. I obtained the data and performed the predictive modeling and analysis of the output, and wrote most of the paper. Michelle Girvan and Geoffrey West acted as mentors and Kelly Crowell helped establish the geospatial

database and provided feedback on the intricacies of the predictive modeling. This work stemmed from an interest that grew in this topic through several undergraduate research projects.

Abstract

Terrestrial vegetation plays a central role in regulating the carbon and water cycles, and adjusting planetary albedo. As such, a clear understanding and accurate characterization of vegetation dynamics is critical to understanding and modeling the broader climate system. Maximum tree height is an important feature of forest vegetation because it is directly related to the overall scale of many ecological and environmental quantities and is an important indicator for understanding several properties of plant communities, including total standing biomass and resource use. We present a model that predicts local maximal tree height across the entire continental United States, in good agreement with data. The model combines scaling laws, which encode the average, base-line behavior of many tree characteristics, with energy budgets constrained by local resource limitations, such as precipitation, temperature and solar radiation. In addition to predicting maximum tree height in an environment, our framework can be extended to predict how other tree traits, such as stomatal density, depend on these resource constraints. Furthermore, it offers predictions for the relationship between height and whole canopy albedo, which is important for understanding the Earth's radiative budget, a critical component of the climate system. Because our model focuses on dominant features, which are represented by a small set of mechanisms, it can be easily integrated into more complicated ecological or climate models.

5.1 Introduction

A critical component for understanding the earth system is determining the interplay between biotic and abiotic factors, such as the interaction between forest characteristics and local meteorology [56, 174, 219, 167, 175, 239, 31, 238, 213, 22, 220]. At present a range of ecological perspectives and techniques are used for interpreting forest structure and dynamics at both the local and regional scale. Historical and ongoing modeling efforts have become increasingly accurate at describing critical forest features such as standing biomass and dynamic transpiration rates [176, 166, 167, 175, 268, 220, 221, 204, 47, 34, 61, 170]. Most of these models explicitly simulate the temporal and/or spatial dynamics of a forest and typically focus on a detailed description of a variety of coupled plant processes including transpiration, competition between trees, seedling dispersal, and mortality.

Another perspective for interpreting ecological features is the use of allometric relationships as a means to characterize the general variation of plant traits across

many species living in a variety of environments [187, 189, 43, 216]. These scaling relationships show that, on the average, many of the dominant physiological traits relevant to forest dynamics and structure are correlated with tree size following approximate power laws (e.g. [189, 190, 177, 191]). As such, size is viewed as the major determinant of variation among trees setting the baseline from which variation due to local, environmental, historical, geographical, and species related factors are considered secondary perturbations. Because of the relative simplicity of these relationships many models rely on basic allometries as part of a more complicated description of plant behavior (e.g. [176, 61]). Furthermore, there are conceptual frameworks from which these scaling laws, at both the individual and community level, have been derived (e.g. [292, 82, 187, 165]). On the other hand there is ongoing debate over the exact value of the empirical exponents of each relationship and the range of tree sizes over which they are valid, and, in general, it is not yet known what sets the dominant variability of the data around a given scaling law (see [216] for a review). Thus, it is unclear how useful the basic power-law relationships are in describing local variation or how applicable they are to modeling endeavors.

Here we create a model of plant physiology that focuses almost entirely on these scaling laws which we connect with an energy budget approach and couple to environmental resources in order to calculate an important component of this variation. In particular, we incorporate the relationships between basal metabolic rate, water availability, incoming solar energy, heat loss and ambient temperature. Because the underlying scaling laws represent the average tendencies across many species we apply a single tree characterization to a variety of environments. In our framework plant diversity is encapsulated according to the average trends across many species and the scaling laws allow us to use a single parameter, tree size, to determine a range of physiological traits. We show that this model successfully predicts the local and regional variation of maximum tree heights from a small number of environmental parameters (Figure 5-1). This coupling of various scaling laws also predicts more complicated relationships for tree traits such as the sigmoidal decrease in canopy albedo with increasing height. Our model can be extended to predict the variation of other plant

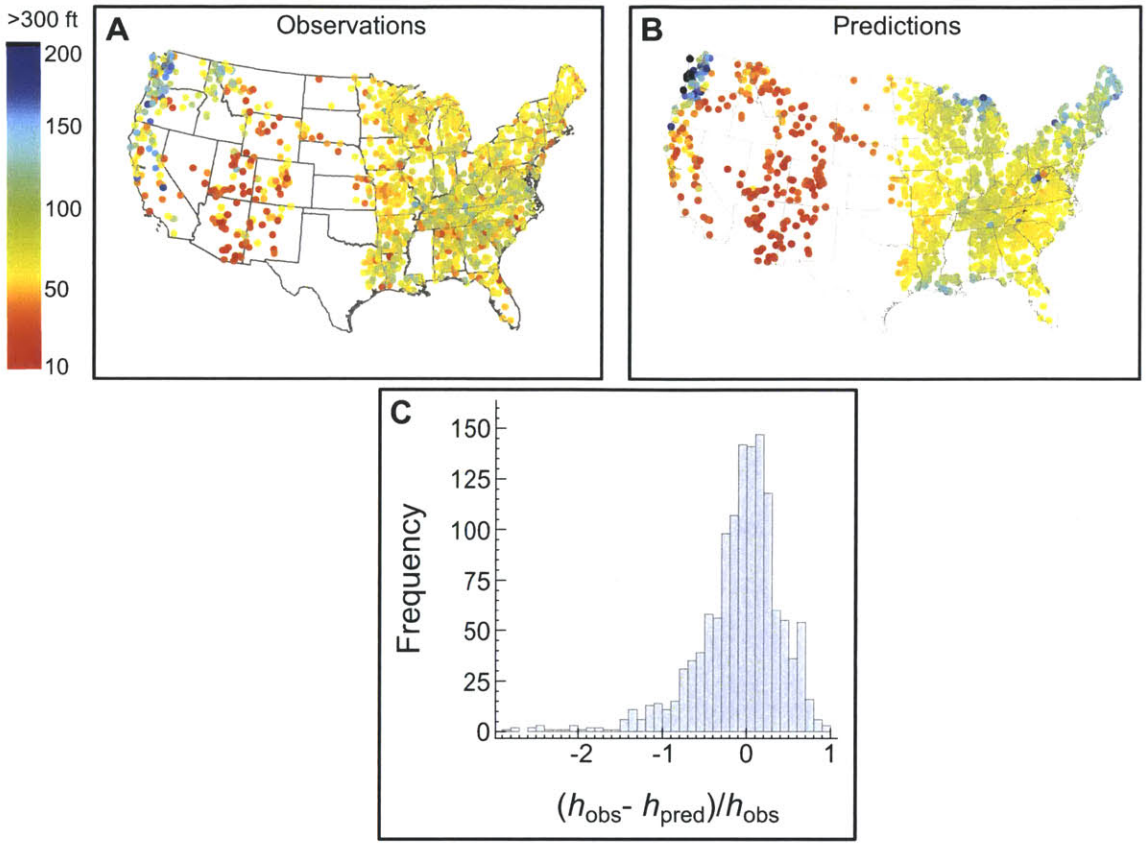


Figure 5-1: Comparisons between observed and predicted maximum tree heights. Maps of the continental United States comparing **(A)** observed and **(B)** predicted maximum heights of trees. **(C)** Histogram showing the distribution of deviations of the predicted maximum tree heights, h_{pred}, from their measured values, h_{obs}, expressed in terms of the dimensionless ratio $(h_{obs} - h_{pred})/h_{obs}$. The median of the entire distribution is $-.013$ and 20 values less than -3.0 were omitted from the histogram.

traits and we show how stomatal density depends on local mean annual temperature.

5.1.1 Previous modeling approaches

To understand the interplay between forest structure and local or regional climate it is necessary to both understand the competitive dynamics of trees within a stand and to couple tree physiology — at the individual or whole forest scale — to environmental conditions. At the regional scale, a common approach has been to focus on vegetation types coupled to atmospheric conditions. These models have successfully captured the geographic distribution of vegetation types and net primary productivity as well

as environmental processes such as moisture flux and runoff [97, 116, 117, 213, 31, 238, 287, 30]. For understanding fine-scale forest structure explicit temporal and spatial modeling and simulation efforts are becoming increasingly accurate at capturing local forest dynamics. Several models which aim to capture local phenomena focus on the small-scale competition of trees represented either as components or patches (e.g. the JABOWA model [34, 47]) or explicitly as individually trees (e.g. SORTIE [204, 61, 47] and TASS [170]). These models predict the gap structure of the canopy [47, 34, 61], the species composition and diversity of a stand [204, 61, 34], the standing biomass [204, 47, 34, 61, 170], and the size distribution of trees [170] at the local scale. In the case of SORTIE, the model tracks individual trees and simulates the coupled dynamics of canopy spatial structure, crown competition, light availability, seedling recruitment, growth, and tree mortality [204, 61].

The drawback of these models is that they are computationally expensive when applied to larger regions. The more recent efforts of the perfect plasticity approximation (PPA) have used basic assumptions about the interaction of individual trees to produce macroscopic equations (analogous to those found in statistical physics) for features such as the equilibrium size distribution of trees [268]. This technique captures the average interaction between competing trees without explicitly modeling each individual and thus can be inexpensively applied to larger regions. For features which represent the average of numerous trees (e.g. total density and average height) the PPA produces very similar results to the models which explicitly track individual trees [268]. The PPA also compares well with measurements for crown characteristics such as depth and radius [220] and the temporal dynamics of stand structure, biomass and successional patterns [221].

The models discussed thus far focus primarily on either the competition dynamics within a stand or the regional coupling of environmental conditions to vegetation. The ecosystem demography model (ED) connects these two approaches in an effort to more accurately understand forest dynamics coupled to the environment at multiple scales [176]. ED relies on plant functional types as a means for capturing local forest diversity and, similar to the PPA discussed above, relates an ensemble aver-

age to the complex dynamics of individual trees including the stochastic processes of mortality and succession. ED then couples this ensemble approach with numerous environmental processes such as atmospheric conditions, fire, evapotranspiration, and carbon sequestration. ED is able to capture important local and regional phenomena such as carbon flux, standing biomass, the stock of soil carbon, or the response of productivity to changing climate [176, 166, 167, 175]. Approaches like this hold much future promise for understanding both small-scale forest structure and regional vegetation patterns as they feedback with climate. However, these models require explicit temporal simulation, and decisions about how to represent plant diversity and physiology.

5.1.2 Steady-state allometric approach

Distinct from the models discussed above, the framework that we develop in this paper consists of a steady-state analytic calculation rather than a temporally and/or spatially explicit simulation. Our framework takes average local meteorology as an input and numerically calculates maximum tree height as an output. In comparison with the models discussed above our framework is not able to characterize detailed local phenomena such as temporal dynamics or species composition, but it does allow us to understand the average tendencies and constraints facing trees across different environments and this provides a useful foundation for incorporating more complicated processes.

We employ a single generalized tree across a range of environments without specific knowledge of local plant functional types commonly used in previous models [213, 31, 238, 287, 30]. In doing so we sacrifice accuracy at the local scale but gain a simple understanding of the average variations across environments. In the context of resources our framework lends insight into the mechanisms underlying deviations from the allometric scaling laws where, for example, we are able to show how different tree traits are suited to a given environment and predict the temperature-based variation of stomatal density (Figure 5-5). These variations in turn modify the size-based scalings for an individual tree species (see Sections C.5.4 and C.5.5). Ongoing work

is beginning to understand departures from the basic zeroth order allometric scaling laws [191, 217, 216, 215], however, it should be noted that the zeroth order theory has yet to be coupled with environments in order to test its predictive power. Our work provides another means for expanding the basic allometric scaling laws to encompass features that are relevant to more detailed modeling efforts.

5.1.3 Ecological relevance of tree height

We demonstrate the utility of our framework by predicting maximum tree height. We choose to focus on tree height because size is a natural quantity within the allometric framework and because height is an important indicator of various consequential features of a forest, such as its total resource use, biomass production rates, spatial distribution, and patterns of mortality and succession [81, 80, 181, 83, 82]. For example, frequency distributions of trees follow characteristically similar relationships across forests in different regions experiencing different resource environments [181, 83, 82]. These frequency distributions follow a power-law over a large range of the data with a drop-off for the tallest trees [181, 83, 82]. This implies that the tallest trees can be used to infer the size structure of forests. Given the significance of maximum tree height our framework offers future extensions for understanding regional and global energy budgets, water and carbon cycles, temperature feedbacks, and ecosystem dynamics in response to changing environmental factors from the perspective of average physiology. It should be noted that our framework can be used to predict the variation of other plant features beyond maximum tree height such as the environmental variation of stomatal density.

Beyond its importance as a predictor of forest demographics, tree height has been shown to influence competition between individual trees for access to light [135, 89, 91, 90]. However, the advantage of being taller comes with the added costs of growth and maintenance and this may set up a complicated evolutionary game between individuals [89, 135]. Maximum height has various correlations and related tradeoffs with other important plant traits [90, 91]. These include seed mass, overall growth rate, leaf mass per area, and wood density, each with environmental consequences

ranging from soil resource use, to biomass production rates, to competitive dynamics within a community [90, 91]. Our framework provides insight into the environmental and physical limitations of these evolutionary dynamics.

In general, tree height is constrained by the interplay between many competing factors including resource limitations, internal metabolic constraints, overall growth rate, maturation processes, the hydrodynamic flow through vascular tubes of the branch network, its geometry and topology, and biomechanical and gravitational forces [139, 91, 90, 194, 189, 241, 292, 191, 192, 194]. This complicated intersection of constraints is not unique to height but is a standard characteristic of most tree traits. Nevertheless, data on many properties of trees (Y) can be encapsulated and summarized in phenomenological scaling laws which typically approximate a simple power law form:

$$Y = Y_0 M^b \tag{5.1}$$

where M is tree mass, Y_0 a normalization pre-factor, and b the scaling exponent. Examples include tree heights ($b = .264$), respiration rates ($b = .78 \pm .04$ or $.81 \pm .02$), overall growth rate ($b = .66 \pm .01$), the frequency distributions of individuals ($b = -.79 \pm .02$), and trunk radii ($b \approx 3/8$) [190, 177, 189, 191].

These scaling laws represent the average variation of a given evolved trait across many species. Because trees have simultaneously negotiated the limitations imposed by multiple physical constraints over their complicated evolutionary trajectory, these scaling laws are likely the manifestation of multiple constraints. For example the evolved canopy structure must be both mechanically stable and able to gather sufficient solar resources in order for the tree to survive and compete. Thus considerations of either or both of these limitations may anticipate an observed empirical scaling law. By focusing on empirical scaling laws these constraints, whether known or unknown, are then implicitly incorporated into our model without needing to specify which limitations — or combination of limitations — are the most important. Both hydraulic (e.g. [139, 241]) and mechanical (e.g. [191, 194]) limitations are argued to constrain maximum tree height and our model incorporates both of these via various

scaling laws including the scaling of basal flow rates and the scaling of the canopy geometry. Beyond the inherent limitations of hydraulics or mechanics, ultimately and locally, maximum tree height is governed by the availability of resources. By connecting scaling laws to an interaction with the local environment we are considering the constraints of both resources and plant structure.

5.2 Results

5.2.1 Model framework

We investigate the survival of an idealized tree with features determined primarily by its size. These features include the number of leaves, canopy shape and size, and the root mass, all of which interact with the environment via the tree's requirements for light and water (Figure 5-3). Trees rely on their phloem and xylem for the internal distribution of nutrients and water. This circulation is a process of trees extracting moisture from the soil and making it available for evaporation, which drives the flow at the leaves. Accordingly, the rate of fluid flow through the vascular system has been a long-standing focus of environmental tree physiology[51, 173].

Our strategy is to compare flow rates that are constrained by resource supply with the flow rates that are required to sustain a tree of a given size in the absence of resource limitations. Both of these types of flow are governed by overall tree height according to scaling laws which relate various tree features to size. A basic assumption of our framework is that the essential tree traits required for building our predictive model scale with tree size according to approximate power laws (including isometric relationships). For many traits this is well supported by existing data. However, it should be noted that these power laws may break down for small trees where more complicated relationships hold (e.g. [191]) and some scaling exponents are known to have different values and confidence levels across different environments (e.g. [217]). These variations are beyond the scope of our efforts here. We focus on power laws because we are interested in the simplest construction of average behavior as a tool

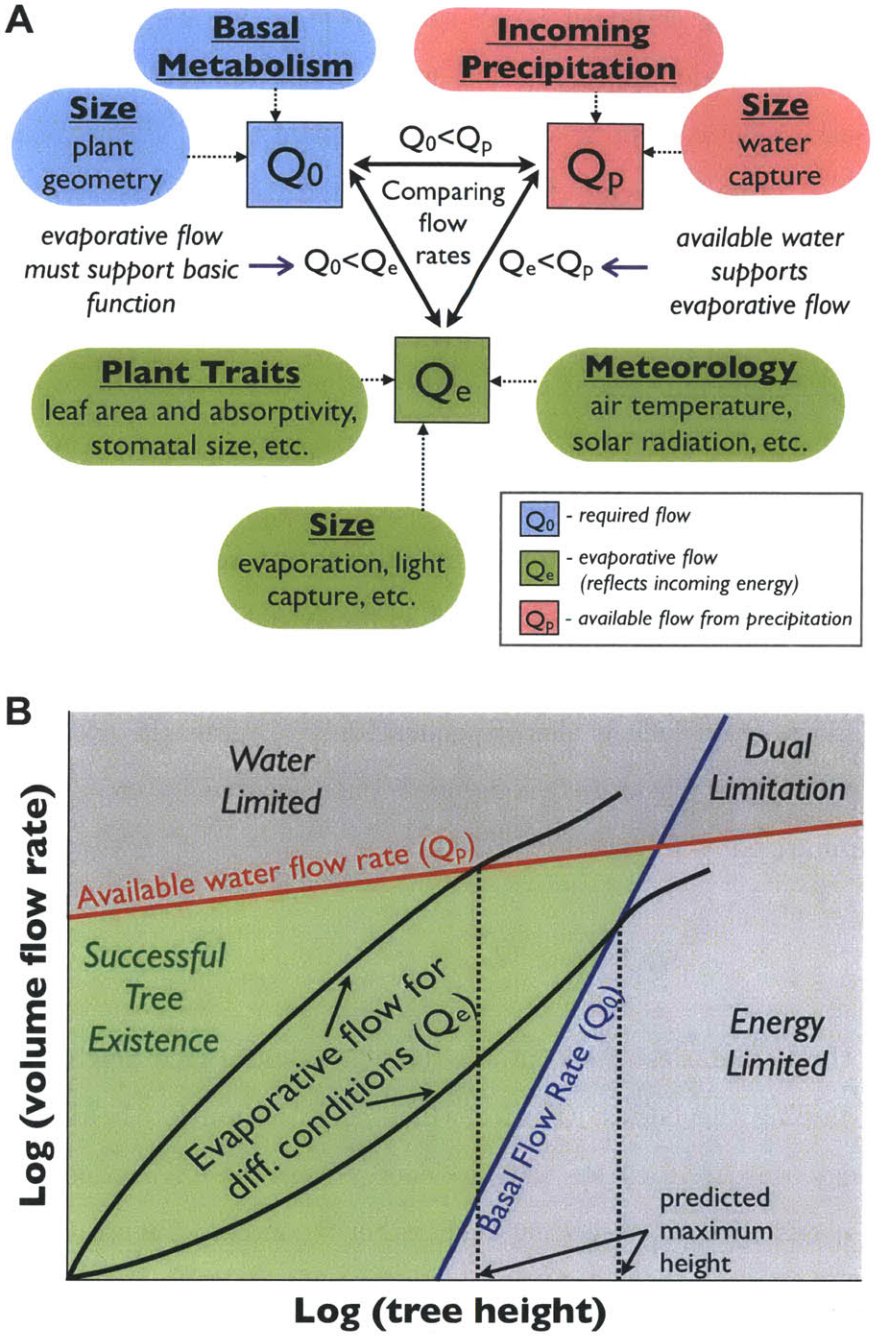


Figure 5-2: Schematics of the modeling framework. **(A)** The relationships between the required flow rate, Q_0, the evaporative flow rate, Q_e, and the available flow rate, Q_p, and the factors which influence them. **(B)** Limitation Diagram. Red Curve: the flow rate of available water, which is a function of precipitation and size, as described in the text. Blue Curve: the required flow rate determined from allometric scaling, which is a function of size but independent of environmental conditions. Black Curve: the calculated evaporative flow rate, which is dependent on both size and meteorological conditions. The intersection of the black curve with either of the other two determines the maximum tree height.

for predicting and understanding variation across species. Thus, we are testing the predictive power of the zeroth order approximation, which in this case are the widely used and studied power laws between body size and various plant traits. Future work should consider the higher order behavior of more complicated trait models.

Scaling relationships quantify how the total *required flow rate* of water in a tree, Q_0, changes with overall body size in order to support its basal metabolism [80, 292, 291]. We examine two principle limitations to the flow rate in trees: available water and energy (light and heat). Energy from the environment results in an *evaporative flow rate* of water through the tree, Q_e, which depends on both body size as well as on meteorological conditions, including air temperature, pressure, relative humidity, and solar radiation. This evaporative flow rate, which is the actual flow rate through a tree, must be met by a sufficient *available flow rate* of water from precipitation captured by the root mass, Q_p, which is also dependent on body size. In addition, Q_e must be sufficient to support basal metabolic needs encapsulated by Q_0. These constraints can be summarized as follows:

$$Q_0 \leq Q_e \leq Q_p. \tag{5.2}$$

Thus, Q_0 and Q_p set the boundaries of acceptable flow. Maximum tree height can then be predicted by finding the largest tree for which this relationship holds. In other words, our strategy searches for trees that use energy from the environment to meet their metabolic needs without exceeding their water resources. Figure 5-2 A summarizes our model, highlighting the factors involved in calculating Q_0, Q_e, and Q_p.

Graphically, Eq. C.4 implies that if we plot $Q_0(h)$ and $Q_p(h)$ as functions of tree height, h, trees can only function in the region $Q_p > Q_0$ (the green-colored region of Figure 5-2 B). If we then plot a curve specific to a given environment, $Q_e(h)$, we can determine which curve, $Q_p(h)$ or $Q_0(h)$, is first intersected by $Q_e(h)$ at lower h. The value of h at this intersection specifies the height of the tallest possible tree. If a tree were to grow larger than this in the given environment, then its evaporative

flow rate would exceed the availability of some resource. In water-rich environments lacking the appropriate incoming energy, Q_e intersects Q_0 before it intersects Q_p, and this determines the maximum tree height. On the other hand, in water-limited environments with ample solar radiation, the reverse is true.

In order to explicitly calculate maximum tree height, we need to relate these various flow rates to tree height by invoking scaling relationships. Reference [292] provides a convenient way to relate height to several other dimensions of trees.

5.2.2 Basal metabolic requirements of a plant (Q_0)

The total basal volume flow rate of internal fluid is well approximated by

$$Q_0 = \beta_1 D^{\eta_1} \rightarrow \beta_2 h^{\eta_2} \tag{5.3}$$

where D is stem diameter, β_1 and β_2 are normalization constants, and η_1 and η_2 are scaling exponents [80, 292]. Empirically, best fits to data give $\eta_1 = 1.8$, $\beta_1 = 0.26$ liter day^{-1} cm$^{-\eta_1}$ with D in cm, $\eta_2 \approx 2.7$ and $\beta_2 \approx 9.2 \times 10^{-7}$ liter day^{-1} cm$^{-\eta_2}$ for h in cm [80] (see Section C.2). In order to convert the empirical relationship in equation C.9, which relates Q_0 to diameter, to a relationship concerning height we employ a calculation which relates various tree dimensions such as height and diameter. For large tree sizes it has been shown in [191] that $h \sim D^{2/3}$ which agrees with our analysis of $\eta_2 = \eta_1 3/2 = 2.7$ (see Section C.2. We rely on an analytic calculation to find β_2 because the model in [191] includes a small tree correction to the basic power law which is outside of the scope of our stated goal. When a direct empirical relationship between two features, such as Q_0 and h, is not known we typically employ an analytic calculation in order to avoid the propagation of error resulting from the combination of two or more empirical relationships. In some situations this is not possible because there are no known analytic derivations. Our overall framework, which is simply the connection of specific set of scaling relationships, does not depend critically on these analytic calculations. For future efforts one can employ our framework and replace any given empirical or analytic scaling relationship with alternative data or calculations.

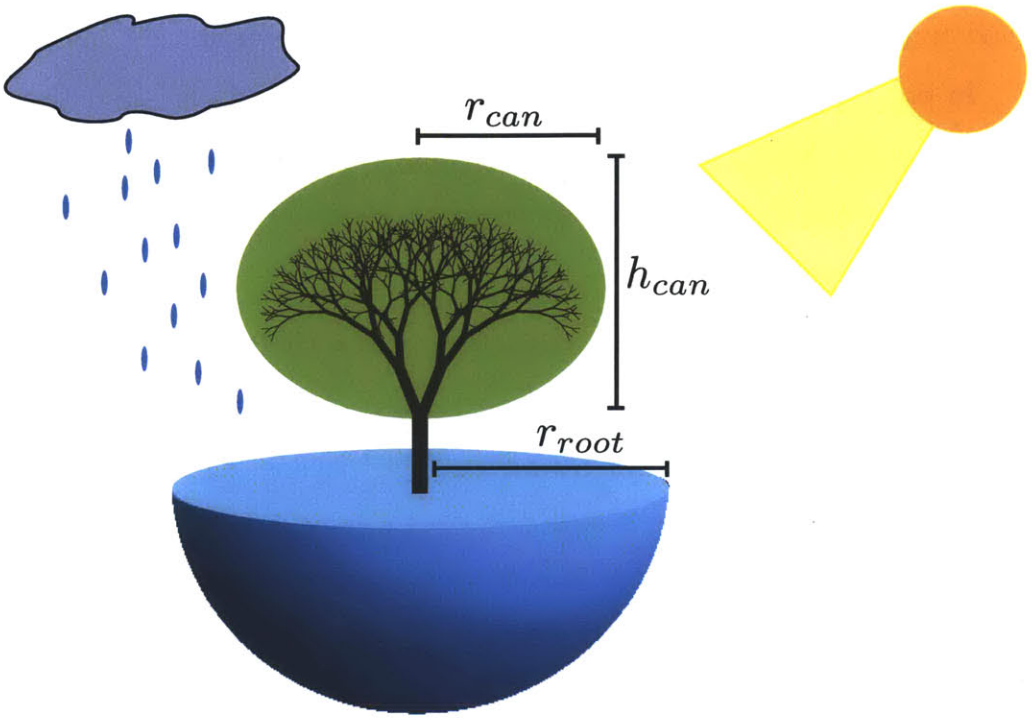


Figure 5-3: The size-based resource gathering capabilities of a tree. The above-ground canopy is shown in green and the below-ground root mass in blue. The essential dimensions of the tree are indicated, where r_{can} is the radius of the canopy, h_{can} is the height of the canopy, and r_{root} is the radius of the root mass. Each of these features scales with height, h, where $r_{can} \propto h^{1.14}$ [83], $h_{can} \propto h$ and $r_{root} \propto h$. The number of leaves scales as h^3 [292]. The scaling of the canopy features determines the collection of solar radiation and the heat exchange with the atmosphere, which can be used to solve for Q_e. The rate of moisture absorption, Q_p, is related to the scaling of the root system and incoming precipitation. Please see Sections C.2 through C.4 of the appendices for a more detailed treatment of these scaling relationships along with derivations for the associated tree physiology.

All that is actually required are the phenomenological scaling relationships themselves which are, or can be, constructed from data (all parameter symbols, definitions, and values can be found in Table S1).

5.2.3 Available flow rate due to precipitation (Q_p)

Given an incoming rate of precipitation, and ignoring hydrology (i.e. water due to runoff, pooling, or subterranean flow and storage), the moisture available to a plant is based on the capture area and capture efficiency of the root system. The capture area for precipitation is defined by the lateral extent of the root system, which can be determined from the geometric properties of the root architecture. From the data and scaling relationships given in [189, 191, 242, 292] the radial extent of roots is approximately given by

$$r_{root} = \beta_3^{1/4} h \tag{5.4}$$

with $\beta_3 = 0.42 \pm 0.02$ (see Section C.3 for a detailed discussion). In our model, trees have access to the total volume of precipitation that falls on the area of flat ground directly above the root system, adjusted by the absorption efficiency of the roots. This can be expressed as

$$Q_p = \gamma\pi \, r_{root}^2 \, p_{inc}, \tag{5.5}$$

where p_{inc} (m year^{-1}) is the rate of precipitation, and γ is the root absorption efficiency.

5.2.4 Evaporative flow rate (Q_e)

Trees act as passive solar pumps with the rate of water escaping due to evaporation equal to the internal flow rate. Hence, Q_e is governed by incoming energy. The basic physiological responses of tree canopies to local meteorology are well-established and are typically summarized using an energy budget [51, 173]. Although an energy budget formulation, which represents the overall conservation of energy, is conceptually simple, each individual energy flux requires a careful calculation based on the physics

relevant to the appropriate tree characteristics, such as the density of stomata on a leaf and the geometry of the canopy. In Section C.4 we provide details of these calculations which include considerations of both the tree size and environmental dependence of evaporation, radiation and conductance in the leaf and canopy microclimate. These are all governed by well-known physical laws, such as the Stefan-Boltzmann law for radiation, whose parameters have been measured or, in the few cases where they are not known, can be derived within our framework.

The basic energy budget requires that the total radiation absorption rate of a canopy, R_{abs}, is the sum of the rates of emitted thermal radiation and the sensible and evaporative heat losses:

$$R_{abs} = La_g + Ha_j + \lambda E a_f. \tag{5.6}$$

Here, L, H, and λE are energy fluxes (W m^{-2}): L is the emitted thermal radiation, H the sensible heat loss, and λE the latent heat loss with λ being the latent heat of vaporization for water and E the evaporative molar flux (mol m^{-2} s^{-1}) [51, 173]. The coefficients a_g, a_j, a_f are effective areas (m^2) over which each heat flux occurrs and are determined by considering how the canopy architecture affects the degree to which each flux is coupled to the atmosphere.

In terms of the molar mass, μ_w (kg mol^{-1}), and density, ρ_w (kg m^{-3}), of water, the evaporative flow rate is related to E by

$$Q_e = a_f E \mu_w / \rho_w. \tag{5.7}$$

From Eq. 5.6, we observe that the dependence of E, and therefore Q_e, on tree height arises entirely from R_{abs} and the effective areas, $a_{f,g,j}$, since H, L, and λ depend only on meteorological conditions. Thus, we can write

$$Q_e = f\left(R_{abs}\left(h\right), a_{f,g,j}\left(h\right), \{m\}\right), \tag{5.8}$$

where $\{m\}$ represents the set of meteorological variables.

Each effective area for heat flux has a linear dependence on the total one-sided leaf area of the canopy, a_L, where $a_L \propto h^3$. The height dependence of R_{abs} can be determined by noting that

$$R_{abs} = \alpha_{can} P_{can} R_{inc}, \tag{5.9}$$

where R_{inc} (W m^{-2}) is the incoming radiation per unit area (normal to the ground), α_{can} is the absorption coefficient for the canopy, and P_{can} is the projected area of the canopy. Both α_{can} and P_{can} depend on tree height via the shape of the canopy and the number and distribution of leaves within that canopy. For a given incoming radiation, $R_{abs} \sim h^3$ for large trees, whereas, for smaller trees, a more complex, but derivable, relation holds (please note that capital the "R" notation refers to absorbed radiation and should not be confused with lower-case "r" which refers to root or canopy radii).

For the average tree whose features are encapsulated in the scaling relationships, these derivations have predictive power beyond determining maximum height. For example, our model predicts the specific form of the decrease in canopy albedo with increasing tree height in excellent agreement with data, as illustrated in Figure 5-4 (please see Section C.4.2 for a derivation). Albedo plays a critical role in many questions related to the earth system and our model framework provides a quantitative means for linking albedo to tree heights and thereby to local resources.

Because evaporation depends on many contributing meteorological variables (solar radiation, air temperature, relative humidity, and wind-speed) and on multiple tree traits (such as average leaf size and stomatal density) it is not possible to write a simple scaling relationship for the evaporative flow rate, $Q_e(h)$. In determining Q_e we picked representative values for tree features that entered into the calculation and used the same values across all locations. (A detailed treatment of Q_e along with the parameter values used can be found in Section C.4 and Table C.1.)

5.2.5 Predicting maximum tree height and other traits

To determine maximum tree heights across the continental United States, we combined meteorological data sets (see Section 5.4) to calculate the functions $Q_e(h)$ and

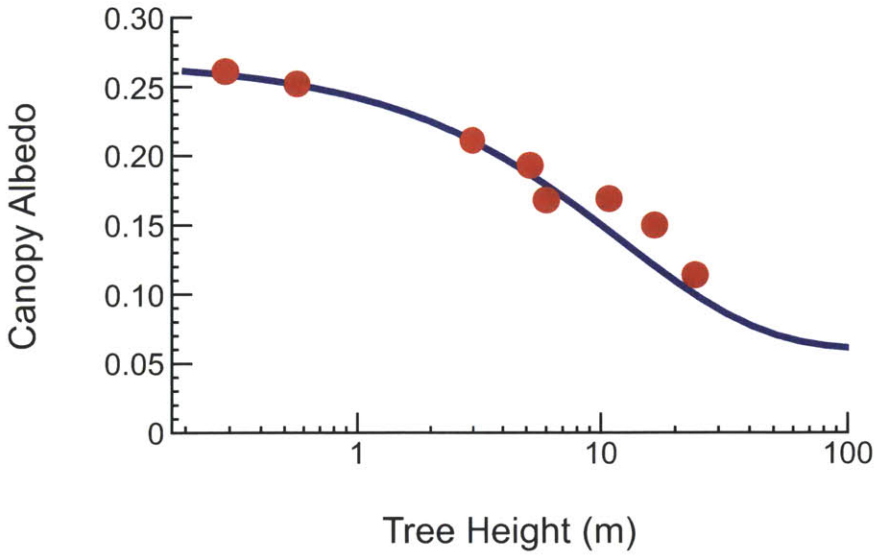


Figure 5-4: The relationship between tree height and the total albedo for the canopy of a single tree. The red points are data [265], and the blue curve is our generalized model for a tree using a soil reflection coefficient of .27 and a deep canopy reflection coefficient of .06 [173] (Section C.4.2). We have not included error bars here because ref. [265] does not provide them for every point.

$Q_p(h)$ for the conditions at each location with $Q_0(h)$ determined from Eq. C.9. As discussed above, our predictions for maximum tree height are found from the first intersection of Q_e with either Q_0 or Q_p. We find that Q_e scales similarly to Q_0 (Figure C-1) and that, in practice, the best predictions are achieved by searching for intersections of Q_e with Q_p once the root absorption efficiency, γ, has been calibrated (see Section C.3).

Because tree height spans nearly two orders of magnitude, we used the relative error, $|h_{obs} - h_{pred}| \,/\, |h_{obs}|$, to compare our predictions, h_{pred}, with observations, h_{obs}, of maximum tree height. As can be seen from the figures, our model gives good agreement with observed maximum tree heights, suggesting that it does indeed capture the essential features of environmental constraints and tree physiology. Figure 5-1 shows a histogram of the relative error prior to taking absolute values $((h_{obs} - h_{pred}) \,/ h_{obs})$ making it possible to determine over- and under-prediction. Error values are relatively narrowly distributed and the center of the distribution is close to zero. (Please see Section C.5.2 and Figure C-2 for a discussion of the slight bimodal nature of this

distribution.)

We tend to *over*-predict maximum tree height in wet environments where there are likely competitive factors limiting tree height. Under-prediction in our model generally occurs in arid environments where trees likely have developed specialized traits which deviate from the average values we used. However, with different, more realistic trait values, such as lower stomatal density in arid environments, we find that these trees obey Eq. C.4. This is to be expected as different trait values are better suited to different environments. We can expand our framework by allowing traits to vary in order to optimize maximum height while still obeying Eq. C.4. For example, holding all other tree parameters constant we can find the stomatal density which maximizes the upper bound on tree height in a given environment. We observe in Figure 5-5 that the optimal stomatal density that we calculate decreases with increasing average annual temperature consistent with observations [21]. We also calculated the optimal leaf size in a similar fashion and found it to decrease with increasing temperature (not shown), which is also a trend suggested by observations [237]. This type of analysis, where the model is used as a point of departure for including sub-dominant effects, including the covariation of other traits, is an important area of investigation. In Section C.5.5 we conduct a similar analysis to determine the optimal allometric scaling of two plant features which we initially took to be constant, the stomatal density and root absorption efficiency. We show that incorporating these additional scaling relationships into our model can reduce the error between predictions and observations (Figure C-5). Understanding the covariation and co-optimization of various plant scalings is an important area of ongoing [217, 216, 215] and future research.

Finally, we explore the effects of environmental shifts on maximum tree heights while holding plant traits constant. Applying the simplest case of a uniform change in mean annual temperature across the United States of ± 2 °C we can solve for the maximum height in that environment. We chose this value because $+2$ °C compares well with the conservative projections for temperature change over the next 100 years according to the frequently cited scenarios summarized by the Intergovernmental Panel on Climate Change (e.g. [138]). We find that for $+2$ °C the average maximum

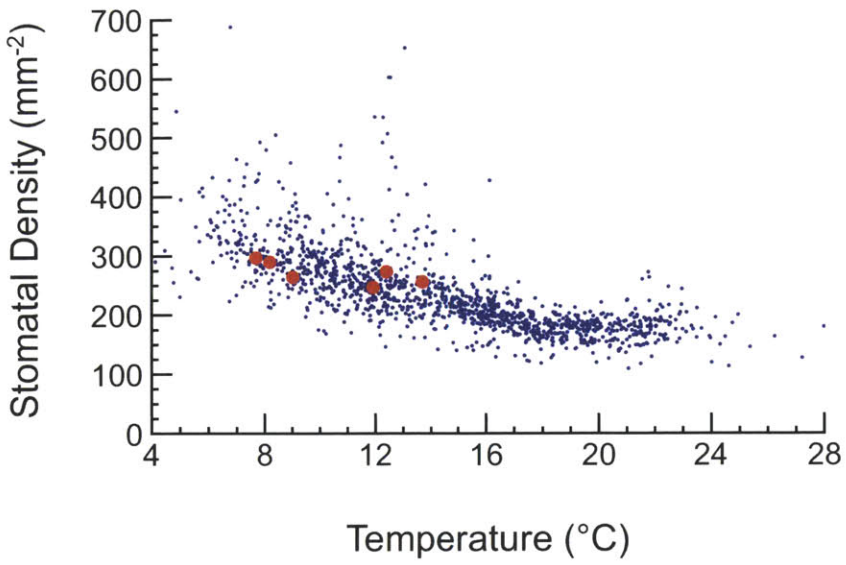


Figure 5-5: The change in the stomatal density as a function of environmental temperature. The values presented are averaged over both sides of the leaf. The blue points are predictions from our model for the optimal stomatal density in each environment, holding all other traits at the average value. The red points are observations from ref. [21]. The observations fall within the range predicted by the model.

height across the continental U.S. decreases by 11% while for -2 °C the average maximum height increases by 13% (Figure C-3).

5.3 Discussion and Conclusions

It is noteworthy that our framework, which uses a general morphology and an average set of tree parameters, can consistently predict maximum tree height over a wide range of environments and tree species. At the same time, it can be easily extended to explore the specific resource tradeoffs associated with each tree trait, and thus predict environment-dependent adaptation. Various plant traits such as stomatal density and leaf size and shape have been suggested as proxies for reconstructing the paleoclimate [21, 237]. Yet some of these traits depend on multiple climatic factors. For example, stomatal density decreases with both increasing temperature and atmospheric CO_2 concentrations [21]. Accurate reconstruction of either temperature or CO_2 concentrations requires disentangling how each factor independently contributes

to stomatal density. Our model provides simple mechanisms for interpreting how single plant traits are suited for different meteorological conditions and with this we can predict optimal plant traits for a given environment. Future work that incorporates the covariation of multiple traits may give insight into both paleo-records and the observed modern geographic variation of plant traits.

Equally important for interpreting the paleo-world is the use of allometry to reconstruct the form of paleoflora where, for example, fossilized tree trunks have been suggested as a means for reconstructing tree height [188]. Because our model makes an explicit and simple connection between local meteorology and tree size this may open up the possibility of supplementing existing proxies with trunk diameters in order to reconstruct both paleoclimate and the structure of local flora.

With respect to present day, our model can be used to anticipate potential changes in maximum tree height as a result of changes in meteorology. As maximum height is connected to local demographics and standing biomass [81, 80, 83, 82] our model may be extended to comment on how changing climate would affect these important forest features.

In short, our model has important implications for understanding tree distributions and dynamics in forests from a resource perspective and presents the possibility for understanding relationships between both paleo and modern climates and dynamic ecology. As such, it has the potential to inform important environmental issues such as migration, climate change, and carbon sequestration.

5.4 Materials and Methods

Scaling Laws For the empirical scaling laws used in this paper we have presented the error associated with scaling constants and exponents when the original reference provided this information.

Height and Meteorological Data For observed maximum tree heights we used the United States Forest Service's Forest Inventory and Analysis (FIA) database,

which records the height and location of individual trees [259, 122]. We are interested in predicting the largest tree in an area given local meteorology. The spatial variation of meteorology can be significant over relatively short distances. Thus, it is important to pair tree sites to meteorological stations which are geographically close to one another. This ensures that the predictions are capturing the conditions experienced by the observed trees. We paired trees with meteorological stations from the National Climatic Data Center (NCDC) [183] for purposes of using station or station interpolate data. Tree-meteorology pairs were separated by no more than 100 m of elevation and 4 km of radial distance. As a result of these stringent criteria we were only able to use a small subset of trees from the FIA database.

We considered all meteorology in terms of long-term annual averages. For precipitation we used the Parameter-elevation Regression on Independent Slopes Model (PRISM) [66, 218] 30-year average (1971-2000) sampled at the location of the meteorological stations. We constructed mean temperatures for individual stations using data from the NCDC [183]. We calculated relative humidity from the PRISM 30-year average [66, 218] for mean dewpoint temperature, minimum temperature and maximum temperature using a method described in ref. [51]. For wind speeds we used data from the National Centers for Environmental Prediction (NCEP) reanalysis [127, 184]. Solar radiation data was obtained from the NREL national grid [64].

Tree Traits Because of our focus on size and its relationship to survival in an environment we chose a single set of plant traits representative of a wide variety of tree species from different environments. This single set of traits was used across all environments to calculate Q_e. For each tree trait we examined the variation across many species, plant sizes, and environments and picked values that were representative of that variation. For several traits we checked that our values compared well to averages from the TRY database [67] which is a comprehensive collection of 65 trait databases and is representative of a large number of species and geographical regions. We picked traits that were appropriate for both angiosperms and gymnosperms, as our model does not distinguish between the two. We checked that the traits we picked

gave rise to related properties, such as LAI, that were similar to observations for both angiosperms and gymnosperms. The values which we used can be found in Table S1 along with means from the TRY database.

Sensitivity analysis It should be noted that many empirical scaling exponents come with associated error bounds and that these exponents can vary across different environments (e.g. [217]). In addition, we have relied on several analytic derivations to inform some of the scaled tree physiology. To deal with the potential variation of exponents we have carried out a basic sensitivity analysis where we perturbed individual exponents away from the value used for our predictions and examined the shift in the median relative error between observations and the new predictions (see Section C.5.4 and Figure C-4 for details).

Chapter 6

The evolutionary consequences of the structure of gene encoding

Preface

An important focus of this thesis is that structure controls function at multiple levels of organization. In this Chapter we turn our attention to the fundamental level of genetic structure where we show how the specific encoding of a gene has the ability to confer implicit functions such as controlling the mutation rate of organisms.

The work in this chapter began with conversations that Ben Allen, a fellow graduate student, and myself had while taking Martin Novak's evolutionary dynamics course at Harvard. I was interested in how mutation rates have differing benefits depending on the nature of the environment: in static environments with a clearly dominant strain low mutation rates are useful for maintaining the best genotype, but in situations with large environmental shifts organisms may need to mutate quickly to escape extinction. I was interested in how organisms manage this benefit and Ben was interested in the possibility of organisms which switch between different mutation rates as an adaptive strategy. Out of these conversations came a collaboration and two group projects which I spearheaded and wrote-up (I was officially taking the class while Ben was sitting in). I proposed the key analyses for searching the parameter space for interesting tradeoffs and evolutionary success.

I did an early mathematical analysis which Ben later advanced into its current state. Our collaboration has been very close, and most of this chapter appears in Ben Allen's thesis from BU [7] where it is cited as co-work with myself. Scott Wylie has been working on similar problems and following conversations with him in the last year we added him as a collaborator (at the time he was a post-doc at Harvard). These conversations have led to the bioinformatic analyses where Scott and I have taken turns in investigating different possibilities and hypotheses using various techniques. Scott conceived of using synonymous changes as a means to detect positive evolutionary selection. I performed the first analyses on the synonymous changes observed in *E. coli*, Scott added a consideration of the phylogenetic tree using the PAML software, and I have analyzed the final output from his investigations.

This chapter represents a mature manuscript entitled "The evolutionary advantage of variable evolvability: Mutator production as an adaptive strategy in bacteria" by B. Allen, C.P. Kempes, and S. Wylie which we plan to submit shortly. B. Allen is the corresponding author.

Abstract

Populations of *Escherichia coli* and other bacteria often contain strains with point mutation rate several orders of magnitude higher than the rest of the population. These "mutator" strains face the negative effects of increased production of deleterious mutations, increased mutational load, but have greater potential for adaptive mutations than their wild-type (low-mutation rate) counterparts. We investigate the hypothesis that evolution may select for the occasional production of mutator offspring, a strategy that we call "mutageneration". We first explore the empirical evidence for this hypothesis where we find that *E. coli* demonstrates a strategy where a strain can switch between the high and low mutation rates. Using a bioinformatic analysis we show that this strategy is a conserved trait across diverse strains and species and is likely under positive selection. We next introduce a simple model of evolution on a fitness landscape with beneficial and deleterious mutations. On this landscape, we consider two competing strains: mutagenerators, who produce mutator offspring with a certain fixed probability, and nonmutagenerators, whose offspring maintain the wild-type mutation rate. Through mathematical analysis and simulation, we investigate the question of which strain is better able to acquire and maintain beneficial mutations, using parameter values derived from empirical results relevant to *E. coli*. We find that mutagenerator strains have a small but persistent advantage in adapting to the fitness landscape, despite their increased mutational load. We show how the extent of this advantage varies with the availability and strength of new adaptive mutations, and with the rate of mutator production in mutagenerators.

6.1 Introduction

The evolution of mutation rates has emerged as an important topic in evolutionary biology. Competitive organisms must manage a tradeoff between acquiring beneficial mutations and avoiding mutational load, the accumulation of deleterious mutations which make the strain less fit. The need to maintain a stable physiology sets an apparent upper bound on mutation rates [75]. However, in novel or fluctuating environments, strains with greatly increased mutation rate can out-compete their wild-type counterparts. The appearance of these "mutator" strains is common (e.g. [144, 164, 199, 200, 206, 298]) and the advantage of mutator strains arises from their increased capacity to acquire beneficial mutations, upon which the mutator allele may "hitchhike" to high abundance [146, 147, 125, 222, 136, 76, 9, 261]. This indirect action of natural selection on mutation rates—by way of its action on the mutations produced at these rates—is known as second-order selection [274].

Second-order selection can be observed, for example, in *Escherichia coli* and other bacteria. Populations of such bacteria often contain mutator strains with mutation rates 2-4 orders of magnitude above those of their wild-type counterparts [144, 164, 199]. These mutator strains play an important role in diverse biological settings ranging from human health [200] to biogeography [206]. For example it has been shown that in patients with cystic fibrosis a large fraction (36 %) of lung infections are dominated by a mutator strain of *Pseudomonas aeruginosa* [200]. For *P. aeruginosa* it has also been shown that these mutator strains have a decreased susceptibility to antibiotics [298]. It has also been hypothesized that mutator strains have allowed the ocean microbe *Prochlorococcus* to rapidly streamline its genome for increased survival in low nutrient environments [206].

Mutator strains can arise due to mutations on genes involved in the methyl-directed mismatch repair (MMR) system or other aspects of DNA replication and maintenance. Mutator strains can be favored in novel or rapidly changing environments, as experiments [55, 162, 263, 262, 108, 252, 300], computer simulations [272, 275], and mathematical analysis [125, 303, 72] have shown. The evolutionary trade-offs involved in the competition between mutators and wild-types have been reviewed by [262], [109], [274], [71], and [261].

Given the importance of mutation we might consider that it is beneficial for organisms to evolve any number of simple or advanced strategies for managing mutation rate. For example mutations producing various mutator phenotypes may be the consequence of unavoidable random errors from which the organisms with single best mutation rate may be selected. Another possibility is loci that regulate the mutation rate itself have an increased susceptibility to genetic alterations [234]. In the latter case, the capacity to produce mutator offspring may itself be considered a selectable trait, which we term "mutageneration". Selection for or against mutageneration would represent a form of "third-order selection" [253]—selection acting on mechanisms for varying the mutation rate. The idea of mutageneration arises frequently in connection to the question of whether amplified mutation rates are a programmed response to stress [26, 98, 236, 102]. However, it is also possible to consider mutagen-

eration, even in low-stress environments, as an evolutionary strategy for discovering beneficial mutations.

The proposed selectable trait of mutageneration consists of two components: a mechanism for producing mutator offspring at an elevated rate, and a pathway for eventual restoration of the wild-type mutation rate [206]. Fixation of this trait could occur over one or more cycles in which:

(a) mutator offspring are produced

(b) a subpopulation of mutators acquires beneficial mutations and hitchhikes to high abundance

(c) a subpopulation of these abundant mutators regains the wild-type mutation rate

(d) this subpopulation rises to high abundance due to its decreased rate of detrimental mutations

Step (b) of this cycle is more likely to occur in novel or fluctuating environments, whereas step (d) would be expected after the environment has stabilized and the population is well-adapted. For this strategy to evolve, its essential features—the genetic capacity for elevated production of mutator alleles and for subsequent wild-type rate restoration—must be maintained through all phases and iterations of this cycle. Step (c) is important because if an environment becomes static a low mutation rate strain can out-compete a high mutation rate strain due to its increased mutational load and this may mean the loss of the mutagenerator strategy if the wild-type rate is not restored.

Previous work suggests that simple sequence repeats (SSRs) in MMR genes (e.g. *mutL*, *mutH*, and *mutS*) and other genes may provide a genetic basis for this trait in *E. coli*, *Salmonella typhimurium* and other bacteria. SSRs are short sequences of base pairs repeated consecutively within a DNA sequence. SSRs are vulnerable to mutation due to slipped-strand mis-pairing, which increases or decreases the number of sequence repeats by one [152, 52]. Such changes in repeat number in certain genes

can produce the mutator phenotype (step (a) above). The restoration of the wild-type mutation rate (step (c)) can occur either through direct reversal of the change in repeat number, or through recombination.

Evidence supporting this potential mechanism comes from a long-term experiment in which 12 *E. coli* populations evolved independently [148, 149, 19]. Mutator alleles fixed in three of these populations within 10,000 generations [263], and in a fourth by 27,000 generations [19]. Two of these mutator alleles were caused by changes in repeat number of a particular sequence at the same position in *mutL*. The ancestor strain REL606 contains, starting from position 210, the sequence GAG CTG GCG CTG GCG CTG GCG, coding for the amino acid sequence ELALALA. This sequence contains three repeated copies of the motif CTG GCG, and also three copies of the motif G CTG GC (obtained by shifting the frame by one base pair). The observed mutator strains were caused by a change from three to four copies in one case, and a change from three to two copies in another. Of the remaining two mutator strains, one was caused by an increase in the number of a mononucleotide repeat in *mutT*, from six Cs to seven [19]. The other was caused by a single base pair insertion in *mutS*, in which no SSRs were involved [252].

Similarly, in archival strains of *S. typhimurium* that had been stocked in sealed agar stabs at room temperature for several decades a mutator strain (LT7) grew to dominance [156]. This strain possessed a change in repeat number of the motif G CTG GC in *mutL* at a homologous location to the *E. coli* genome. Moreover, [112] and [59] both observed cases in which the the wild-type mutation rate was regained via a further slipped-strand mis-pairing that restored the original repeat number, and this new wild-type strain rose to high abundance (steps (c) and (d) of the mutagenerator cycle).

The threefold repetition of the G CTG GC motif important in each of these contexts has been shown to be conserved across strains of *E. coli* and species of Salmonella [58]. More generally, numerous other SSRs have been found in *mutL*, and *mutS* [58], and it has been shown that mononucleotide repeats in the *mutT* gene of *E. coli* K12 are more abundant than what would be expected from a randomized gene

sequence [234]. Though this evidence cannot be regarded as conclusive, it suggests that evolution may have selected for SSRs in MMR genes. These SSRs provide a mechanism for the production of mutator offspring at an elevated rate. They also provide a mechanism for restoration of the wild-type mutation rate, by reversal of the slipped-strand mis-pairing that produced the mutator allele (as was observed in the *S. typhimurium* studies described above).

It has also been proposed [262, 42, 109, 71] that mutators may regain MMR functionality through recombination with wild-types. MMR-deficient mutators have, compared to other mutator alleles, a high rate of recombination [226, 70]. Moreover, the MMR gene *mutS* is itself a frequent target of recombination [70, 42]. It is also worth noting that MMR mutators are observed more commonly than other heritable mutator phenotypes. [71] suggest that evolution may have selected for MMR-deficient mutators over other mutator alleles, in part because of the increased probability (relative to other mutator alleles) of wild-type rate restoration through recombination.

The previous work reviewed here suggests that there is likely a "switch" in the mutL gene producing high and low mutation strains, and we interpret other work to suggest that this strategy may be under positive selection. However it has not yet been explicitly shown that mutageneration is a strategy which is being positively selected for, nor has it been understood why and in what contexts this trait has an evolutionary advantage. Here we use bioinformatics to concretely demonstrate the positive selection of this trait across diverse strains and species, and we us a mathematical model to illustrate that mutageneration is a beneficial strategy for a well-defined set of biological and environmental contexts. Our bioinformatic analysis compares the expected likelihood of disrupting the repeated codon motif (responsible for the mutation rate "switch") to what is observed across divergent strains. We find that this genetic structure is likely under positive selection. Our mathematical model considers two competing types: mutagenerators, who produce mutator offspring at a fixed stochastic rate, and nonmutagenerators, who do not produce mutators. These types evolve and compete on a fitness landscape which includes both beneficial and deleterious mutations. Using this model, we analyze the question of which type is

favored to achieve the first beneficial mutation that survives stochastic drift. This corresponds to step (a) and the beginning of step (b) in the process outlined above. We find that mutagenerators have a small but persistent advantage in this contest, despite their increased mutational load. All parameter values in our model correspond to empirically verified quantities regarding *E. coli* evolution. Our model also predicts which parameter alterations would make mutageneration a detrimental strategy and this provides avenues for experimentally testing our theory.

6.2 Bioinformatics analysis of mutagenerator selection.

Our basic premise for the mutagenerator strategy is that there is an implicit function [52] associated with the encoding of the mutL gene. The 6 nucleotide string which is repeated 3 times (referred to here as the 3×6 repeat) allows the gene to easily add or drop the repeated motif and in doing so either disrupt or restore the mismatch repair function of mutL. This is an unusual structure (e.g. *E. coli* K-12 has only three genes with a 3×6 repeat) which leads to a special mutational property. If this property is beneficial then we should see selection for this specific genetic encoding opposed to synonymous encodings of the same amino acid sequence in mutL. A simple test of this is to consider the synonymous changes for all codons in mutL between various strains of *E. coli* and determine what the likelihood is that that the 3×6 repeat is preserved across all strains compared to what is observed. Comparing every strain of *E. coli* from BLAST, for which *mutL* has been annotated, we find that of 40 genome sequences only 1 does not contain the 3×6 repeat. This is interesting because it demonstrates that losing the repeat motif is possible for a viable organism, but rare. We compared each of these strains to K-12 as a reference sequence and calculated the synonymous change rate over all of *mutL*. We then applied this synonymous change rate to each sequence and asked how likely it would be for the 3×6 repeat to be preserved. We found that across all species the average likelihood that the repeat

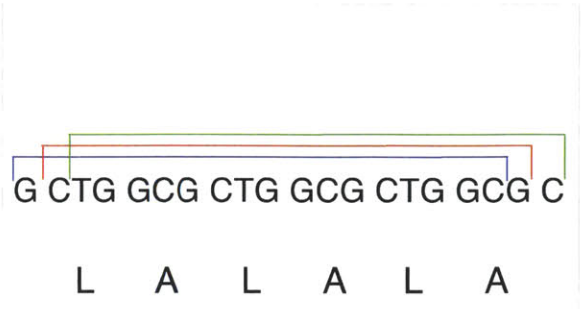


Figure 6-1: The three windows with a 3×6 repeat along with the amino acid sequence represented by these codons.

would be lost is $78\% \pm 13\%$ compared to the observed value of 2.5%. This simple analysis shows that the motif is lost in genomes much more infrequently than random synonymous changes would suggest.

However this analysis is slightly flawed in that it does not account for the relatedness of the strains involved. A simple example of why this could be problematic is to consider a situation where a mutation occurred in an ancestor and was propagated to all descendants, in which case analysis of how frequently a change occurred would be much larger if this was not taken into consideration. In a more advanced analysis we have first constructed a phylogenetic tree of all of the strains and increased our search to include all genomes with the specific 3×6 repeat which includes species other than *E. coli* (e.g. *Salmonella typhimurium*). We used the Phylogenetic Analysis by Maximum Likelihood (PAML) software to both construct phylogenetic trees and to calculate the number of synonymous changes that occurred over the repeat region. PAML's analysis of synonymous change likelihoods takes into consideration many detailed features of the genome such as codon bias which is the observation that in genomes certain synonymous codons are more likely to be used than others.

Using PAML we first calculate the expected number of synonymous changes that should occur in the repeat region given its length and the synonymous change rate of codons over the entire mutL gene. There are three windows Figure 6-1 (each separated

by a distance of one nucleotide) which contain a 3×6 repeat. Using PAML, we find that for the middle repeat window we would expect to have 25 synonymous changes across 55 genomes, while we calculate the actual number of synonymous changes in this region to be 15. This is much lower than expected and if we take synonymous changes to be an independent random process we can use the Poisson distribution to calculate the probability that we would see 15 changes or less given the expected number. We find that the p-value for this situation is $p = .03$ demonstrating that it is unlikely that the repeat is conserved only by chance. It is likely that the specific structure of these codons is being conserved due to the related evolutionary function that the repeat confers.

For the other two regions we find $p = 3.01 \times 10^{-7}$ (the expected number of synonymous changes is 33 while the observed number is 8) and $p = .73$ respectively. It is interesting that each window has a varying degree to which it is conserved across all of the genomes. The first window is incredibly conserved while the last is likely not being conserved. From the perspective of the mutational switch it is only necessary to have a single repeat and thus it is not clear how or why multiple repeat windows would be conserved to varying degrees.

6.3 Mathematical model development

Our bioinformatic analysis confirms that the 3×6 repeat motif in mutL is likely under positive selection along with its function of a mutation rate "switch". However we still do not understand why mutageneration is an effective strategy nor under what conditions it will be successful. Here we develop a model which analyzes competition between mutagenerators and strains with a single fixed mutation rate. We look at the likelihood that a strain achieves the first beneficial mutation. Focusing on this question avoids the combinatorial difficulties posed by modeling multiple mutations and clonal interference (the long-term competition between strains with different positive mutations).

Fitness landscape In order to highlight the evolutionary tradeoffs of mutator production within an analytically tractable framework, we consider a simple fitness landscape model with beneficial and deleterious mutations, illustrated in Figure 6-2. This model has the advantage that each parameter value corresponds to an empirically evaluated quantity.

In this fitness landscape model, genotypes are classified as neutral, advantageous, or disadvantageous. Each mutation from a neutral genotype results in an advantageous genotype with probability b, a disadvantageous genotype with probability d, and otherwise another neutral genotype. Each mutation from an advantageous genotype results in a disadvantageous genotype with probability d and otherwise another advantageous genotype of the same fitness. (Since our focus is on which type sustains the first beneficial mutation, we do not consider beneficial mutations beyond the first.) Mutations from disadvantageous genotypes result in other disadvantageous genotypes. The implicit assumption that deleterious mutations cannot be reverted or compensated for by evolution at other loci is reasonable for large genome sizes and relatively stable environments [196].

Disadvantageous genotypes in our model have fitness $1 - s_d$. The assumption that all disadvantageous genotypes have the same fitness has little effect on our results: since disadvantageous genotypes are evolutionary dead ends in our model, their fitness is relevant only to the question of determining their abundance in mutation-selection equilibrium.

Advantageous genotypes have fitness $1 + s_b$, where the fitness effect s_b is sampled independently, each time a neutral-to-advantageous mutation occurs, from an exponential distribution of mean $\langle s_b \rangle = \beta$. The choice of an exponential distribution is standard, motivated by results in extreme value theory [203].

Mutagenerators and nonmutagenerators On this fitness landscape, we consider a population of fixed size $N \gg 1$, divided into mutagenerator and nonmutagenerator strains. The mutagenerator strain is itself subdivided into wild-types (standard mutation rate) and mutators.

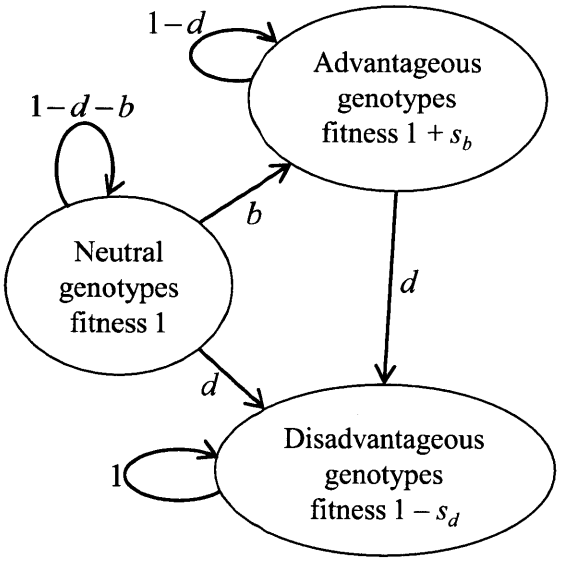


Figure 6-2: The fitness landscape model. The arrows between genotype categories are labeled with transition probabilities per genomic mutation. For example, if a single mutation occurs in the replication of a neutral genotype, the result is an advantageous genotype with probability b, a disadvantageous genotype with probability d, and otherwise another neutral genotype.

Evolution proceeds according to an asexual Wright-Fisher process. Each new offspring is born with a random number of mutations, u. This number is sampled from a Poisson distribution, with mean u_w for offspring of nonmutagenerators or wild-type mutagenerators, and mean u_m for offspring of mutators. Each of these mutations is randomly designated as neutral, beneficial, or deleterious, and the resulting offspring is then designated as neutral, advantageous, or disadvantageous, according to the fitness landscape model described above.

We represent mutageneration by assuming that each offspring of a wild-type mutagenerator has probability μ of being born a mutator (and is otherwise another wild-type mutagenerator). Nonmutagenerators only produce nonmutagenerator offspring.

We do not incorporate restoration of the wild-type mutation rate in our model. Although such restoration is an essential component of the proposed mutagenerator trait, its mechanisms and rates are not yet sufficiently well-understood to be modeled [262, 71, 59]. In any case, the focus of our model is on the acquisition of the first

beneficial mutation, and the restoration of wild-type rates is unlikely to be relevant to this question.

6.3.1 Success criterion

We quantify evolutionary success in terms of probabilities of achieving adaptations. We use the term "adaptation" as shorthand for "beneficial mutation that survives stochastic drift". We focus on adaptations, rather than fixation or other measures of evolutionary success, in order to avoid the complications that arise when multiple adaptations occur in a single lineage, or in competing lineages (clonal interference).

Specifically, we consider the case of an initially rare mutagenerator strain invading a population of nonmutagenerators. The initial frequency of mutagenerators, $x_g(0)$, should satisfy $1 \ll Nx_g(0) \ll N$.

We quantify the success of mutagenerators in terms of the probability π_g that they acquire an adaptation before nonmutagenerators do. We consider mutagenerators to be favored if this first adaptation probability exceeds their initial frequency, $\pi_g > x_g(0)$. This criterion is derived from the general fact that, in an evolving population, a neutral substrain (identical to the rest of the population in fitness and mutation rate) has first adaptation probability equal to its initial frequency. Thus $\pi_g > x_g(0)$ indicates that mutagenerators have an advantage in adaptation, relative to a neutral invading strain.

6.3.2 Simulations

To gauge the effect of the approximations made in our analysis we performed an exact simulation of our model. In this simulation, the state of the population at each time step is represented by the number of individuals of each of eight types: neutral, advantageous, and disadvantageous nonmutagenerators, neutral and advantageous wild-type mutagenerators, neutral and advantageous mutators, and disadvantageous mutagenerators of either subtype. At each time-step, the population is replaced by an equal number of new individuals (offspring). The number of offspring produced

Table 6.1: Default parameter values used in analysis and simulation

Symbol	Description	Default Value	Reference
N	Population size	2×10^6	
u_w	Wild-type mutagenerator mutation rate*	2.5×10^{-3}	[75]
u_m	Mutator mutation rate*	2.5×10^{-1}	
a	Rate of mutator production	5×10^{-6}	[29]
d	Proportion of mutations that are deleterious	0.08	[133]
s_d	Fitness effect of deleterious mutation	0.012	[133]
b	Proportion of mutations that are beneficial	8×10^{-3}	[207]
$\langle s_b \rangle$	Mean fitness effect of beneficial mutation	0.013	[207]

* Mutations rates are in mutations per replication per genome.

by each parental type is sampled from a multinomial distribution, where each type is assigned a probability proportional to fitness times abundance. Once the number of offspring of each type is determined, the numbers of mutations between types are sampled from appropriate binomial distributions; from these sampled values the new frequencies of the eight types are calculated. Using 2×10^5 trials per parameter combination, we recorded the frequencies with which each type sustained the first beneficial mutation. In general, these simulations agreed closely with the analytical results which we present later.

6.3.3 Parameters

Table 6.1 gives the default parameter values that we use in analytical and simulation results unless otherwise stated. All parameter values are derived from published empirical results. The fraction d of mutations that are deleterious is derived from the finding of [133] that deleterious mutations occur at a rate of 2×10^{-4} per genome replication in wild-type *E. coli*; the equation

$$1 - e^{-u_n d} = 2 \times 10^{-4}$$

yields $d \approx 0.08$. The fraction b of mutations that are beneficial comes from the result of [207] that approximately 2×10^{-5} beneficial mutations occur per individual per generation. Dividing by the overall mutation rate u_n yields the probability $b = 0.008$ that a random mutation is beneficial.

We note, however, that [207] consider only the first 2000 generations as an *E. coli* population is adapting to a novel landscape. According to theory [202] and experiment [149, 19], new beneficial mutations decrease in abundance and/or strength as a population adapts to a landscape over time. Therefore, we consider the values of b and $\langle s_b \rangle$ in table 6.1, adapted from [207], to be upper bounds, representing the initial phase of adaptation. The typical b and $\langle s_b \rangle$ values encountered by *E. coli* in stable environments may be significantly smaller.

6.3.4 Dynamics of neutral and disadvantageous genotypes

We begin our analysis by considering the temporal dynamics of competition between mutagenerators and strains with a single mutation rate. Ultimately we are interested in how these dynamics lead to a probability that a given strain will fixate on the first beneficial mutation and how this probability depends on the key parameters summarized in Table 6.1.

We first consider the case $b = 0$ (no beneficial mutations available). To analyze this case, we use a deterministic approximation to the Wright-Fisher process, which is accurate for large population size. In this approximation, each subtype produces a number of offspring per unit time equal to its abundance times fitness, mutations between subtypes occur in proportion to their probability, and death affects all subtypes equally at a rate equal to average population fitness. The dynamics of subtype frequencies in this approximation can be described by quasispecies equation of [79], and variants thereof.

We observe that for a subtype of mutation rate u, the probability of avoiding deleterious mutation (i.e. the probability that an offspring of a neutral type will again be neutral) is e^{-ud}. This arises from the fact that, in our fitness landscape model, the number of total mutations per reproduction is Poisson-distributed with

expected value u, and therefore the number of deleterious mutations is also Poisson-distributed, with expected value ud. We use the notation $q_w = e^{-u_w d}$ and $q_m = e^{-u_m d}$ to denote these avoidance probabilities for to the wild-type and mutator mutation rates, respectively.

We first compute the internal mutation-selection equilibria of the nonmutagenerator and mutagenerator strains, and then discuss the dynamics of competition between these strains.

Mutation-selection equilibrium for nonmutagenerators

For the nonmutagenerator strain, we describe the dynamics of fixation using a quasi-species equation which is a deterministic approximation commonly used in theoretical biology [197, 78]. The quasi-species formulation considers the likelihood of a strain mutating into another and the relative reproduction of each strain based on its fitness. This can be described as

$$x_i = \sum_{j=0}^{N} x_j f_j q_{j,i} - \phi x_i \tag{6.1}$$

where x_i is the proportion of the population composed of a given strain, $q_{j,i}$ is the probability that strain j mutates into strain i, and f_i is the relative fitness of a given strain. The last term ensures a constant population size where ϕ is the average fitness.

For our system the mutational probabilities are given by

$$\mathbf{q} = \begin{pmatrix} q_w & 1-q_w \\ 0 & 1 \end{pmatrix} \tag{6.2}$$

and fitness by

$$\mathbf{f} = \begin{pmatrix} 1 & 1-s_d \end{pmatrix} \tag{6.3}$$

Thus equation 6.1 becomes

$$\begin{pmatrix} \dot{x}_n^0 \\ \dot{x}_n^- \end{pmatrix} = \begin{pmatrix} q_w & 0 \\ 1-q_w & 1-s_d \end{pmatrix} \begin{pmatrix} x_n^0 \\ x_n^- \end{pmatrix} - \phi \begin{pmatrix} x_n^0 \\ x_n^- \end{pmatrix}. \tag{6.4}$$

Above, x_n^0 and x_n^- denote the frequencies of neutral and disadvantageous genotypes, respectively. $\phi = x_n^0 + (1 - s_d)x_n^-$ denotes the average population fitness (it should again be noted that $x_n^0 + x_n^- = 1$ as these are frequencies).

We now assume that $q_w > 1 - s_d$: the fitness of the neutral genotype multiplied by the probability of remaining in that genotype is greater than the fitness of the disadvantageous genotype multiplied by the likelihood of remaining in that state. If this assumption is violated then an "error catastrophe" [79, 196] occurs, in which deleterious mutations are produced at such a high rate that the neutral genotypes are evolutionarily lost.

The mutation-selection equilibrium in nonmutagenerators is given by the dominant eigenvector of the matrix in (6.4), which under the above assumptions can be calculated as

$$\begin{pmatrix} x_n^0 \\ x_n^- \end{pmatrix} = \begin{pmatrix} \frac{q_w - (1 - s_d)}{s_d} \\ \frac{1 - q_w}{s_d} \end{pmatrix}. \tag{6.5}$$

The corresponding eigenvalue is q_w, which gives the average fitness of the nonmutagenerators at equilibrium.

Mutation-selection equilibrium for mutagenerators We now consider mutagenerators alone. The deterministic approximation in this (again using the framework of equation 6.1) is described by the equation

$$\begin{pmatrix} \dot{x}_w^0 \\ \dot{x}_m^0 \\ \dot{x}_g^- \end{pmatrix} = \begin{pmatrix} q_w(1-\mu) & 0 & 0 \\ q_w\mu & q_m & 0 \\ 1 - q_w & 1 - q_m & 1 - s_d \end{pmatrix} \begin{pmatrix} x_w^0 \\ x_m^0 \\ x_g^- \end{pmatrix} - \phi \begin{pmatrix} x_w^0 \\ x_m^0 \\ x_g^- \end{pmatrix}. \tag{6.6}$$

Here x_w^0, x_m^0, and x_g^- denote, respectively, the frequencies of neutral wild-types, neutral mutators, and disadvantageous genotypes (of either subtype). ϕ again denotes average fitness, which is given in this case by $x_w^0 + x_m^0 + (1 - s_d)x_g^-$.

We now assume $(1 - \mu)q_w > 1 - s_d$ and $(1 - \mu)q_w > q_m$. If the first inequality is violated, an error catastrophe occurs as described above. If the second inequality is

violated, there is a "mutator catastrophe", in which the wild-type mutation rate is lost to mutators.

As above, the mutation-selection balance in mutagenerators is given by the dominant eigenvector of the matrix in (6.6), which under the given assumptions can be calculated as

$$\begin{pmatrix} x_w^0 \\ x_m^0 \\ x_g^- \end{pmatrix} = \begin{pmatrix} \dfrac{\big((1-\mu)q_w - q_m\big)\big((1-\mu)q_w + s_d - 1\big)}{s_d(q_w - q_m)} \\ \dfrac{\mu q_w\big((1-\mu)q_w + s_d - 1\big)}{s_d(q_w - q_m)} \\ \dfrac{1-(1-\mu)q_w}{s_d} \end{pmatrix}. \tag{6.7}$$

The average fitness of mutagenerators at equilibrium is equal to the corresponding eigenvalue, which is $(1-\mu)q_w$.

Dynamics of rare mutagenerators Still assuming no beneficial mutations ($b = 0$), we now suppose that a population dominated by nonmutagenerators is invaded by a strain of mutagenerators. We suppose the initial frequency of mutagenerators, $x_g(0)$ satisfies $1 \ll Nx_g(0) \ll N$. We also suppose that the subtypes of the mutagenerators are initially in mutation-selection equilibrium.

Comparing the average fitnesses of mutagenerators and nonmutagenerators, we observe that mutagenerators have a selective disadvantage, with selection coefficient $-\mu q_w$, due to excess mutational load. Under the deterministic approximation (and using the assumption $1 \ll Nx_g(0) \ll N$), the frequency of the mutagenerator strain declines exponentially through time:

$$x_g(t) = x_g(0)e^{-\mu q_w t}.$$

The deterministic approximation further implies that the relative frequencies of the subtypes within each strain are constant over time. Since the nonmutagenerators are dominant, we approximate their frequency as 1.

This highlights, in the extreme, that mutageneration is not beneficial in the case

where there are extremely rare positive mutations. In this situation mutageneration simply accelerates the rate a which offspring with deleterious mutations are produced.

Probability of surviving stochastic drift We turn now to the case $b > 0$ in which beneficial mutation is possible. When such a mutation appears, it may either grow to sustainable abundance or disappear due to stochastic drift (organisms with fitter genotypes are more likely, but not guaranteed, to survive). We apply the standard result that, for the haploid Wright-Fisher model with large population size, the probability that a beneficial mutation escapes loss due to drift is approximately equal to twice the mutation's selection coefficient s. This approximation is accurate for $1/N \ll s \ll 1$ [115, 134].

The relevant selection coefficient in this case is not merely the fitness increase, s_b, of advantageous genotypes relative to neutral genotypes. This is because the beneficial mutation may be lost in one of two ways—by death or by mutation to a deleterious genotype—and we must take both of these possibilities into account. Thus the relevant selection coefficient is the rate at which advantageous genotypes are faithfully reproduced, divided by the average population fitness, minus one. To compute this, we note that advantageous genotypes are faithfully reproduced at rate $(1+s_b)q$, where q equals q_w, $q_w(1-a)$ or q_m, depending on which strain has achieved this mutation. Since nonmutagenerators are dominant, the average population fitness is q_n. Thus the selection coefficient is $(1+s_b)q/q_w - 1$, with q equal to the appropriate probability of avoiding deleterious mutation.

Recalling that s_b is exponentially distributed with mean β, we compute the probability that a random beneficial mutation will survive stochastic drift as follows:

$$\begin{aligned} \rho &= \frac{2}{\beta}\int_{q_w/q-1}^{\infty}\left((1+s_b)\frac{q}{q_w}-1\right)e^{-s_b/\beta}\,ds_b \\ &= 2\beta\frac{q}{q_w}\exp\left(\frac{q-q_w}{\beta q}\right). \end{aligned} \tag{6.8}$$

Substituting the appropriate values of q and replacing β with $\langle s_b \rangle$, we have

$$\begin{aligned} \rho_n &= 2\langle s_b \rangle \\ \rho_w &= 2\langle s_b \rangle (1-a) \exp\left(\frac{-a}{\langle s_b \rangle (1-a)}\right) \\ \rho_m &= 2\langle s_b \rangle \frac{q_m}{q_w} \exp\left(\frac{q_m - q_w}{\langle s_b \rangle q_m}\right) \end{aligned} \tag{6.9}$$

for nonmutagenerators, mutagenerators with the wild-type and mutator mutation rates respectively. The probability that a beneficial mutation will survive stochastic drift is proportional to the mean strength of beneficial mutations, $\langle s_b \rangle$, for nonmutagenerators, while in mutagenerators this same success depends on the rate at which the mutators are produced, a, and the relative likelihoods of avoiding deleterious genotypes.

Probability of first adaptation We now compute adaptation probabilities. An adaptation at time t is defined here as a beneficial mutation that arises at time t and eventually survives stochastic drift.

For nonmutagenerators, the probability of acquiring an adaptation in the infinitesimal time interval $[t, t+dt)$ equals the number of offspring produced by neutral nonmutagenerators in that time, $Nx_n^0\ dt$, times the probability of being born with an advantageous genotype, $bu_w q_w$, times the probability ρ_n of eventually surviving drift. This instantaneous probability can be written as

$$Nx_n^0\, bu_w q_w\, \rho_n\ dt.$$

To shorten this expression, we introduce the constant

$$\alpha_n = x_n^0\, bu_w q_w\, \rho_n, \tag{6.10}$$

which gives the average per-capita rate of adaptation in nonmutagenerators. Their instantaneous probability of adaptation is then $N\alpha_n\ dt$. The probability that no adaptation has occurred in nonmutagenerators by time t is $e^{-N\alpha_n t}$. Since nonmu-

tagenerators are dominant in the population, this quantity also approximates the probability that neither strain has acquired an adaptation by time t.

For mutagenerators, the instantaneous probability of adaptation in the interval $[t, t+dt)$ is given by

$$Nx_g(t)\left(\frac{x_w^0}{x_g}bu_w q_w \rho_w + \frac{x_m^0}{x_g}bu_m q_m \rho_m\right)\, dt.$$

The two terms above correspond to the adaptations by the two sub-populations of wild-types and mutators, respectively. $x_g(t)$ is the total frequency of mutagenerators, given by $x_g(t) = x_g(0)e^{-aq_w t}$ as described above. The ratios x_w^0/x_g and x_m^0/x_g refer to the relative frequencies of wild-types and mutators, respectively, within the wild-type strain. Introducing the constant

$$\alpha_g = \frac{x_w^0}{x_g}bu_w q_w \rho_w + \frac{x_m^0}{x_g}bu_m q_m \rho_m, \tag{6.11}$$

we can rewrite the instantaneous probability of mutagenerators to a positive adaptation as

$$Nx_g(0)\, e^{-aq_w t}\, \alpha_g\, dt.$$

We are interested in the probability π_g that mutagenerators acquire an adaptation before nonmutagenerators. We obtain this by integrating the instantaneous adaptation probability $Nx_g(0)\, e^{-aq_w t}\, \alpha_g\, dt$ for mutagenerators against the probability $e^{-N\alpha_n t}$ that no adaptation has occurred before this time. This yields

$$\begin{aligned}\pi_g &= \int_0^\infty Nx_g(0)\, e^{-\mu q_w t}\, \alpha_g\, e^{-N\alpha_n t}\, dt \\ &= \frac{Nx_g(0)\alpha_g}{\mu q_w + N\alpha_n}.\end{aligned} \tag{6.12}$$

Recalling our success criterion $\pi_g > x_g(0)$, we arrive at a simple yet powerful rule: mutagenerators have an evolutionary advantage over nonmutagenerators if and only if

$$N\alpha_g - \mu q_w > N\alpha_n. \tag{6.13}$$

This rule can be rewritten as

$$N\alpha_g - \big(1 - (1-\mu)q_w\big) > N\alpha_n - (1 - q_w),$$

so that each side has the form

$$\text{population size} \times \text{per-capita adaptation rate} - \text{mutational load}.$$

This rule makes quantitatively clear that the success of mutation strategies depends on the tradeoff between adaptation rate and mutational load (production of deleterious mutations). The applicability of this rule to general evolvability strategies is the subject of future exploration.

Dependence on strength and availability of beneficial mutations Having derived condition (6.13) for the success of mutagenerators, we now investigate how their success depends on the population size and the properties of beneficial mutations. We first observe that, using (6.10), (6.11), and (6.12), the probability of mutagenerators finding an adaptation before nonmutagenerators can be written as

$$\frac{\pi_g}{x_g(0)} = \frac{Nb\left((x_w^0/x_g)\, u_w q_w \rho_w + (x_m^0/x_g)\, u_m q_m \rho_m\right)}{\mu q_w + Nb\,(x_n^0\, u_w q_w \rho_n)}. \tag{6.14}$$

where we have rearranged the equation to normalize by the initial frequency of the mutagenerator population. In this expression, N and b only appear together as a product. Therefore, for the purposes of determining mutagenerator success, the product Nb can therefore be considered as a single parameter. This product represents the availability of beneficial mutations to the population as a whole.

Using this insight, we explore how the success of mutagenerators depends on the strength, $\langle s_b \rangle$, and availability, Nb, of beneficial mutations (Figure 6-3). Not surprisingly, mutagenerator success increases as beneficial mutations become stronger and more plentiful. For the empirically-derived parameter set (Table 6.1), the probability of mutagenerator adaptive success normalized by initial frequency, $\pi_g/x_g(0)$, equals

approximately 1.0004. This number is greater than one implying that mutagenerators are expected to outcompete nonmutagenerators.

We parenthetically remark that it is not surprising to find values of $\pi_g/x_g(0)$ very close to 1 in Figure 6-3, considering that the mutagenerator population is composed almost entirely of wild-types. Indeed, we observe from (6.7) that the fraction of mutators within the mutagenerators is of order $\mu = 5\times10^{-6}$. Thus the vast majority of the mutagenerator strain has the same mutation rate, u_w, as the nonmutagenerators. It is only the mutator subtype, comprising a small fraction of the mutagenerators, that gives mutagenerators an advantage in adaptation.

The plentiful and scarce regimes From (6.13) and (6.14) we observe that, for $Nb \gg \mu q_w$, the condition for mutagenerator success reduces to $\alpha_g > \alpha_n$, or equivalently,

$$\frac{x_w^0}{x_g} u_w q_w \rho_w + \frac{x_m^0}{x_g} u_m q_m \rho_m > x_n^0 u_w q_w \rho_n.$$

In other words, the load term $-\mu q_w$ in (6.13) becomes dwarfed by the adaptation rates $N\alpha_g$ and $N\alpha_n$, and the factor Nb cancels from both sides. Intuitively, in this regime, adaptation occurs on a faster timescale than decline due to load; thus the probabilities of first adaptation depend only on the instantaneous adaptation rates.

Conversely, for $Nb \ll \mu q_w$, condition (6.13) implies that mutagenerator success is impossible. This is because, in this regime, decline due to load occurs on a faster timescale than adaptation. Adaptation is unlikely to occur before the mutagenerator strain is greatly diminished due to load.

Dependence on the rate of mutator production The rate μ of mutation from wild-type mutagenerator to mutator defines, in a sense, the mutagenerator strategy; it is our hypothesis that this parameter may be acted upon by selection. Figure 6-4 shows that increasing μ amplifies either the success or failure of the mutagenerator type, depending on the fitness landscape.

We can also investigate the effect of increasing μ on the success of the subtypes within the mutagenerator strain. We denote by π_w and π_m the probabilities that

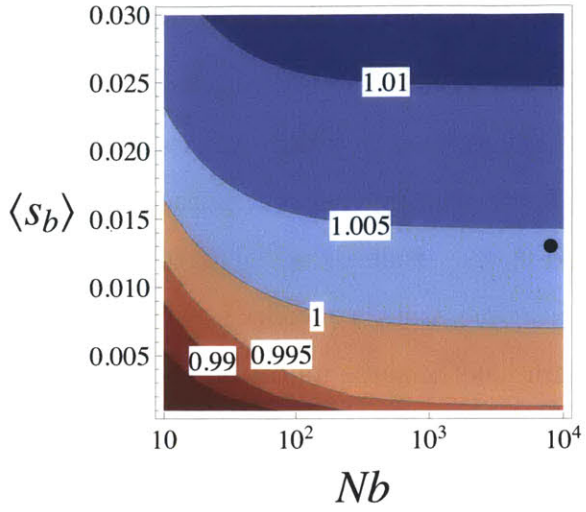


Figure 6-3: Mutagenerator success as a function of the expected strength, $\langle s_b \rangle$, and availability, Nb, of beneficial mutations. Success of mutagenerators is quantified by the ratio $\pi_g/x_g(0) = N\alpha_g/(\mu q_w + N\alpha_n)$, derived in (6.12). A value larger than 1 indicates that mutagenerators are favored to acquire the first adaptation. Mutagenerator success increases with both the strength and availability of beneficial mutations. The point • indicates the empirically-derived parameter values (Table 6.1).

wild-type mutagenerators and mutators, respectively, achieve the first adaptation in the population. The ratios $\pi_w/x_g(0)$ and $\pi_m/x_g(0)$ can be obtained in similar fashion to the derivation of (6.14), yielding

$$\begin{aligned} \frac{\pi_w}{x_g(0)} &= \frac{N\,(x_w^0/x_g)\,bu_w q_w \rho_w}{\mu q_w + N\,x_n^0\,bu_w q_w \rho_n} \\ \frac{\pi_m}{x_g(0)} &= \frac{N\,(x_m^0/x_g)\,bu_m q_m \rho_m}{\mu q_w + N\,x_n^0\,bu_w q_w \rho_n}. \end{aligned} \tag{6.15}$$

Figure 6-5 explores the dependence of these quantities on μ. We observe a tradeoff: for fitness landscapes that favor mutagenerators, their success increases with μ, but this success comes from cases where the mutator subtype adapts. The probability π_w that wild-types acquire the firs adaptation decreases with μ. Since wild-types have greater long-term evolutionary stability, this tradeoff may constrain the upward evolution of μ.

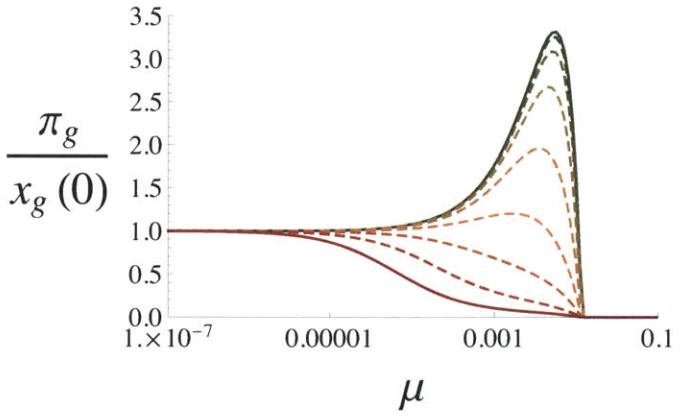


Figure 6-4: Mutagenerator success, $\pi_g/x_g(0)$, as a function of the mutator production rate μ and the availability of beneficial mutations, Nb. Values of Nb range from 1 (bottom line) to 10^4 (top line), in logarithmically spaced intervals. Increasing μ magnifies either the advantage or disadvantage to mutagenerators, depending on the landscape. The "mutator catastrophe", in which the wild-type mutation rate is lost to mutagenerators, occurs when $\mu \geq 1 - q_m/q_w \approx 0.02$.

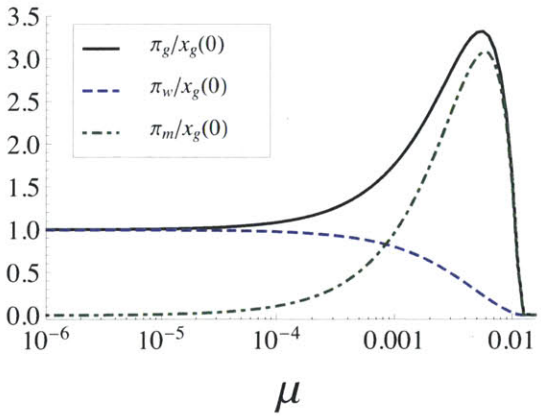


Figure 6-5: First adaptation probabilities, in relation to initial mutagenerator abundance, for the mutagenerator strain and its subtypes, as a function of the rate μ of mutator production. Other parameter values are as in Table 6.1. With increasing μ, the overall success of mutagenerators increases, but the likelihood of wild-type fixation decreases.

6.4 Discussion

Overview Our results show that mutageneration, as a selectable trait, is plausible both mechanistically and evolutionarily in *E. coli* and related organisms.

On the mechanistic side, previous work has highlighted the role of SSRs in MMR genes to act as switches between the wild-type and mutator alleles. In particular, mononucleotide repeats in *mutT* and a 6pb motif repeated three times in *mutL* have led to the production of mutator alleles in *E. coli* and *S. typhimurium* [253, 112, 19]. In particular, the repeated 6bp motif has been validated as a genetic switch for the mutation rate, and spontaneous restoration of the wild-type mutation rate through reverse mutation has been observed, in *S. typhimurium* [112, 59]. Our bioinformatic analysis demonstrates that this repeated 6bp motif is not forced by the constraints of the amino acid sequence and codon bias. Synonymous changes in the codons are possible and do occur accross various strains. However synonymous changes occur much more infrequently than would be expected highlighting that the specific repeat structure is likely important for function and is possibly under positive selective pressure.

Our analysis of the evolutionary model shows that this hypothesis is credible from an evolutionary standpoint as well. Under conditions that are reasonable for *E. coli*, mutagenerators are more likely than nonmutagenerators to achieve and sustain the first beneficial mutation. Moreover, mutagenerators are able to out-compete nonmutagenerators for conditions under which pure mutators would not be favored. Thus, mutageneration is an effective hedge strategy, able to adapt when possible and maintain stability when necessary.

Fixation of mutagenerators Our model focuses on the short-term question of sustaining the first beneficial mutation. Moving to a longer time-scale, we can ask how a mutagenerator strain might achieve fixation in a population initially dominated by nonmutagenerators. The fixation of mutator alleles in the long-term *E. coli* experiment [263, 19] suggests that mutagenerator fixation would most likely occur in a single iteration of the four-step process outlined in the introduction, rather than mul-

tiple iterations. Furthermore, analysis of this experiment by [252] and [18] suggests that mutator selective sweeps (step (b)) are likely to occur on the basis of multiple beneficial mutations, rather than just one.

For the evolution of mutageneration to be complete, the wild-type mutation rate must be restored (steps (c) and (d)) through reverse mutation or recombination, as described in the introduction. For reasons that are unclear, this did not occur in the long-term *E. coli* experiment, even after the rate of adaptation in the mutator-dominated populations had slowed significantly [253]. It did, however, occur in the *S. typhimurium* experiments through spontaneous reverse mutation and subsequent fixation of the wild-type [112, 59]. Further research is necessary to sort out the mechanisms and rates of restoration of the wild-type rate, and to address why this reversal did not occur in the *E. coli* populations.

Mathematical analysis of the full process of mutagenerator fixation would be complicated by the effects of clonal interference, multiple mutations, and exhaustion of the supply of beneficial mutations. However, this process could readily be investigated through simulation.

It is also possible to address this question experimentally, by setting up competition between mutagenerator and nonmutagenerator strains of bacteria. Nonmutagenerator strains could be obtained using the method of [58], who engineered nonmutagenerator strains of *S. typhimurium* by disrupting SSRs through synonymous nucleotide substitutions. These nonmutagenerator strains were verified to have a lower population average mutation rate, likely due to the absence of mutator subpopulations.

MMR-deficient mutators are also recombinators Strains with defects in the MMR pathway have not only a high mutation rate, but also an increased rate of homeologous recombination [226, 70]. Thus far, we have mentioned this increased recombination rate only in terms of its potential to facilitate restoration of the wild-type mutation rate. However, this increased recombination rate may have many other effects on the evolution of MMR-deficient strains [70]. In particular, it may

facilitate adaptation by mitigating the effects of clonal interference [93, 65]. It may also guard against further, possibly untenable, increases in the mutation rate [104]. These effects may have important consequences for the evolution of MMR-mediated mutageneration, which merit further study.

The rate of mutageneration If it is accepted that mutageneration is indeed a selectable trait, the question then arises of how evolution has shaped the rate of mutator production (parameter a in our model). There are several tradeoffs to be considered here.

First, the benefits and drawbacks mutator production depend on the fitness landscape, as shown in figure 6-4. Rapid production of mutators magnifies both the advantages (when beneficial mutations are plentiful) and disadvantages (when beneficial mutations are scarce) of mutageneration. Thus the answer to this question depends on the variety of environments and fitness landscapes faced by an organism over the course of its evolutionary history. Organisms facing greater variability and harshness in their environments may evolve a higher rate of mutator production. In this light, it is unsurprising that [58] found a number of SSRs in the *mutL* and *mutS* genes of almost all common bacterial pathogens. An important open question is whether pathogenic bacteria have a significantly greater number (in the statistical sense) of SSRs in MMR genes than their non-pathogenic relatives.

Second, it is important to consider constraints imposed by the genetic basis for mutageneration. If indeed SSRs in MMR genes are the primary genetic mechanism for this trait, then the amino acid sequence restricts the possible number and length of these repeats. Reversibility is also an important constraint: if a mutator acquires two independent mutations in MMR genes, this mutator may become "locked"—unable to produce wild-type offspring—since independent reversal of both of these mutations may be highly unlikely. The need for reversibility may therefore constrain the number of SSRs in MMR genes, and with it the rate of mutator production.

Conclusions Here we have seen that mutageneration forms an evolutionary advantageous strategy for organisms. The mechanism for mutator production is connected with a special codon pattern within the mutL gene. We find that this pattern seems to be under positive selection as synonymous changes to the repeat occur much less frequently than expected.

The success of mutatagenerators is dependent on several key parameters including the production rate of mutator strains, and the strength and availability of beneficial mutations within the genetic landscape. Experimentally determined values of the key parameters show that, for *E. coli*, mutageneration is a advantageous trait explaining why mutators are often observed in natural populations [144, 164, 199]. However, we also find that for certain reasonable parameter regimes mutageneration would be disadvantageous and would likely be excluded. For example, if the there are few beneficial mutations then mutageneration is not the best strategy.

One of our most fundamental results is the tradeoff associated with the mutator production rate. This rate is set to compromise the relative benefits and disadvantages of producing mutators dependent on the availability of beneficial mutations. Moving forward it will be important to determing if, and how, this rate is being selected for by evolution.

Understanding the rate of mutator production, or testing the effect of any important parameter, requires an experimental system where mutagenerators can be directly competed against nonmutagenerators. To do this we would like to collaborate in the future with experimentalists to create a nonmutagenerator strain of *E. coli*. This can be done by replacing the repeat in mutL with synonymous codons with no repeat structure. We can then directly compete mutagenerators and nonmutagenerators in a variety of contexts such as a situation where there are few beneficial mutations. Such a situation can be imposed by selecting a growth medium that allows for the survival of only a few well-defined genotypes. The challenge in these experiments may be that the timescales for the competitive, adaptive and exclusionary processes may be incredibly long, especially as mutators are produced slowly.

Mutators are quite successful in human infections and also are shown to have

decreased antibiotic susceptibility [200, 298]. Thus understanding the general constraints faced by strains which produce mutators has implications for better treating infections.

From an evolutionary perspective it is interesting that there are species which produce mutators but do not use the same repeated codon mechanism found in *E. coli*. For example, *P. aeruginosa* does not have the same codon repeat pattern in mutL but this species frequently produces mutator strains [200, 298]. Moving forward we are interested in investigating the variety of mechanisms which produce mutators in a diversity of species and determining whether there are any universal similarities amongst these mechanisms. This can be done by sequencing mutator strains in a number of species and comparing them to the wild-type in order to determine how the elevated mutation rate was produced. If we find that the mechanisms are all quite different then this would suggest that mutageneration is important and that there are a large number of ways to evolve this strategy. If the mechanisms are similar then this implies common structural functions across diverse genomes.

Chapter 7

Conclusion

In this thesis we have seen that the structure of organisms has wide reaching implications for biological function at scales ranging from the genetic encoding of genes to the architecture of tree canopies. The evolution of life through various degrees of complexity and across vastly different scales is accompanied by critical changes in morphology. Indeed our work highlights that some of the major innovations of evolution are morphological whether those be the addition of mitochondria to unicellular organisms, the formation of structured unicellular communities, or the optimization of tree canopy shape. Structure is important because it both regulates the internal metabolic processes of organisms and mediates an interaction with the external environment and resource supply.

By focusing on basic physical constraints and idealized structure we have developed and explored a series of first-order hypotheses and theories in a variety of case studies where we are able to successful predict features ranging from regional speciation to the dynamics driving mutation rate. Importantly, each is associated with a set of testable predictions which can be used to guide future experimental work.

In our work on microbial energetics and evolution we proposed several testable hypotheses. The model predicts the minimum size of prokaryotes. For the maximum size of prokaryotes and the minimal size of eukaryotes it is important in the future to examine the scaling of many features such as the total ribosomal and mitochondrial volumes found within each of these classes of organisms as these may determine the

observed limits. Our work suggests that metabolic scaling and partitioning may be directly connected with both resource use of single cells and the partitioning of transcriptional resources within that cell. Both of these possibilities could be examined by tracking the growth of a single cell concurrently with its resource consumption and internal transcriptional composition.

In our work on biofilms we saw how emergent raised structures lead to a metabolic and growth advantage for the colony. We have shown how the structure of single features can be explained by the effect that vertical geometry has on resource acquisition. The spatial pattern formation results from the interplay between diffusive dynamics and resource supply management. However, we interpret the vertical structure and horizontal patterning mostly in terms of top-level resource supply considerations. We still do not understand the specifics of how cellular physiologies regulate emergent structure. For example, although we understand that a given ridge width is optimal for resource acquisition and growth we do not know how this width is maintained constant as the ridges grow taller. Various genetic knockouts reveal key physiological process for pattern formation such as motility, and we have identified graph theoretic metrics for categorizing mutants. A full mechanistic interpretation of pattern formation is still elusive. Moving forward we plan to automate analysis of the entire mutant library in order to uncover the "program" which gives rise to patterning. Once we uncover the key physiologies involved in this "program" we can then proceed with targeted experiments to measure the chemical and physiological state of the colony over the time-series of development. A key technique will be to tag a given gene of interest with fluorescent proteins and monitor its expression both spatially and temporally.

Our work suggests that some aspects of the emergent patterns are optimal for community metabolism. For this to be the case it is possible that cellular traits have been selected for such that optimal features are produced. We suggest that the selection of these traits can be verified via experimental competition between strains. This can be done my mixing together two or more genetic knockouts and observing the effects on patterning and the success of each strain within the community. It

will be interesting to find out whether a single strain can outcompete another or if coexistence is possible in this context. Given competition dynamics and possible coexistence, it is also important to determine whether mixed strain colonies have characteristically different patterning than observed for either strain growing in isolation. Such experiments could provide much insight into the processes by which community organization and multicellularity evolve and are continually modified.

Our model for tree structure and physiology is successful in its ability to predict the regional variation in the important trait of maximum tree height. We suggested that this maximum tree height, which represents an upper bound, is the key variable in describing the overall structure of a forest. Recent theoretical work has shown the concrete connection between plant size and the spacing of trees [294, 83]. This provides an avenue for us to link local resources to predict the distribution of trees and standing biomass of a forest via our model for maximum tree height. Our model is founded on a connection between size, canopy structure, and a radiative energy balance. As such once we have determined the overall standing biomass and distribution of tree sizes we can also predict the albedo of an entire forest region which we can verify using existing remote sensing data. The connection between albedo and resources is important for understanding the feedback between terrestrial vegetation and overall climate or local weather. We have begun to look at these dynamics by adding our vegetation dynamics to climate models in order to determine the fundamental feedback between vegetation and longterm climate trends.

Our work on mutagenerators showed that a mutational switch is advantageous to populations of cells and that an observed mechanism for switching is highly conserved across similar species. However, there are many other species, such *Prochlorococcus* or *Pseudomonas aeruginosa*, which exhibit mutator strains but do not have any repeated codon modules which resemble those found in *E. coli*. Thus there must be another mechanism that is producing mutators and several questions stem from this: Are these other mechanisms also switches? Across diverse species, what is the variety in mechanisms for producing mutators and are there any universal similarities amongst these mechanisms? These questions can be easily addressed by sequencing numerous

mutator strains from diverse organisms and determining which portion, if any, of the mutation genes (homogous genes to mutL or mutH, mutS, and mutL) are repeatedly damaged to produce mutators. For any possible mutator mechanisms we can again use our bioinformatic technique to test whether there is positive selection for the ability to produce mutators. This has far reaching implications for understanding how mutation rates are managed and how they perhaps form one of the most central traits of an organism.

Moving forward it is important to consider how multiple levels of structure and organization within an organism interact to form physiology and fitness and how this package is dynamically selected by evolution (e.g. Figure 4-3). Within the biofilm system we see a great potential for future study. The biofilm system represents most clearly analogies for all of the distinct levels of organization prevalent in the biological world: At the top level the cooperation and competition of cells and various strains form an ecological system. At the intermediate level, the overall structure of the biofilm represent the rudiments of complex multicellularity or cell differentiation, and individual features may behave like individuals within a forest governed by resource flow and individual metabolism. At the most fundamental level the rate of mutation may control the overall rate of morphological adaptation and mutation may give rise to divergent patterning in a given region which is analogous to cancer formation in complex multicellular organisms. Biofilm morphology is governed by the key processes of genetics, environmental control, and stochastic spatial dynamics and the accessibility of this system to a variety of measurements along with easy modifications of environmental, structural, and genetic situations form an ideal model system with far reaching implications.

Appendix A

Appendix for Chapter 2

A.1 Metabolic partitioning, bioenergetic constants, and chemostat measurements

In the main text we compared the energetics of a populations of cells with the parameter values that we obtained from fits of single-cell growth trajectories. Analyses of populations rests on the Pirt model [211] which linearly relates the consumption rate Q of a limiting resource to the specific growth rate μ of a population along with its yield coefficient Y and maintenance metabolism P:

$$Q = \frac{\mu}{Y} + P. \tag{A.1}$$

Although the consumption rates measured in chemostat experiments may range from glucose to oxygen to light absorption it is always possible to represent this relationship in the normalized units of percentage growth and maintenance

$$1 = \frac{\mu}{Y}\frac{1}{Q} + \frac{P}{Q} \tag{A.2}$$

where this equation can be compared to the single cell analysis as represented by equation 5 of the main text and discussed below. Equations 3 and 4 from the main text detail the correspondence between the Pirt population model and the single cell

framework. This allows us to interpret population based values as the averages of single cell values and in particular we note that

$$\bar{E}_m = \frac{N}{Y} \tag{A.3}$$

$$\bar{B}_m = PN \tag{A.4}$$

and thus

$$\bar{b} \equiv \frac{B_m}{E_m} = PY \tag{A.5}$$

independent of the type of consumption being measured by Q. Here the bar notation denotes average values over a population of cells.

Thus we can calculate b from information obtained from population experiments. We first located those experiments that correspond to each of the individual species analyzed in this paper. The resulting b values are useful for informing the initial conditions of the single cell growth trajectory fits (see below), and for subsequent comparison with the best fit value of b for each trajectory. For four species (*E. coli*, *B. subtilis*, *C. albicans*, and *T. weissflogii*) we are able to directly compare our single cell analysis to chemostat experiments. For the diatom *Lauderia borealis* (eukaryotic autotroph) we used the value from *T. weissflogii* for comparison, and for *Calanus pacificus* and *Pseudocalanus sp.*, two species of copepods (multicellular heterotrophs), we used chemostat results from the rotifer *Brachionus calyciflorus*. Table A.1 reports Y, and P along with the calculated value of b for each experiment along with notes on each of the sub-experiments that we used.

A.1.1 Compilation of b estimates from diverse species

In general, a large number of experiments exist where it is possible to calculate b for a wide variety of species, and we have created a general compilation of b values from a survey of the literature (this compilation is attached as a supplementary data file and also as a PDF). Figure A-1 gives the distribution of b values for both prokaryotes and eukaryotes. For both groups the b values are approximately log-normally distributed

Species	Evolutionary Life-history Grouping	Chemostat Temperature	Y	P	b		notes	Refs.
					Raw Value	Normalized to 20 ° C		
B. subtilis	Prokaryote	37 ° C	87.57 (dry g cells · mol glucose^{-1})	1.22×10^{-7} (mol glucose · s^{-1} · dry g cells^{-1})	1.07×10^{-5} (s^{-1})	2.90×10^{-6} (s^{-1})	we used the wild type values for growth on glucose	[273]
E. coli	Prokaryote	30 ° C	59.7 (dry g cells · mol O_2^{-1})	1.22×10^{-7} (mol O_2 · s^{-1} · dry g cells^{-1})	7.30×10^{-6} (s^{-1})	3.33×10^{-6} (s^{-1})	we used the glucose experiment to compare well with the growth trajectories of refs. [231, 232]	[92]
C. albicans	Eukaryote	30 ° C	.40 (dry g cells · g maltose^{-1})	4.23×10^{-6} (g maltose · s^{-1} · dry g cells^{-1})	1.68×10^{-6} (s^{-1})	7.66×10^{-7} (s^{-1})	we used the measurements for growth on maltose to be most comparable to ref.[119]	[254]
T. weissflogii	Eukaryote	18 ° C	16.70 (dry g cells · mol O_2^{-1})	2.37×10^{-8} (mol O_2 · s^{-1} · dry g cells^{-1})	3.97×10^{-7} (s^{-1})	4.67×10^{-7} (s^{-1})	estimated from the reported gross photosynthesis	[85]
Brachionus calyciflorus	Metazoan	25 ° C	—	—	3.61×10^{-6} (s^{-1})	2.43×10^{-6} (s^{-1})	measurements are given in specific ingestion rates so the units of Y and P are unclear but b can still be calculated	[33]

Table A.1: Chemostat growth energetics for various organisms spanning three major evolutionary life-history transitions.

The units of consumption rate Q are the same as the maintenance term P.

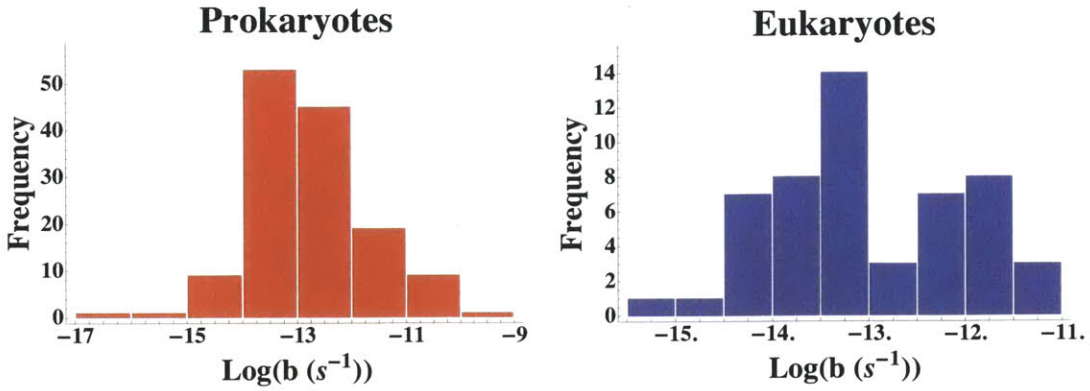


Figure A-1: The distribution of b values. The values have been normalized to 20° C for both prokaryotes and eukaryotes. The median value for prokaryotes is $b = 2.42 \times 10^{-6}$ s^{-1} and for eukaryotes it is $b = 1.84 \times 10^{-6}$ s^{-1}. The mean value for prokaryotes is $b = 5.79 \pm 9.99 \times 10^{-6}$ which is indistinguishable from the eukaryotic value of $b = 3.39 \pm 3.17 \times 10^{-6}$.

and have means that are indistinguishable from one another ($b = 5.79 \pm 9.99 \times 10^{-6}$ for prokaryotes and $b = 3.39 \pm 3.17 \times 10^{-6}$ for eukaryotes). We were able to pair many of the compiled b estimates with measurements of cell size and estimates of body mass. These are the data presented in Figure 2-3 D. Here we find that there is no obvious relationship between b and body mass for both prokaryotes and eukaryotes.

In compiling experiments for estimating b we included a diverse set of growth conditions in order to cover a large number of species and body sizes. These conditions include growth on a range of substrates, in different experimental setups (chemostats, batch culture time evolutions, and recycling fermenters), and at various temperatures. These conditions contribute to the relatively large spread in the value of b. Some of the growth condition deviations can be systematically eliminated, using, for example, temperature normalization (see below), while others, such as growth on different substrates or in different culture setups, are more complicated to standardize. For example taking only chemostat experiments the prokaryotic mean becomes $b = 4.23 \pm 5.12 \times 10^{-6}$, which is similar to the value listed above but with much less variance. Similarly, growth on different substrates contributes to the variation in b where, for *E. coli*, ref. [92] uses the same experimental setup but alters the growth medium (Table A.2) resulting in a value of b which ranges over a factor of about 2.

Substrate	Y (dry g cells · mol O_2^{-1})	P (mol O_2 · s^{-1} · dry g $cells^{-1}$)	b (s^{-1}) (Normalized to 20 ° C)
Acetate	20.3	3.33×10^{-7}	3.09×10^{-6}
Pyruvate	34.3	2.03×10^{-7}	3.18×10^{-6}
Lactate	35	2.03×10^{-7}	3.24×10^{-6}
Glucose	59.7	1.22×10^{-7}	3.33×10^{-6}
Glycerol	50.9	1.61×10^{-7}	3.75×10^{-6}
Arabinose	57.8	1.78×10^{-7}	4.69×10^{-6}
Fructose	56	1.97×10^{-7}	5.05×10^{-6}
Fumerate	40.4	3.14×10^{-7}	5.79×10^{-6}
Galactose	58.2	2.69×10^{-7}	7.16×10^{-6}

Table A.2: *E. coli* chemostat growth energetics for different substrates. All data is from Ref. [92].

Many of the studies that we compiled do not report an estimate of Y and P and in such cases we fit a linear relationship to digitized data of Q vs. μ. For multiple studies we fit only the early portion of the data (slow growth rates) where there is a clear linear relationship between Q and μ and the data agree with equation 1 of the main text. At high growth rates nonlinearities, such as saturation or accelerating consumption, can appear in the data which we do not consider in our framework or data compilation. For one of the data points b has a negative value which is not realistic. This is likely due to noise in the data or some unaccounted for physiological response of the particular species.

Experiments which can be used to estimate b often measure multiple consumption rates within a single experiment, for example substrate consumption and oxygen consumption. For our purposes here it is important to measure a resource which is directly proportional to the overall metabolic rate. In some cases it is essential to measure the limiting resource. For example in ref. [254] the carbon source is the limiting resource for growth and estimates of b using substrate consumption are an order of magnitude smaller than estimates obtained from oxygen consumption. This matters less in other studies where the two estimates can be nearly identical (e.g. ref. [251]).

For species where the conversion N between consumption rate and metabolic

energy production is known, then it is possible to directly calculate E_m and B_m. For *E. coli* ref. [92] provides the number of moles of ATP produced per mole of oxygen consumed, denoted here as n. Combining values of n with an understanding of the ATP synthesis the energy production conversion is given by $N = n\Delta G_{phos}$ where ΔG_{phos} $\left(\text{J} \cdot \text{mol ATP}^{-1}\right)$ is the phosphorylation potential of ATP. From ref. [277] for *E. coli* growing on glucose $\Delta G_{phos} = 4.65 \times 10^4$ $\left(\text{J} \cdot \text{mol ATP}^{-1}\right)$, from ref. [92] the aerobic energy conversion efficiency is $n = 4.31$ (mol ATP $\cdot$ mol O_2^{-1}), and the *E. coli* value for Y can be found in Table A.1. Given these values we calculate that $E_m = n\Delta G/Y = 3345$ (J $\cdot$ dry g^{-1}). This value is comparable to values found previously for multicellular organisms which range from $E_m = 800$ to $E_m = 13000$ (J $\cdot$ dry g^{-1}) for embryos and juveniles of several species of birds, fish, and mammals [179] (also see supplement of ref. [121]). This suggests that the commonality in the unit energetics that we found in microbes may also extend to larger multicellular organisms.

Similarly, the maintenance cost for *E. coli* is given by $B_m = P\Delta G_{phos} n = .025$ (W $\cdot$ dry g cells $^{-1}$) for the experiment in ref. [92] which is carried out at 30° C. We mention the temperature of the experiment here because within the metabolic theory there are well documented temperature effects and normalizations which we discuss below.

A.1.2 Metabolic partitioning from individual cells and chemostat populations

Using equation 11 from the main text it is possible to estimate $\bar{\gamma}$ given an estimate of μ for each species. We also compiled estimates of μ_{max} for many of the species where we have already estimated b. We use μ_{max} to calculate $\bar{\gamma}$ because this represents a limiting value and also compares well to the single cell growth trajectories where the conditions are such that these individuals are often growing near their maximum rate. The data presented in figure 2-3 B are the result of pairing b with measurements of μ_{max} and cell size.

A.2 Temperature normalization

Following refs. [105, 106, 247] temperature has been shown to affect the total metabolic rate of an organism according to

$$B(T) = B(T_0)\, e^{E(T-T_0)/kTT_0} \tag{A.6}$$

where it is assumed that this temperature dependence is carried by the normalization constant as

$$B_0(T) = B_0(T_0)\, e^{E(T-T_0)/kTT_0}, \tag{A.7}$$

where E is an average activation energy for biochemical reactions, k is the Boltzmann constant, T is the operating temperature of an organism, and T_0 is a standard temperature of interest [105, 106, 247]. This has implications for several of the parameters that we use in our model which can be normalized to a common reference temperature. Following equation 3 of the main text $Q\bar{m}N$, and $B_m = PN$ will have the same temperature dependence as B while $\bar{E_M} = N/Y$ will be independent of temperature consistent with previous assumptions [106] (specific growth carries the temperature dependence). These two relationships along with equation A.7 demonstrate that the parameter $b = B_m/E_m$ from our model will depend on temperature while $\rho = \frac{B_m}{B_0} m^{1-\alpha}$ and $\gamma = 1 - \rho$ will not:

$$b(T) = b(T_0)\, e^{E(T-T_0)/kTT_0} \tag{A.8}$$

$$\rho(T) = \rho(T_0) \tag{A.9}$$

$$\gamma(T) = \gamma(T_0)\,. \tag{A.10}$$

For those data where the operating temperature T is reported we use these normalizations. In this paper we used $E = .6$ eV which has been shown to be the average value across a diverse set of organisms [105]. The most important feature of the normalization was that we consistently chose a common E and reference temperature T_0.

A.3 Derivation of the growth trajectory

Typically the partitioning of equation 4 of the main text is rewritten as

$$\frac{dm}{dt} = am^{\alpha} - bm \tag{A.11}$$

with $a = B_0/E_m$ ($g^{1-\alpha}\ s^{-1}$) and $b = B_m/E_m$ [293, 179]. The mass trajectory for a free value of α can then be solved as

$$m(t) = \left[1 - \left(1 - \frac{b}{a}m_0^{1-\alpha}\right)e^{-b(1-\alpha)t}\right]^{1/(1-\alpha)} \left(\frac{a}{b}\right)^{1/(1-\alpha)}. \tag{A.12}$$

Recognizing that $\frac{b}{a} = \frac{B_m}{B_0}$ it follows that $\gamma_0 = 1 - \frac{b}{a}m_0^{1-\alpha}$ and we can rewrite this equation as

$$m(t) = \left[1 - \gamma_0 e^{-b(1-\alpha)t}\right]^{1/(1-\alpha)} \left(\frac{1}{1-\gamma_0}\right)^{1/(1-\alpha)} m_0 \tag{A.13}$$

which is the form from the main text. This form is appealing because γ_0 is a nondimensional number bounded between zero and one, and the initial mass now appears as a simple factor.

Both systems of parameters are useful in different contexts and each makes certain interpretations of data more conceptually explicit (e.g. metabolic partitioning vs. unit costs).

A.4 Normalized growth trajectories

Normalizing the lifespan, or rate of growth, reveals clearly the general shifts in the metabolic partitioning between these taxa. Choosing the dimensionless time variable $\tau = b(1-\alpha)t - \ln(\gamma_0)$ Eq. A.13 becomes $\gamma = e^{-\tau}$. The dimensionless temporal parameter τ accounts for differences in the overall metabolic and bioenergetic rates. This relationship is plotted in Figure A-2 of the main text along with each of the 6170 datapoints from the individual growth trajectories. All of the data lie tightly along the predicted curve reflecting an underlying commonality in the form of growth, and the goodness of fit of the model simultaneously across diverse species. On this curve taxa

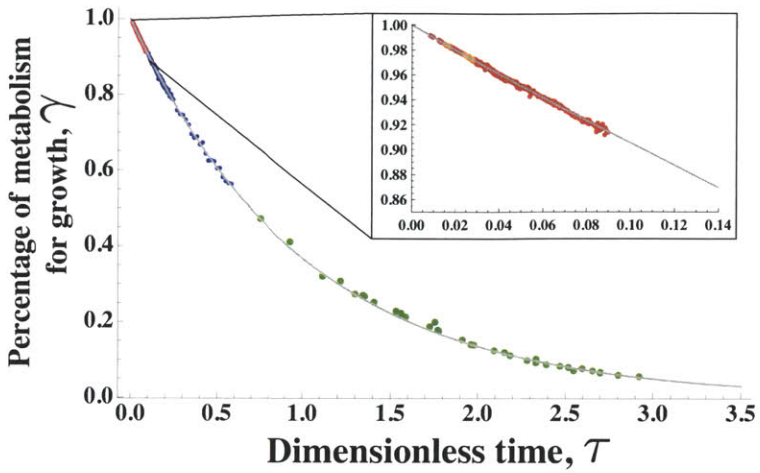


Figure A-2: The universal metabolic partitioning curve for all of the individuals analyzed in Chapter 2. Colors denote the three evolutionary life-history groups: prokaryotes (red), eukaryotes (blue), and small metazoans (green). The insert shows the two prokaryotic species with *E. coli* in red and *B. subtilis* in orange. The points in this plot are from 45 individual growth trajectories. Not all of the 6170 datapoints are individually distinguishable because of the tight clustering within this transformation.

are separated based on differences in the relative metabolic power devoted to growth, γ. There is a constant decrease in γ as we move to longer normalized timescales which also corresponds to moving across the three evolutionary life-history transitions. This view highlights that prokaryotes and small unicellular eukaryotes live over relatively short normalized timescales while using almost all of their metabolism for growth; they truly "live fast and divide young" (Figure A-2).

In previous work similar normalizations have been used where the plot in ref. [293] can be interpreted as the fraction of metabolism devoted to maintenance (rather than growth as in Figure A-2) for a fixed value of $\alpha = 3/4$. Our free-α version accounts for differences in growth trajectories between individuals related to the overall metabolic scaling of an organism in addition to variation of the the unit bioenergetic costs.

A.5 Population Growth Rate

In the main text we discussed the population growth rate, μ, of an organism which is based on our derivation of the generation time, G, along with the fecundity, f, and

percentage of the population to reach the age of reproduction, L. In most of the single organism studies f and L are not measured in addition to the growth trajectory. In order to deal with this issue, along with the differing reproductive strategies across taxa, we introduce a "mass fecundity" where we consider the total mass production over the life cycle of an organism. That is, we define population growth rate by

$$\mu = \ln(\epsilon)/G \tag{A.14}$$

where ϵ quantifies the factor change in body mass over a life-cycle, $M_d = \epsilon m_0$. These are the values reported in Figure 2-3 of the main text.

A.6 Fitting the interspecific relationships for the population growth rate and the fraction of metabolism devoted to growth

In the main text we present an interspecific fit for the dependence of population growth rate on body mass and from this we are able to predict the fraction of metabolism devoted to growth for both prokaryotes and eukaryotes. It is useful to explicitly give the mass dependence of generation time which determines μ. Using equation A.12 we can rewrite equation 9 from the main text as

$$G = \frac{1}{b(1-\alpha)} \ln\left[\frac{1-\frac{b}{a}m_0^{1-\alpha}}{1-\frac{b}{a}(\epsilon m_0)^{1-\alpha}}\right]. \tag{A.15}$$

This form has the appealing features that the mass dependence is explicit and that the parameter a is based on unit costs and thus should be constant across organisms of different size similar to b and in contrast to γ_0 which depends on the initial mass of a cell. Using versions of the equations which depend on the parameter a is best suited for interspecific fits.

In fitting the interspecific data for prokaryotes and eukaryotes we fix the value of

b to the average for each group using our compilation (Figure A-1). We then find the best fit values of a and α for the interspecific data using a reduced major axis regression. The results for the best fit values of α are given in the main text.

We were unable to fit the interspecific data for the metazoans because it is unclear how features such as fecundity or ϵ, the ratio of reproductive mass to initial mass, change with body size. As discussed in the main text these alterations may be critical for allowing metazoans to grow larger while avoiding the limit where the fraction of energy devoted to biosynthesis goes to zero, as is the case for unicellular eukaryotes.

Similar to the generation time, the fraction of metabolism devoted to growth can be written in terms of the parameter a and cell mass using equation 7 (or 6 and 8) of the maintext:

$$\gamma = 1 - \frac{B_m}{B_0} m^{1-\alpha} = 1 - \frac{b}{a} m^{1-\alpha}. \tag{A.16}$$

Using the taxonomic average value of b and the best fit values of a and α from the interspecific fit of population growth rate from above we are able to predict γ as a function of body size for prokaryotes and eukaryotes and these are the curves drawn in Figure 2-3 B of the main text.

As described above we cannot fit the interspecific data of growth rate for the metazoans using our framework for G and μ. In order to predict the interspecific relationship of $\bar{\gamma}$ we instead fit a power law to growth rate [69] and use equation 11 from the main text where we use the eukaryotic average for $\bar{b}$.

A.7 The growth of buds within the yeast complex

Our hypothesis for the yeast complex is that when a new bud forms nearly all of the growth energy from the entire complex is devoted to that bud. This is to say that a bud grows with the metabolic energy of a much larger organism than its own size. In order to test this hypothesis we first fit our growth model for $m(t)$ to the entire complex and we then use those fit parameters to describe the growth trajectory of each bud in agreement with data. Fitting $m(t)$ to any region of total complex will give a prediction of the mass at any subsequent time and from this the total growth

rate of the entire complex at any given time is given by

$$\frac{dm}{dt} = am_{tot}(t)^{\alpha} - bm_{tot}(t) \tag{A.17}$$

where it should be noted that $a = \frac{b}{\rho} m_0^{1-\alpha}$ (see equation A.11) using m_0 of the entire complex. If all growth energy is devoted to the bud then the growth dynamics of the bud are described by

$$\begin{cases} \frac{dm_{bud}}{dt} = am_{tot}(t)^{\alpha} - bm_{tot}(t) \\ m_{tot}(t) = m_{tot}(t_b) + m_{bud} \\ m_{bud}(t_b) = m_0^{bud} \end{cases} \tag{A.18}$$

or

$$\begin{cases} m_{bud}(t + \Delta t) = m_{bud}(t) + \Delta t \left[am_{tot}(t)^{\alpha} - bm_{tot}(t) \right] \\ m_{tot} = m_{tot}(t_b) + m_{bud} \\ m_{bud}(t_b) = m_0^{bud} \end{cases} \tag{A.19}$$

where t_b is the time when a bud starts growing and m_0^{bud} is its initial size. The bud curves in Figure 2-1 D are the result of numerically integrating these dynamics for each bud. Thus each bud trajectory is not a fit but a prediction based on the growth dynamics of the entire complex. The fact that this agrees so well with the data for each bud supports the hypothesis that all growth energy is being devoted to a newly formed bud.

A.8 The reproductive strategy of diatoms

Another example of an altered reproductive strategy is that used by some diatoms (including some of those discussed in [201]), which interrupt single cell growth with long resting phases [201]. Such cell types could not be analyzed with our framework, which only considers continuous growth. During a resting phase our model cannot interpret the energetics of the cell because there are no changes in size even though biosynthesis may be continuing. The two diatom species that we did examine were considered to grow continuously by ref. [201], yet we find here that an asymptotic

mass just before division could signal a resting phase. Thus the low value of alpha found for these species could be the result of a different reproductive strategy (resting) rather than a metabolic constraint. Supporting this hypothesis, a terminal plateau has been previously observed in plant cells, where following nuclear division and separation of internal components by the cellular membrane the cell continues to build the cell wall for final cleavage [171]. This could lead to the apparent asymptote where biosynthesis is occurring but not being reflected by changes to the overall size of the cell. This is similar to the daughter buds of the yeast complex which approach an asymptote related to the dynamics of the entire complex rather than the value of the metabolic exponent α.

However, it should be noted that the low value of alpha could instead be due to the decrease in the ability for these organisms to harvest light resources with increasing cell size based on the packaging of photosynthetic pigments [193]. Similarly, the metabolic scaling exponent has been shown to significantly decrease in light-limited growth conditions for phytoplankton [96]. There is not enough information yet to decipher if either of these effects, resting phases or changes in photosynthetic capacity, is leading to the low value of α.

A.9 Fitting routine for the growth trajectory

We considered a least squares analogy for each individual growth trajectory where we minimize the sum

$$\sum_i \left[m_i - m\left(t_i, \gamma_0, b, \alpha\right)\right]^2 \tag{A.20}$$

where m_i is the measured mass at measured time t_i, and $m\left(t_i, \gamma_0, b, \alpha\right)$ is given by equation 8 of the main text.

For the fits presented in the main text we use three free parameters (α, b, and γ_0). It should be noted that the statistical confidence in each best fit parameter is greatly increased by reducing the number of parameters fit. Below we discuss reduced parameter fits which yield slightly different results.

The optimization of equation A.20 involves many local minima and for this reason we employed a heuristic algorithm. We used the Nelder-Mead simplex algorithm as implemented by the "NelderMead" method from the numerical minimization function "NMinimize" of the Mathematica software. All of the options were set to the Mathematica defaults. We seeded the algorithm with 48 distinct random initial values and allowed the algorithm to minimize the function thus finding a local minimum. We then compared each of these local minima and selected as the best fit the minimum with the lowest sum of square residuals. In many cases the various random seeds yielded very similar sets of parameters with only slightly different sums.

The method requires bounds to be set for each of the three parameters to be fit (α, b, γ_0). By definition gamma is required to be between 0 and 1. For α we know the experimental range across taxa from [69] and we take the bounds to be generously larger than this for the search. For b we allowed the search to include values that were two orders of magnitude larger and smaller than an initial guess. The initial guess was based on chemostat experiments where we were able to find previously published chemostat experiments under similar growth conditions as the single cell experiments for *E. coli* [92], *B. subtilis* [245], *C. albicans* [254], and *T. weissflogii* [85]. We were unable to locate chemostat experiments for *L. borealis* where we instead used *T. weissglogii* [85] as a guess; similarly for the two copepods we used *Brachionus calyciflorus* [33]. The Nelder-Mead algorithm also requires that we choose a range of parameter values in which to choose the initial points of the simplex. For each parameter this range was the same as these bounds except for b. We found that if b is sampled over too large a range then the fits are not tightly constrained. The initial guesses should represent the relatively small variation in b that we observed in the chemostat experiments between species or taxa (Table A.1), and various growth conditions (Table A.2). For a species such as *E. coli* b was observed to vary by about a factor of 2. Thus for those species where we have a chemostat estimate for b our initial simplex sampling includes points that are 1.5 times larger or smaller than the initial guess, leading to a sampling that covers a factor of roughly 2. When we do not have a species estimate (diatoms and copepods) we allow more flexibility in b and

sample initial points over an order of magnitude centered on the initial guess. As the minimization runs the values of b are still allowed to range between the bounds described above (over 4 orders of magnitude).

We used a fixed initial mass for each fit. For data where mass is given at $t = 0$ we take this to be m_0, for instances where the initial time point is close to zero relative to the length of the time series we use a linear fit to the first third of the data to estimate m_0. For cases where the initial timepoint is not close to zero (*Calanus pacificus*) we used an exponential fit to the first third of the data.

A.9.1 Reduced parameter fits.

Reducing the number of parameters in the model greatly increases the statistical confidence of the best fit value for each parameter. Yet this comes with a decision about which parameter to fix. In previous studies α is taken to be fixed [293, 179, 106, 121], but ref. [69] illustrates that this exponent varies between taxa and it is thus reasonable to consider that it might also vary between species. The parameter b can be estimated from population studies which allows us to examine the value for a single species independent of the growth trajectory fits. This is not commonly possible for α at present.

We first tested the effect of imposing a fixed value of α for each of the growth trajectory fits. For each fit we choose the appropriate value of α based on taxonomy ($\alpha = 1.96$ for prokaryotes, $\alpha = 1.06$ for eukaryotes, and $\alpha = .79$ for metazoans [69]). We found that these fixed-α fits can cause the best fit value of b to disagree with the population estimates by an order of magnitude or more. For example in *E. coli* using the free value of α yields an average b of $5.50 \pm 2.31 \times 10^{-6}$ which compares well with chemostat estimates of $b = 3.35 \times 10^{-6}$ for cells growing under very similar conditions; while using the fixed interspecific value of $\alpha = 1.96$ gives $b = 1.47 \pm 3.22 \times 10^{-5}$ which does not agree as closely with chemostat estimates and has greater variance in the best fit values. These fits also have lower R^2 value compared to the free α and fixed-b (discussed below) fits.

Given that for most species we can estimate b from measurements of the same

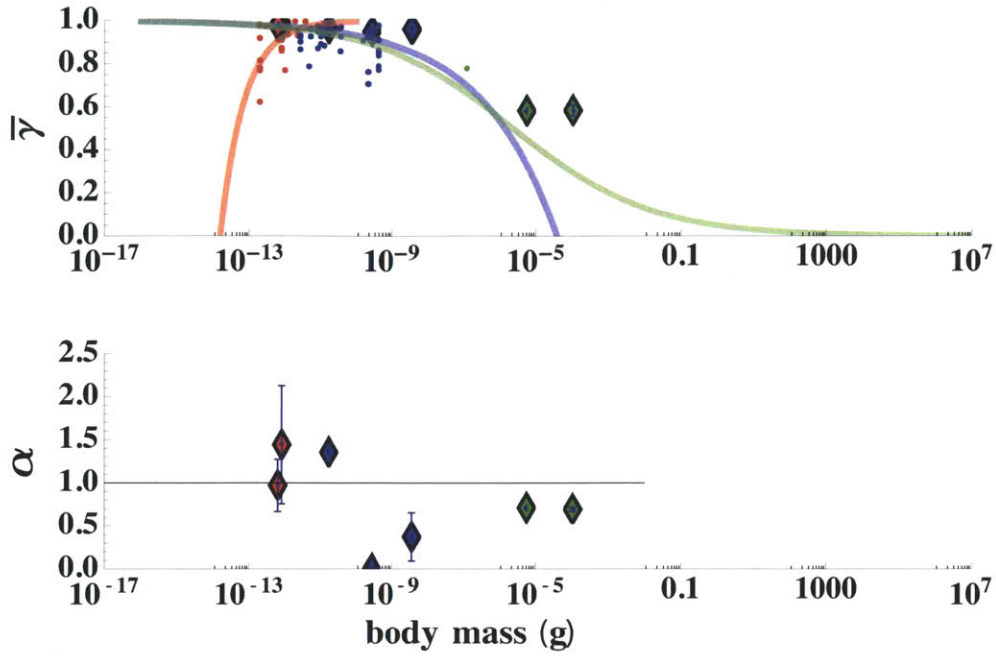


Figure A-3: Results from reduced parameter fits to the individual growth trajectories. The best fit values for the metabolic scaling exponent, α, and the average fraction of metabolism devoted to growth, $\bar{\gamma}$, are shown given b fixed to chemostat estimates for each species. All other lines and data are the same as Figure 2-3.

species in a chemostat it perhaps makes more sense to treat these values as constant in a reduced parameter fit. Here we explicitly fix b to chemostat estimates and fit the growth trajectories using only α and γ_0. The resulting best-fit α and γ_0 values are given in Figure A-3 where it can be seen that these fits are similar to those found using a free value of b, but we find that the statistical confidence in each parameter is much higher. We fixed the value of b to the population estimate for the same species except for *L. borealis* where we instead used the value for *T. weissglogii* [85] and for the two copepods we used *Brachionus calyciflorus* [33].

A.10 Conversions between measurements of size

In this study we considered all mass data in terms of dry mass. However, the original sources presented their data in terms of many different units for measuring size: wet, dry and carbon weights, cell volume, and bouyant mass. Here we present our methodology for converting between these quantities. These conversions are summarized in

Table A.3: Allometric conversions

Property	**Relationship**	**Ref.**
Dry, m_{dry}, and wet weight m_{wet}	$m_{dry} = \beta_1 m_{wet} \quad \beta_1 \approx .22$	[37]
Volume, V_c, and dry weight m_{dry}	$m_{dry} = \beta_2 V_c^{\eta_1}$	
E. coli	$\eta_1 = .86 \quad \beta_2 = 435$ (dry fg μm$^{-\eta_1}$)	[157]
Bacteria	$\eta_1 = .91 \quad \beta_2 = 162$ (dry fg μm$^{-\eta_1}$)	[195]
Carbon content C and cell volume, V_c, and dry weight m_{dry}	$C = \beta_3 V^{\eta_2}$	
	$\eta_2 = .89 \quad \beta_3 = 224$ (dry fg C μm$^{-\eta_2}$)	[159]
Buoyant mass m_b and dry weight m_{dry}	$m_{dry} = \beta_2 \left(\frac{C}{\beta_3}\right)^{\eta_1/\eta_2}$	derived from refs. [159, 157, 195]
Carbon content C and dry weight m_{dry}	$m_{dry} = \frac{\beta_1 m_b}{1-\frac{\delta_f}{\delta_c}} \quad \delta_c = 1.1 \times 10^6$ (g m^{-3})	[37]

Table A.3.

We convert between volume and dry weight, m_{dry}, using the relationship

$$m_{dry} = \beta_2 V^{\eta_1} \tag{A.21}$$

where empirically, in the units of the original paper, $\eta_1 = .86$ and $\beta_2 = 435$ (dry fg μm$^{-\eta_1}$) for *E. coli* [157] and $\eta_1 = .91$ and $\beta_2 = 162$ (dry fg μm$^{-\eta_1}$) for a variety of bacteria [195].

The carbon content of a cell C has been shown to follow

$$C = \beta_3 V^{\eta_2} \tag{A.22}$$

[159] which given the relationship for dry weight implies that

$$m_{dry} = \beta_2 \left(\frac{C}{\beta_3} \right)^{\eta_1/\eta_2}. \tag{A.23}$$

Empirically, $\eta_2 = .89$ and $\beta_3 = 224$ (dry fg C $\mu m^{-\eta_2}$) [159] and thus $m_{dry} \approx 2.33 C^{.97}$ (fg).

We assume that dry weights scale isometrically with wet weight, m_{wet}, such that

$$m_{dry} = \beta_1 m_{wet}. \tag{A.24}$$

It is observed that an appropriate average for several species including *E. coli* and *B. subtilis* is $\beta_1 = .22$ [37] which agrees with other estimates ($\beta_1 \approx .3$) for *E. coli* and a mammalian cell [53, 185, 38].

Bouyant mass m_b is related to wet weight as

$$m_b = m_{wet} \left(1 - \frac{\delta_f}{\delta_c} \right) \tag{A.25}$$

where δ_f is the density of the fluid in which the cell is suspended, and δ_c is the density of the cell. The dry mass is then given by

$$m_{dry} = \frac{\beta_1 m_b}{1 - \frac{\delta_f}{\delta_c}}. \tag{A.26}$$

We use a cell density of $\delta_c = 1.1 \times 10^6$ (g m^{-3}) which is representative of several bacterial species [37].

A.11 Data compilation for single-cell growth trajectories

For the single cell analysis we obtained growth trajectories from several previously published sources [231, 232, 201, 119, 110, 281]. We were able to obtain the original data and numerical values from the authors for *E. coli* from refs. [231, 232] and *B.*

subtilis from ref. [110]. For the two species of diatoms (*Thalassiosira weissflogii* and *Lauderia borealis*) given in ref. [201], *C. albicans* from ref. [119], and the two species of copepods (*Calanus pacificus*, and *Pseudocalanus sp.*) from ref. [281] we obtained values by digitizing the growth trajectory figures from each paper using the software GraphClick. Below is a more detailed discussion of the growth data from each study. We have included a data file summarizing all of the data used in this study including the original units along with our conversions to mass.

A.11.1 Calculating mass for *E. coli*

For the analysis of a single *E. coli* cell we used the data presented in ref. [231, 232]. In order to calculate cell volume we employed the shape model from ref. [232] where *E. coli* growth is divided into two periods: first, the cell is treated as cylinder with two hemispherical caps, and new biomass results in the simple elongation of the cylindrical portion of the cell; second, the middle of the cell undergoes a constriction and the relationship between biomass production and volume becomes more complicated, where the constricted region can be treated as two intersecting hemispheres of equal size to the caps [232]. The cylinder and the two caps are taken to have the same radius. We obtained data from the authors for the cell length and relative "waist width" time-series presented in ref. [231, 232] along with the noted onset of constriction. From this we extracted the time series of cell volume given the reported average diameter of $d = .933$ (μm) [231, 232]. We convert this to mass using the allometric relationship in Table A.3.

A.11.2 Calculating mass for budding yeast

For the analysis of a budding yeast complex (*Candida albicans*) we used the data from ref. [119]. These data relied on optical methods for tracking cell size. The yeast data is reported in volume units and we convert to mass using the scaling law described in Table A.3.

A.11.3 Calculating mass for the diatoms

Data for single diatoms came from ref. [201] where changes in relative length were measured optically. We first convert to volume using a cylindrical cell shape model along with the constant diameter and initial length of the cell given in ref. [201]. We assume that each cell has the same initial length. We analyzed the two species *T. weissflogii* and *L. borealis*. Our theory concerns the time required to produce a unit of mass, and thus we did not analyze the three diatom species (*S. turris*, *B. aurita*, and *Coscinodiscus sp.*) presented in ref. [201] because they have long periods of dormancy which is not addressed in our model.

A.11.4 Calculating mass for *B. subtilis*

In Ref. [110] the buoyant mass of single *B. subtilis* cells are measured over the course of life cycle with very high temporal resolution. The dry mass of the cell can be calculated using equation A.26 where we approximate the fluid density with that of water.

A.11.5 Copepod data mass

Two species of copepods are analyzed in ref. [281], *C. pacificus* and *Pseudocalanus sp.*. The data represent the population average growth trajectory for the mass of a single individual. This is distinct from tracking one individual through a growth cycle, but represents the tracking many individuals and then averaging. Each of the curves represents this average growth trajectory in different nutrient (prey) conditions. This study varied the type of prey, prey size, and prey concentrations. We picked the fastest growing curve from each set of prey conditions. The original study also varied the growing temperature but we considered only those curves grown at a temperature most similar to the unicellular studies (15° C).

A.12 Parameter values from the individual growth trajectories

Table A.4 provides the best fit parameter values for each of the individual growth trajectories. These are results from allowing α, b, and γ_0 to be free parameters as described earlier in this appendix.

Species	Type	m_0	γ_0	$\bar{\gamma}$	b	α	Number of cells	Refs.
		Initial mass (dry g)	Initial percentage of metabolism for growth	Average percentage of metabolism for growth	Metabolic cost ratio rate (B_m/E_m) (s^{-1}) normalized to 20° C	Metabolic scaling exponent		
B. subtilis	Prokaryote	$4.5\pm0.41\times10^{-13}$	0.978±0.006	0.977±0.005	$3.08\pm0.76\times10^{-6}$	0.97±0.31	3	[110]
E. coli	Prokaryote	$6.21\pm0.56\times10^{-13}$	0.963±0.018	0.966±0.014	$5.5\pm2.31\times10^{-6}$	1.43±0.67	30	[231, 232]
C. albicans	Unicellular eukaryote	$8.7\pm1.\times10^{-12}$	0.947±0.001	0.955±0.004	$9.49\pm0.55\times10^{-7}$	1.35±0.1	2	[119]
T. weissflogii	Unicellular eukaryote	$1.91\pm0.02\times10^{-10}$ †	0.912±0.016	0.873±0.02	$1.29\pm0.24\times10^{-6}$	0.01±0.01	3	[201]
L. borealis	Unicellular eukaryote	$2.54\pm0.05\times10^{-9}$ †	0.82±0.048	0.768±0.097	$3.62\pm1.89\times10^{-6}$	0.47±0.23	3	[201]
Pseudocalanus sp.	Metazoan	$1.8\pm0.\times10^{-6}$	0.471±0.	0.366±0.	$5.95\pm0.\times10^{-6}$	0.82±0.	1	[281]
C. pacificus	Metazoan	$4.95\pm0.98\times10^{-6}$	0.265±0.022	0.139±0.008	$2.42\pm0.01\times10^{-5}$	0.93±0.	4§	[281]
Multi-species Population Compilation:								
	Prokaryote average	——	No data	mass dependent	$5.79\pm9.99\times10^{-6}$	No data	138	see SI
	Eukaryote average	——	No data	mass dependent	$3.39\pm3.17\times10^{-6}$	No data	52	see SI

† Data presented in ref. [201] are given in units of relative mass, these are based on the reported lower bound on size
§ This average is conducted across individuals living in different nutrient conditions.

Table A.4: The energetic constants obtained by fitting our model to single cell growth trajectories.

A.13 Definition of symbols

Individual growth curves Here we present fits for each individual cell that we examined. In each plot the dots represent the compiled data for cell mass against time, and the red line is the best fit of Eq. 8 (of the main text) to the growth trajectory.

Symbol	Name	Units
a	Ratio of the metabolic normalization constant (B_0) to the unit biosynthetic cost (E_m)	$g^{1-\alpha}\ s^{-1}$
α	Metabolic scaling exponent	Dimensionless
b	Ratio of the maintenance metabolic rate (B_m) to the unit biosynthetic cost (E_m)	s^{-1}
B	Total metabolic rate	W
B_0	Size-normalized metabolic constant	$(W\ g^{-\alpha})$
B_m	Metabolic expenditure to support a unit of mass	$(W\ g^{-1})$
ϵ	Ratio of division mass to initial mass	Dimensionless
E_m	Energy to synthesize a unit of mass	$(J\ g^{-1})$
γ	Percentage of metabolism devoted to growth	Dimensionless
γ_0	Percentage of metabolism devoted to growth at the initial size	Dimensionless
$\bar{\gamma}$	Percentage of metabolism devoted to growth averaged over a population in a chemostat	Dimensionless
$\hat{\gamma}$	Percentage of metabolism devoted to growth averaged over the life-cycle of an individual	Dimensionless
G	Generation time	s
ΔG	Phosphorylation potential	$(J \cdot ATP^{-1})$
m	Mass of the cell at any given time	g
m_0	Initial mass of the organism	g
m_d	Division mass of the organism	g
μ	Specific growth rate	s^{-1}
N	Conversion constant for resource use to metabolic power production	(J mol resource $^{-1}$)
n	ATP yield from oxygen	Mol ATP $\cdot$ Mol $O_2{}^{-1}$
P	Maintenance requirement	(mol resource $\cdot\ s^{-1}\ \cdot$ g cells $^{-1}$)
Q	Specific consumption rate	(mol resource $\cdot\ s^{-1}\ \cdot$ g cells $^{-1}$)
ρ	Percentage of metabolism devoted to maintenance	Dimensionless
Y	Yield coefficient	(g cells $\cdot$ mol resource $^{-1}$)

Table A.5: Symbol Definitions.

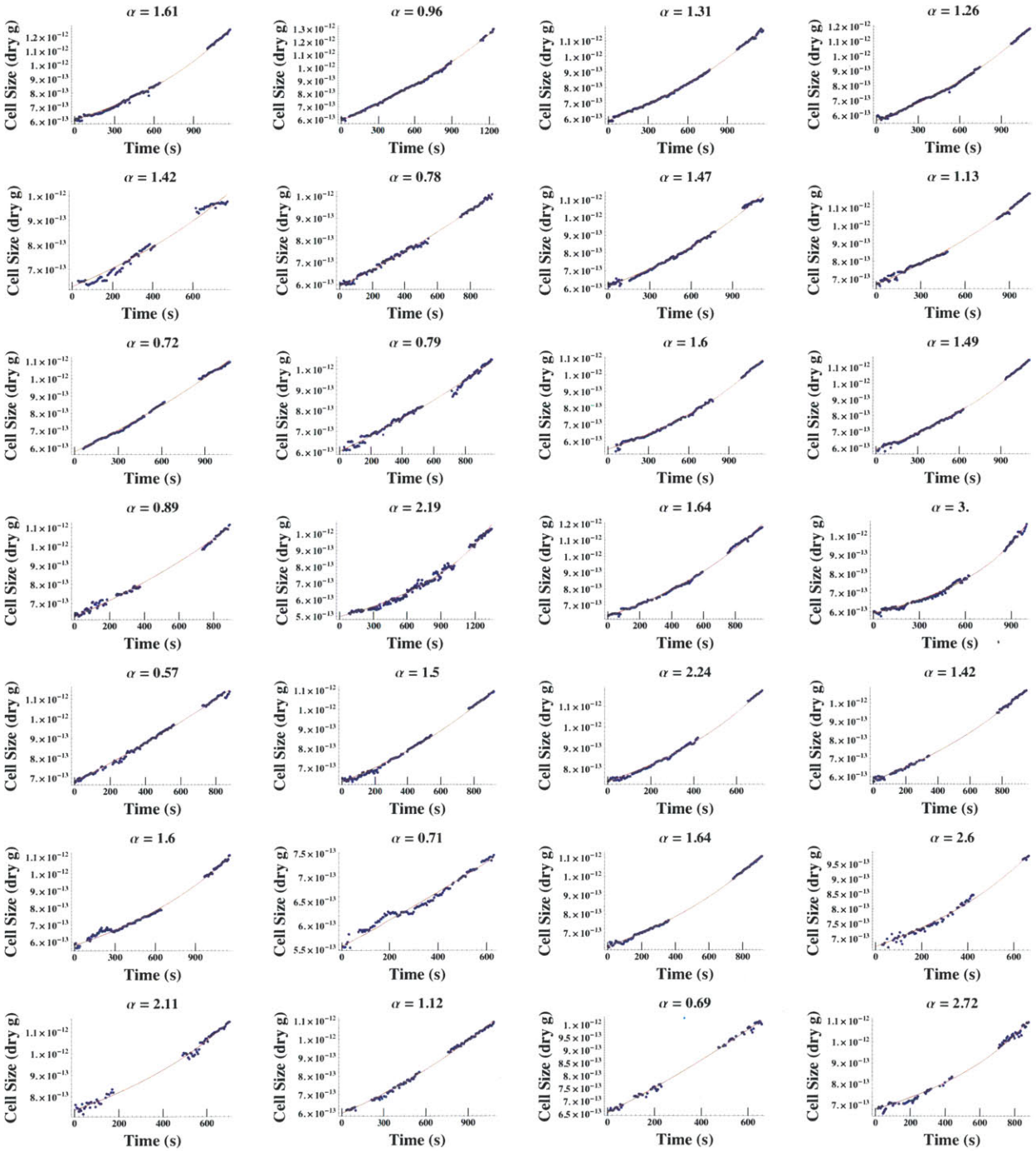


Figure A-4: *E. coli* growth trajectories. Data from [231, 232].

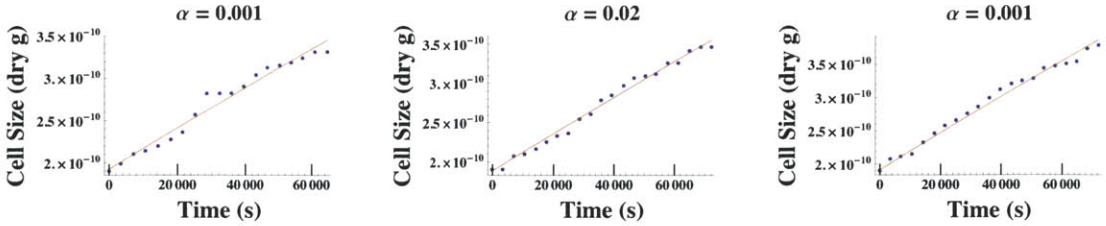


Figure A-5: *T. weissflogii* growth trajectories. Data from [201].

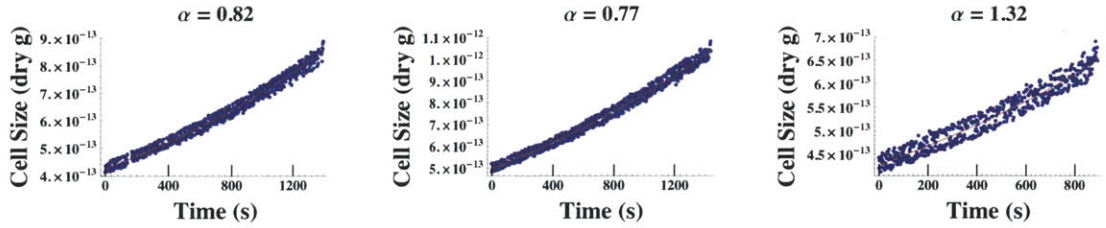


Figure A-6: *B. subtilis* growth trajectories. Data from [110].

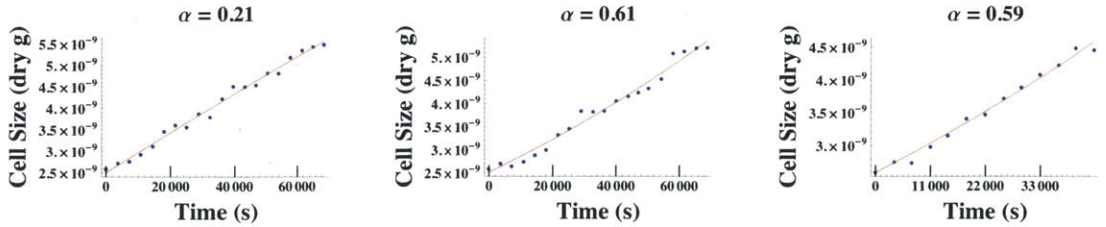


Figure A-7: *L. borealis* growth trajectories. Data from [201].

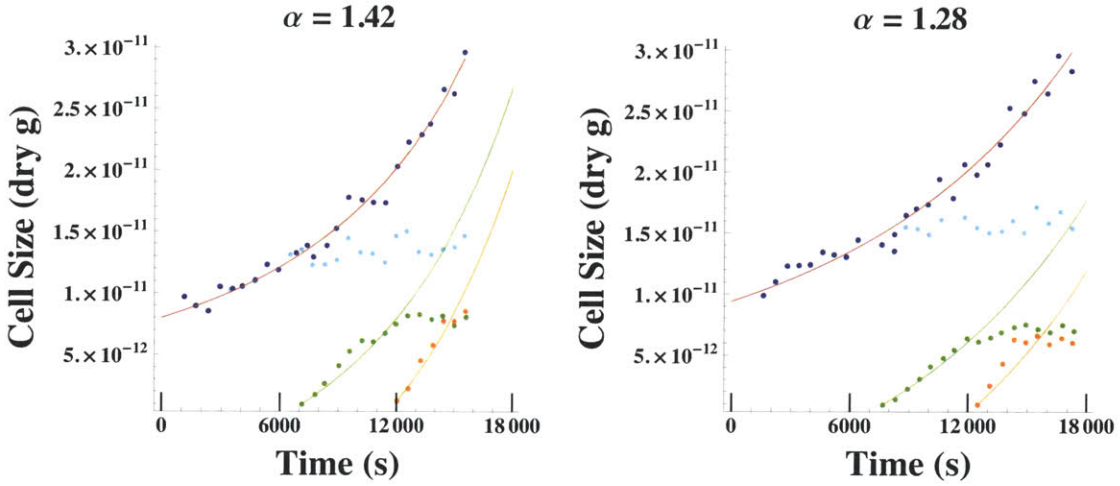


Figure A-8: Entire *C. albicans* complex growth trajectory. This mirrors Figure 2-1 D where the red curve is the fitted growth trajectory for the entire complex of budding yeast (dark blue points) which initially starts from a single mother bud (cyan points). The significance of the mother bud is that its growth slows with the formation of the first (green points) and second (orange points) daughter buds whose fitted trajectories are also shown using the same color scheme. The fitted exponent is for the entire complex. Data is from ref. [119].

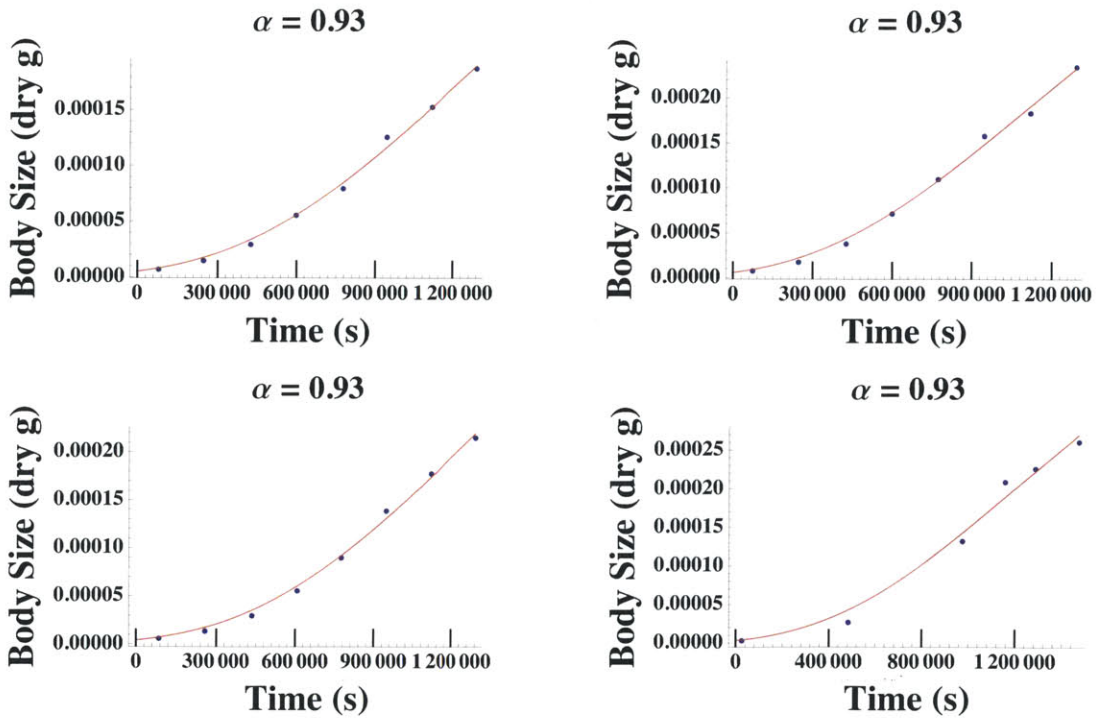


Figure A-9: *C. pacificus* growth trajectories. Data from [281].

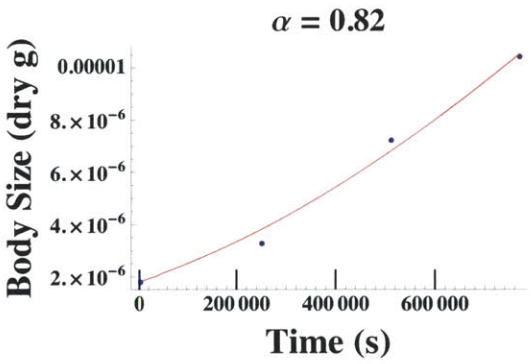


Figure A-10: *Pseudocalanus sp.* growth trajectories. Data from [281].

A.14 Compiled estimates of b and γ

In Table A.6 we provide our compilation for cell mass, b, and γ. The γ values are calculated from b and μ_{max}. Temperature normalizations have been carried out for b and μ_{max}. A digital spreadsheet of Table A.6 is included as part of the supplementary material of ref. [131] and it should be noted that this copy includes additional information such as the calculated Y and P values, culture conditions, the temperature of each reference, and notes on several of the calculations.

Table A.6: Compiled values for cell mass, b, and γ

Species Name	Mass (g)	b (s^{-1})	μ_{max} (s^{-1})	$\bar{\gamma}$	ref. for mass	ref. for b	ref. for μ_{max}
		normalized to 20° C	maximum specific growth rate normalized to 20°	average fraction of metabolism devoted to growth			
Acromonas punctata		3.52×10^{-6}				[118]	
Acromonas punctata		8.45×10^{-6}				[118]	
Arthrobacter globiformis		1.5×10^{-6}				[118]	
Azotobacter chroococcum	1.2×10^{-11}	7.12×10^{-6}	1.61×10^{-4}	0.958	[69]	[118]	[69]
Azotobacter chroococcum	1.2×10^{-11}	6.56×10^{-6}	1.61×10^{-4}	0.961	[69]	[118]	[69]
Azotobacter chroococcum	1.2×10^{-11}	2.03×10^{-6}	1.61×10^{-4}	0.988	[69]	[118]	[69]
Azotobacter vinilandii		4.87×10^{-6}				[118]	
Azotobacter vinilandii		8.74×10^{-6}				[118]	
Bacillus amyloliquefaciens		1.62×10^{-6}				[273]	
Bacillus cereus	3.7×10^{-12}	1.32×10^{-6}	2.9×10^{-4}	0.995	[69]	[270]	[69]
Bacillus cereus	3.7×10^{-12}	1.05×10^{-6}	2.9×10^{-4}	0.996	[69]	[35]	[69]
Bacillus coagulans		2.53×10^{-6}				[118]	
Bacillus licheniformis	$8.\times10^{-13}$	4.58×10^{-6}	4.73×10^{-5}	0.912	[69]	[99]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	6.88×10^{-6}	4.73×10^{-5}	0.873	[69]	[99]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.15×10^{-6}	4.73×10^{-5}	0.976	[69]	[99]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	4.7×10^{-6}	4.73×10^{-5}	0.91	[69]	[49]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	4.23×10^{-6}	4.73×10^{-5}	0.918	[69]	[99]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	4.23×10^{-6}	4.73×10^{-5}	0.918	[69]	[100]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.04×10^{-6}	4.73×10^{-5}	0.979	[69]	[100]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	7.71×10^{-6}	4.73×10^{-5}	0.86	[69]	[100]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.12×10^{-7}	4.73×10^{-5}	0.998	[69]	[49]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.09×10^{-6}	4.73×10^{-5}	0.977	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.03×10^{-6}	4.73×10^{-5}	0.979	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	9.64×10^{-7}	4.73×10^{-5}	0.98	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	8.99×10^{-7}	4.73×10^{-5}	0.981	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	8.17×10^{-7}	4.73×10^{-5}	0.983	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	7.07×10^{-7}	4.73×10^{-5}	0.985	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.91×10^{-6}	4.73×10^{-5}	0.961	[69]	[49]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.52×10^{-6}	4.73×10^{-5}	0.969	[69]	[100]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.52×10^{-6}	4.73×10^{-5}	0.969	[69]	[99]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.62×10^{-6}	4.73×10^{-5}	0.947	[69]	[49]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	9.62×10^{-7}	4.73×10^{-5}	0.98	[69]	[100]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.57×10^{-6}	4.73×10^{-5}	0.968	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.34×10^{-6}	4.73×10^{-5}	0.953	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.12×10^{-6}	4.73×10^{-5}	0.957	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.56×10^{-6}	4.73×10^{-5}	0.949	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.04×10^{-6}	4.73×10^{-5}	0.959	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.94×10^{-6}	4.73×10^{-5}	0.961	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.85×10^{-6}	4.73×10^{-5}	0.962	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.74×10^{-6}	4.73×10^{-5}	0.965	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.21×10^{-6}	4.73×10^{-5}	0.955	[69]	[101]	[69]

Species Name	Mass (g)	b (s^{-1})	μ_{max} (s^{-1})	$\bar{\gamma}$	ref. for mass	ref. for b	ref. for μ_{max}
		normalized to 20° C	maximum specific growth rate normalized to 20°	average fraction of metabolism devoted to growth			
Bacillus licheniformis	$8.\times10^{-13}$	1.46×10^{-6}	4.73×10^{-5}	0.97	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.28×10^{-6}	4.73×10^{-5}	0.954	[69]	[100]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.14×10^{-6}	4.73×10^{-5}	0.957	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	2.05×10^{-6}	4.73×10^{-5}	0.958	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.07×10^{-6}	4.73×10^{-5}	0.978	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.96×10^{-6}	4.73×10^{-5}	0.96	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.85×10^{-6}	4.73×10^{-5}	0.962	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.65×10^{-6}	4.73×10^{-5}	0.966	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.58×10^{-6}	4.73×10^{-5}	0.968	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.67×10^{-6}	4.73×10^{-5}	0.966	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.69×10^{-6}	4.73×10^{-5}	0.966	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.21×10^{-6}	4.73×10^{-5}	0.975	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.74×10^{-6}	4.73×10^{-5}	0.965	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	1.74×10^{-6}	4.73×10^{-5}	0.965	[69]	[101]	[69]
Bacillus licheniformis	$8.\times10^{-13}$	3.31×10^{-6}	4.73×10^{-5}	0.935	[69]	[273]	[69]
Bacillus megaterium	2.28×10^{-12}	7.31×10^{-6}	9.74×10^{-5}	0.93	[269]	[118]	[225]
Bacillus megaterium	2.28×10^{-12}	4.68×10^{-6}	9.74×10^{-5}	0.954	[269]	[118]	[225]
Bacillus pumilus		1.96×10^{-6}				[273]	
Bacillus subtilis	6.3×10^{-13}	8.61×10^{-6}	1.19×10^{-4}	0.933	[110]	[245]	[288]
Bacillus subtilis	6.3×10^{-13}	1.14×10^{-5}	1.19×10^{-4}	0.912	[110]	[245]	[288]
Bacillus subtilis	6.3×10^{-13}	2.33×10^{-6}	1.19×10^{-4}	0.981	[110]	[273]	[288]
Bacillus subtilis	6.3×10^{-13}	2.6×10^{-6}	1.19×10^{-4}	0.979	[110]	[245]	[288]
Bacillus subtilis	6.3×10^{-13}	4.11×10^{-6}	1.19×10^{-4}	0.967	[110]	[245]	[288]
Bacillus subtilis	6.3×10^{-13}	3.58×10^{-6}	1.19×10^{-4}	0.971	[110]	[273]	[288]
Bacillus subtilis	6.3×10^{-13}	2.9×10^{-6}	1.19×10^{-4}	0.976	[110]	[273]	[288]
Beneckea natriegens		3.81×10^{-6}				[118]	
Beneckea natriegens		3.36×10^{-5}				[118]	
Cellulomonas LC-1O		3.24×10^{-6}				[270]	
Corynebacterium glutamicum	6.19×10^{-13}	2.4×10^{-6}	7.23×10^{-5}	0.968	[69]	[62]	[69]
Corynebacterium glutamicum	6.19×10^{-13}	1.03×10^{-6}	7.23×10^{-5}	0.986	[69]	[62]	[69]
Escherichia coli	1.2×10^{-12}	2.01×10^{-6}	4.17×10^{-4}	0.995	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	3.5×10^{-6}	4.17×10^{-4}	0.992	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	3.65×10^{-6}	4.17×10^{-4}	0.991	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	8.31×10^{-7}	4.17×10^{-4}	0.998	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	2.81×10^{-5}	4.17×10^{-4}	0.937	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	9.29×10^{-6}	4.17×10^{-4}	0.978	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	2.57×10^{-6}	4.17×10^{-4}	0.994	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	1.33×10^{-6}	4.17×10^{-4}	0.997	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	3.62×10^{-6}	4.17×10^{-4}	0.991	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	1.47×10^{-5}	4.17×10^{-4}	0.966	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	1.67×10^{-5}	4.17×10^{-4}	0.962	[69]	[118]	[69]
Escherichia coli	1.2×10^{-12}	2.05×10^{-5}	4.17×10^{-4}	0.953	[69]	[118]	[69]
Klebsiella aerogenes		3.49×10^{-6}				[118]	
Klebsiella aerogenes		3.13×10^{-6}				[118]	
Klebsiella aerogenes		6.5×10^{-6}				[118]	
Klebsiella aerogenes		3.44×10^{-6}				[118]	
Klebsiella aerogenes		3.44×10^{-6}				[118]	
Klebsiella aerogenes		7.72×10^{-6}				[118]	
Klebsiella aerogenes		4.28×10^{-6}				[118]	
Lactobacillus casei	1.9×10^{-12}	4.38×10^{-7}	1.37×10^{-4}	0.997	[69]	[285]	[69]
Lactococcus lactis	$2.\times10^{-13}$	1.08×10^{-5}	1.24×10^{-4}	0.92	[69]	[142]	[69]
Lactococcus lactis	$2.\times10^{-13}$	2.01×10^{-6}	1.24×10^{-4}	0.984	[69]	[142]	[69]
Lactococcus lactis	$2.\times10^{-13}$	7.54×10^{-5}	1.24×10^{-4}	0.622	[69]	[142]	[69]
Lactococcus lactis	$2.\times10^{-13}$	3.69×10^{-5}	1.24×10^{-4}	0.771	[69]	[227]	[69]
Lactococcus lactis	$2.\times10^{-13}$	3.48×10^{-5}	1.24×10^{-4}	0.781	[69]	[227]	[69]
Lactococcus lactis	$2.\times10^{-13}$	3.72×10^{-5}	1.24×10^{-4}	0.769	[69]	[227]	[69]
Lactococcus lactis	$2.\times10^{-13}$	2.81×10^{-5}	1.24×10^{-4}	0.815	[69]	[227]	[69]
Lactococcus lactis	$2.\times10^{-13}$	3.56×10^{-5}	1.24×10^{-4}	0.777	[69]	[227]	[69]
Lactococcus lactis	$2.\times10^{-13}$	2.85×10^{-6}	1.24×10^{-4}	0.978	[69]	[168]	[69]
Methylcoccus sp.		1.63×10^{-6}				[118]	
Methylcoccus sp.		8.89×10^{-7}				[118]	

Species Name	Mass (g)	b (s^{-1})	μ_{max} (s^{-1})	$\bar{\gamma}$	ref. for mass	ref. for b	ref. for μ_{max}
		normalized to 20° C	maximum specific growth rate normalized to 20°	average fraction of metabolism devoted to growth			
Metylmonas methanolica		1.2×10^{-5}				[118]	
Metylmonas methanolica		1.31×10^{-5}				[118]	
Micrococcus denitrificans		1.73×10^{-6}				[118]	
Micrococcus denitrificans		2.32×10^{-6}				[118]	
Micrococcus denitrificans		6.41×10^{-7}				[118]	
Micrococcus denitrificans		1.27×10^{-7}				[118]	
Micrococcus denitrificans		3.94×10^{-6}				[118]	
Micrococcus denitrificans		5.52×10^{-6}				[118]	
Micrococcus denitrificans		2.19×10^{-6}				[118]	
Micrococcus denitrificans		2.12×10^{-6}				[118]	
mixed bacterial culture		1.08×10^{-6}				[118]	
mixed bacterial culture		5.93×10^{-7}				[118]	
mixed bacterial culture		4.86×10^{-6}				[118]	
mixed bacterial culture		4.71×10^{-6}				[118]	
mixed culture bacterium		2.04×10^{-6}				[118]	
mixed culture bacterium		2.44×10^{-6}				[118]	
mixed culture bacterium		3.75×10^{-6}				[118]	
mixed culture bacterium		5.72×10^{-6}				[118]	
mixed culture bacterium		8.1×10^{-6}				[118]	
Neisseria meningitidisB	$3.\times10^{-13}$	1.33×10^{-6}	5.13×10^{-5}	0.975	[69]	[11]	[69]
Pseudomonas 1		1.94×10^{-6}				[118]	
Pseudomonas aeruginosa	$6.\times10^{-13}$	1.25×10^{-6}	1.03×10^{-4}	0.988	[69]	[60]	[69]
Pseudomonas C		4.27×10^{-6}				[118]	
Pseudomonas I		1.58×10^{-6}				[118]	
Pseudomonas I35		2.83×10^{-6}				[118]	
Pseudomonas I35		2.85×10^{-6}				[118]	
Pseudomonas methyltropha		6.46×10^{-6}				[118]	
Pseudomonas oxalaticus		7.58×10^{-7}				[118]	
Pseudomonas oxalaticus		2.04×10^{-6}				[118]	
Rhizobium leguminosarum	$6.\times10^{-13}$	3.71×10^{-7}	2.4×10^{-5}	0.985	[69]	[224]	[69]
Rhizobium leguminosarum	$6.\times10^{-13}$	6.81×10^{-7}	2.4×10^{-5}	0.972	[69]	[224]	[69]
Rhodopseudomonas sheperoides		8.37×10^{-7}				[118]	
Rhodopseudomonas sheperoides		3.46×10^{-6}				[118]	
Streptococcus faecalis	$1.\times10^{-12}$	3.78×10^{-6}	1.37×10^{-4}	0.973	[69]	[198]	[69]
Streptococcus faecalis	$1.\times10^{-12}$	3.21×10^{-6}	1.37×10^{-4}	0.977	[69]	[198]	[69]
Streptococcus faecalis	$1.\times10^{-12}$	4.1×10^{-5}	1.37×10^{-4}	0.769	[69]	[198]	[69]
Candida Albicans	1.71×10^{-11}	2.48×10^{-7}	1.49×10^{-5}	0.984	[119]	[254]	[20]
Candida Albicans	1.71×10^{-11}	7.66×10^{-7}	1.49×10^{-5}	0.951	[119]	[254]	[20]
Candida boidinii	4.22×10^{-10}	1.63×10^{-6}	1.65×10^{-5}	0.91	[286]	[118]	[210]
Candida lipolytica	4.22×10^{-10}	8.73×10^{-7}	4.15×10^{-5}	0.979	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	6.8×10^{-6}	4.15×10^{-5}	0.859	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	6.6×10^{-6}	4.15×10^{-5}	0.863	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	3.96×10^{-6}	4.15×10^{-5}	0.913	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	1.08×10^{-6}	4.15×10^{-5}	0.975	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	7.99×10^{-6}	4.15×10^{-5}	0.838	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	1.23×10^{-5}	4.15×10^{-5}	0.772	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	1.21×10^{-5}	4.15×10^{-5}	0.775	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	7.2×10^{-6}	4.15×10^{-5}	0.852	[158]	[118]	[244]
Candida lipolytica	4.22×10^{-10}	1.11×10^{-5}	4.15×10^{-5}	0.789	[158]	[118]	[244]
Candida utilis	$1.\times10^{-11}$	2.09×10^{-6}	4.46×10^{-5}	0.955	[40, 50]	[118]	[50]
Candida utilis	$1.\times10^{-11}$	1.88×10^{-6}	4.46×10^{-5}	0.96	[40, 50]	[118]	[50]
Candida utilis	$1.\times10^{-11}$	1.58×10^{-6}	4.46×10^{-5}	0.966	[40, 50]	[118]	[50]
Candida utilis	$1.\times10^{-11}$	3.52×10^{-6}	4.46×10^{-5}	0.927	[40, 50]	[118]	[50]
Candida utilis	$1.\times10^{-11}$	2.5×10^{-6}	4.46×10^{-5}	0.947	[40, 50]	[118]	[50]
Candida utilis	$1.\times10^{-11}$	5.44×10^{-6}	4.46×10^{-5}	0.891	[40, 50]	[118]	[50]
Chlamydomonas reinhardtii	3.69×10^{-11}	5.29×10^{-7}	1.74×10^{-5}	0.97	[36]	[57]	[290]
Chlamydomonas reinhardtii	3.69×10^{-11}	1.41×10^{-6}	1.74×10^{-5}	0.925	[36]	[137]	[290]
Chlorella ellipsoidea	3.4×10^{-11}	7.15×10^{-7}	1.93×10^{-5}	0.964	[94]	[113]	[94]
Chlorella pyrenoidosa	1.25×10^{-11}	2.19×10^{-6}	1.66×10^{-5}	0.884	[114]	[113]	[111]
Chlorella regularis	5.22×10^{-12}	8.12×10^{-7}	1.96×10^{-5}	0.96	[123, 243]	[118]	[243]
Chlorella regularis		8.12×10^{-7}				[118]	

Species Name	Mass (g)	b (s^{-1}) normalized to 20° C	μ_{max} (s^{-1}) maximum specific growth rate normalized to 20°	$\bar{\gamma}$ average fraction of metabolism devoted to growth	ref. for mass	ref. for b	ref. for μ_{max}
Chlorella regularis	5.22×10^{-12}	1.56×10^{-6}	1.96×10^{-5}	0.926	[123, 243]	[118]	[243]
Chlorella regularis		1.56×10^{-6}				[118]	
Chlorella sorokiniana	2.65×10^{-12}	1.52×10^{-6}	1.87×10^{-5}	0.925	[178]	[113]	[178]
Chlorella vulgaris	4.68×10^{-12}	4.2×10^{-6}	1.55×10^{-5}	0.787	[307]	[145]	[307]
Dunaliella tertiolecta	2.13×10^{-11}	6.3×10^{-6}			[87]	[87]	
Hansenula polymorpha	2.65×10^{-12}	1.27×10^{-6}	1.66×10^{-5}	0.929	[280]	[118]	[150]
Hansenula polymorpha	2.65×10^{-12}	1.2×10^{-6}	1.66×10^{-5}	0.932	[280]	[118]	[150]
Hansenula polymorpha	2.65×10^{-12}	1.81×10^{-6}	1.66×10^{-5}	0.902	[280]	[118]	[150]
Hansenula polymorpha	2.65×10^{-12}	2.57×10^{-6}	1.66×10^{-5}	0.866	[280]	[118]	[150]
id		1.28×10^{-6}	8.55×10^{-6}	0.87		[113]	[113]
Isochrysis galbana	1.26×10^{-11}	7.45×10^{-7}			[85]	[85]	
Ochromonas sp.	1.56×10^{-10}	5.09×10^{-6}	$5.\times10^{-5}$	0.908	[69]	[95]	[69]
Paecilomyces varioti		9.75×10^{-6}				[118]	
Pen. chrysogenum		1.74×10^{-6}				[118]	
Pen. chrysogenum		2.09×10^{-6}				[118]	
Pen. chrysogenum		6.42×10^{-6}				[118]	
Pen. chrysogenum		6.71×10^{-6}				[118]	
Prorocentrum micans	6.57×10^{-10}	5.31×10^{-7}			[85]	[85]	
Saccharomyces cerevisiae	8.25×10^{-12}	1.02×10^{-6}	4.1×10^{-5}	0.976	[126]	[118]	[27]
Saccharomyces cerevisiae	8.25×10^{-12}	2.17×10^{-6}	4.1×10^{-5}	0.95	[126]	[118]	[27]
Saccharomyces cerevisiae	8.25×10^{-12}	6.08×10^{-6}	4.1×10^{-5}	0.871	[126]	[233]	[27]
Scenedesmus obliquus	3.57×10^{-11}	1.64×10^{-6}	9.86×10^{-6}	0.858	[257]	[113]	[163]
Skeletonema costatum	1.98×10^{-11}	-1.26×10^{-6}			[87]	[87]	
Thalassiosira weisflogii	2.25×10^{-10}	4.67×10^{-7}	1.01×10^{-5}	0.956	[85]	[85]	[271]
Trichoderma viride	2.16×10^{-10}	8.93×10^{-7}	1.43×10^{-5}	0.941	[235]	[118]	[41]
Trichoderma viride	2.16×10^{-10}	4.47×10^{-6}	1.43×10^{-5}	0.762	[235]	[118]	[41]
Trichoderma viride	2.16×10^{-10}	1.2×10^{-6}	1.43×10^{-5}	0.923	[235]	[118]	[41]
Trichoderma viride	2.16×10^{-10}	5.96×10^{-6}	1.43×10^{-5}	0.706	[235]	[118]	[41]
Brachionus calyciflorus	1.25×10^{-7}	2.87×10^{-6}	1.01×10^{-5}	0.778	[33]	[33]	[33]

Appendix B

Appendix for Chapter 3

B.1 Estimates of rescaled parameters

In Chapter 3 we present a model for the physical diffusion and biological consumption of oxygen within the colony. We showed that this system is greatly simplified by nondimensionalization. The key parameters in this nondimensionalization are

$$t_{fac} = \frac{a\mu_{max}}{k_s Y}, \qquad x_{fac} = \left(\frac{a\mu_{max}}{k_s Y D}\right)^{1/2}, \quad \text{and} \quad k_s \tag{B.1}$$

where t_{fac} rescales time, x_{fac} rescales space, and k_s rescales the concentration of oxygen. Once the system is rescaled there there is one free parameter which summarizes the maintenance term: $g = YP/\mu_{max}$. The literature provides estimates for all of these key parameters (Table B.1). We combined every combination of these parameters for each of x_{fac}, k_s, and g and use the mean or median value as a best estimate. Doing this we find that the mean value of x_{fac} is $68,135$ s^{-1} and the median is $50,392$ s^{-1}. For k_s the mean estimate is 0.29 g oxygen m^{-3} and the median is 0.20 g oxygen m^{-3}. The mean estimate for g is 0.043 and the median is 0.037. Note that t_{fac} is not critical to our analysis where we examine steady states, but it can be found easily from the parameters presented here.

Parameter	Symbol	Values	Refs.
Half Saturation Constant	k_s		
		.25 g oxygen m^{-3}	[153]
		.045 g oxygen m^{-3}	[267]
		.4 g oxygen m^{-3} (bottom of range)	[25]
		1.18 g oxygen m^{-3} (top of range)	[25]
Maximum growth rate	μ_{max}		
		1.11×10^{-4} s^{-1}	[13]
		7.89×10^{-5} s^{-1}	[255]
		2.78×10^{-5} s^{-1}	[255]
		6.11×10^{-5} s^{-1}	[267]
		5.28×10^{-5} s^{-1}	[25]
		8.89×10^{-5} s^{-1}	[25]
		2.22×10^{-4} s^{-1}	[266]
Yield Coefficient	Y		
		.65 g cells · g $oxygen^{-1}$	[267]
		.85 g cells · g $oxygen^{-1}$	[266]
		.635 g cells · g $oxygen^{-1}$	[25]
Maintenance Coefficient	P		
		3.89×10^{-6} g oxygen · g cells $^{-1}$ · s^{-1}	[25]
Diffusivity	D		
		1.76×10^{-9} m^2 s^{-1}	[153]
		1.53×10^{-9} m^2 s^{-1}	[266]
Cell Density (g cells · m^{-3})	a		
		12000 g m^{-3}	[266]

Table B.1: Values for key biological and physical parameters.

Appendix C

Appendix for Chapter 5

C.1 Relationships within the height limitation framework

Our framework for predicting height is based on a tree's ability to collect sufficient water and sunlight to meet its basal metabolic needs without exceeding the availability of these resources. These requirements can be summarized by inequalities relating the required flow rate Q_0, the potential evaporative flow rate Q_e, and the available flow rate Q_p:

$$Q_p - Q_0 \geq 0 \tag{C.1}$$

$$Q_p - Q_e \geq 0 \tag{C.2}$$

$$Q_e - Q_0 \geq 0 \tag{C.3}$$

Eq. C.1 ensures that the tree must receive enough water to maintain its basal metabolic flow. It is through this constraint that we couple size to water resources. Eq. C.2 states that precipitation must meet or exceed evaporative flow. Eq. C.3 ensures that the energy a tree receives from its environment, which is translated into evaporative flow, meets its basal metabolic needs. These statements are summarized by a single condition:

$$Q_p \geq Q_e \geq Q_0. \tag{C.4}$$

Thus, Q_0 and Q_p set the boundaries of acceptable flow in an environment. Maximum tree height can then be predicted by finding the largest tree for which this relationship holds. That is, our strategy searches for trees that meet their metabolic needs without exceeding their water or solar resources.

C.2 Scaling relationships

Recent work has described many allometric relationships in terms of the branching architecture of trees [292, 291]. This work focuses on the hierarchy and hydrodynamic optimization of branching bundles of vascular tubes, where the mass M of a tree is related to the number of branching generations, N, by $N \propto \ln M$ [292]. This framework also describes the scaling of branch lengths and radii which, along with mass, allow for conversion between various observed allometric relationships. For a tree of height h with n daughter branches at each branching generation, k, whose radius and length are given by r_k and l_k, respectively, it can be shown that

$$h \approx \frac{l_0}{1-n^{-1/3}}, \quad \frac{r_k}{r_N} = n^{(N-k)(a/2)}, \quad \text{and} \quad \frac{l_k}{l_N} = \left(\frac{r_k}{r_N}\right)^{2/3a} \tag{C.5}$$

where, typically, $n = 2$, and $a \approx 1$ is the branching exponent [292]. For terminal units, $r_N \approx 0.4$ mm, and $l_N \approx 4$ cm [292]. Note that plant height can also be expressed as

$$h \approx \frac{n^{N/3} l_N}{1-n^{-1/3}} \tag{C.6}$$

$$h \propto n^{N/3}. \tag{C.7}$$

Many observed allometric relationships are given in terms of the basal stem diameter, D. In the relationships above $D = 2r_0$, and thus $D \propto h^{3/2}$, which agrees with the findings of reference [191] ($h \propto D^{2/3}$ for large trees). This allows us to re-write scaling relationships for stem diameter in units of tree height. For example, the basal flow rate of fluid through a tree given by

$$Q_0 = \beta_1 D^{\eta_1} \tag{C.8}$$

[80] becomes

$$Q_0 = \beta_1 \left(2n^{Na/2} r_N\right)^{\eta_1} \tag{C.9}$$

$$\equiv \beta_2 h^{\eta_2} \tag{C.10}$$

with $\eta_2 = 3\eta_1 a/2 \approx 2.7$ and $\beta_2 = \beta_1 \left[2r_N \left(\frac{1-n^{-1/3}}{l_N}\right)^{3a/2}\right]^{\eta_1} \approx 9.2 \times 10^{-7}$ (liter day^{-1} cm$^{-\eta_2}$; for h in cm), given the parameters shown in Table 1. It should be noted that these values are found from measurements of trees *in situ* which may not be operating at the absolute basal metabolism. Future work should focus on determining the minimal requirement of a tree and this will likely adjust the value of β_1 but not the exponent η_1. This branching architecture also defines several geometric characteristics of the canopy and root systems required for calculating Q_p and Q_e.

C.3 Relationships governing the available flow Q_p

The available flow rate is determined by the incoming rate of precipitation, p_{inc}, along with the capture area and efficiency, γ, of the roots. We define the capture area in terms of the radial extent of the root system, r_{root}, so that

$$Q_p = \gamma \pi r_{root}^2 p_{inc}. \tag{C.11}$$

We find that on average Q_p matches Q_0 for the tallest trees taking $\gamma = 1/3$, that is, solving for $Q_p(h_{max}) = Q_0(h_{max})$ yields a mean value of $\gamma = 1/3$, and this is the value that we use in our model. Calibrating γ in this way allows us to later solve for height using only the single limitation of Q_p. Adding variations in soil type and hydrology to calculate γ locally is an important area of future research.

The root vascular system is assumed to obey similar constraints as the canopy except that gravitational biomechanical limitations, important for canopy branches, are not important for roots. Roots may exhibit a different architecture, or overall shape, from that of the canopy, but, nevertheless, obey similar hydrodynamic and minimization constraints as the above-ground branching network. Branches and stems grow

in diameter through a process of older vascular tubes becoming heartwood and new vascular tubes forming new sapwood. Since vascular tubes run the entire length of the tree, the above- and below-ground networks are not independent. As such, branch tube formation and mortality (above-ground sapwood and heartwood formation) impose constraints on root structure. This is supported by data which show that the branching exponent a for the root system has a mean value close to unity at each branching generation, similar to that for above-ground branches [242]; (our reanalysis of data for the two species presented in Figure 1 of ref. [242] gives $a = 1.03 \pm .33$ and $a = 1.12 \pm .45$ for $n = 2$). Combining all of these constraints results in the scaling and topology of the root system being essentially identical to that of the canopy, although the detailed architectures may be different. Thus, the total mass of the roots M_R should scale isometrically with the mass of the stems M_S, a prediction supported by the data: $M_R = \beta_3 M_S$ where $\beta_3 = 0.423 \pm 0.02$ (dimensionless) [189, 191].

For optimal access to moisture the roots maximize the volume of soil occupied. The radial extent of the roots should therefore be equal to the longest tubes found in the root system. For stems, the length of a tube running from the base of the trunk to the leaves is given by $L_S \equiv h = \beta_4 M_S^{1/4}$ [292] so, given the above argument for roots, it follows that $L_R \equiv r_{root} = \beta_4 M_R^{1/4}$. Thus the radial extent of the roots is given by

$$r_{root} = \beta_3^{1/4} h. \tag{C.12}$$

Later we relax these assumptions regarding the scaling of the root system with overall tree height and show how optimizing the root scaling exponent can significantly reduce the error between observed maximum tree height and our predictions.

C.4 Relationships governing the evaporative flow Q_e

Our relationship for the evaporative flow rate of a tree is determined by the interaction of the canopy with the local meteorological conditions. Here we will analyze the

energy budget of the canopy, and in doing so, determine how the allometric properties of the canopy govern various heat fluxes.

C.4.1 Energy Budget

Our analysis of the energy budget follows ref. [51]. Conservation of energy requires:

$$R_{abs} - La_g - Ha_j - \lambda E a_f = 0 \tag{C.13}$$

where R_{abs} is the total rate at which radiation is absorbed by the canopy (W) and each of the other terms is an energy flux (W m^{-2}) away from the canopy multiplied by an effective area, $a_{i=g,j,f}$, over which that flux occurs (the scaling of these areas are described below). L is the emitted thermal radiation, H is the sensible heat loss, λE is the latent heat loss with λ being the latent heat of evaporation (J mol^{-1}) and E the evaporative molar flux (mol m^{-2} s^{-1}). In the definitions below it can be seen that each of the flux terms depends on the leaf temperature, T_L (in degrees Celsius; temperatures in degrees Kelvin will be denoted using boldface $\mathbf{T}$). It is generally desirable to eliminate T_L from the system of equations in order to focus on common measurements such as the local air temperature. The Penman-Monteith method gives an approximation of T_L by linearizing each component of the energy budget in terms of $\Delta T = T_A - T_L$ and solving for T_L, where T_A is the air temperature.

Assuming $\Delta T << T_L$ and $\Delta T << T_A$ we can approximate the emitted thermal radiation as

$$\begin{aligned} L &= \epsilon_l \sigma \mathbf{T}_L^4 && \text{(C.14)} \\ &\approx \epsilon_l \sigma \mathbf{T}_A^4 + c_p g_r \cdot (T_L - T_A) && \text{(C.15)} \\ &\equiv g_2 + g_1 \cdot (T_L - T_A) && \text{(C.16)} \end{aligned}$$

where ϵ_l is the emmisivity of the leaf, σ (W m^{-2} K^{-4}) is the Stefan-Boltzmann constant, c_p (J mol^{-1} C^{-1}) is the specific heat of air, and $g_r = 4\epsilon_l \sigma \mathbf{T}_a^3 / c_p$ (mol m^{-2} s^{-1}) is the radiative conductance (we will discuss various conductances below).

The sensible heat loss is similarly defined in terms of ΔT and is given by

$$H = c_p g_{Ha} \cdot (T_L - T_A) \tag{C.17}$$

$$\equiv j_1 \cdot (T_L - T_A) \tag{C.18}$$

where g_{Ha} (mol m^{-2} s^{-1}) is the heat conductance of air [51].

For the latent heat loss we make the approximation

$$\lambda E = \lambda g_v \frac{e_s(T_L) - e_s(T_A)}{p_a} \tag{C.19}$$

$$\approx \lambda g_v \frac{1}{p_a} \frac{de_s(T_A)}{dT_A} \cdot (T_L - T_A) + \lambda g_v \frac{D_v}{p_a} \tag{C.20}$$

$$\equiv f_1 \cdot (T_L - T_A) + f_2 \tag{C.21}$$

where g_v (mol m^{-2} s^{-1}) is the vapor conductance, T_D is the dew point temperature, p_a is the air pressure (kPa), and $D_v \equiv e_s(T_A) - e_s(T_D)$ (kPa) is the vapor pressure deficit, and

$$e_s(T_A) = b_1 \exp\left(\frac{b_2 T_A}{b_3 + T_A}\right) \tag{C.22}$$

is the Tetans formula for saturation vapor pressure (kPa) with $b_1 = .611$ (kPa), $b_2 = 17.502$ (dimensionless), and $b_3 = 240.97$ (°C) [51].

Solving this system of equations we obtain

$$T_L - T_A = \frac{R_{abs} - f_2 a_f - g_2 a_g}{f_1 a_f + g_1 a_g + j_1 a_j}, \tag{C.23}$$

which, using the approximation in Eq. C.21, leads to an equation for the evaporative flux

$$E = \frac{f_1}{\lambda}\left[\frac{R_{abs} - f_2 a_f - g_2 a_g}{f_1 a_f + g_1 a_g + j_1 a_j}\right] + \frac{f_2}{\lambda}. \tag{C.24}$$

The total volume flow rate, Q_e, is obtained by multiplying E_{can} by the area a_f, and then converting from a molar rate to a volume flow rate. Thus $Q_e = a_f E \mu_w / \rho_w$ given the molar mass, μ_w (kg mol^{-1}), and density, ρ_w (kg m^{-3}), of water. This can be

written in full as

$$Q_e = a_f \frac{\mu_w}{\rho_w} \left(\frac{f_1}{\lambda} \left[\frac{R_{abs} - f_2 a_f + g_2 a_g}{f_1 a_f + g_1 a_g + j_1 a_j} \right] + \frac{f_2}{\lambda} \right). \tag{C.25}$$

C.4.2 Allometric canopy geometry and solar radiation

In this section we derive the rate of absorbed solar radiation, R_{abs}, a necessary component of Q_e. This absorption rate represents the relationship between available solar resources, plant structure, and, ultimately, tree height.

The collection of solar radiation by a plant can be described by coefficients for canopy transmission, τ, canopy reflection, ξ_c, and soil reflection ξ_s all of which can be encapsulated as the canopy absorption coefficient

$$\alpha_{can} = 1 - \xi_c - (1 - \xi_s)\,\tau \tag{C.26}$$

[173]. A commonly used model [51, 173], which we find connects easily to the branching architecture discussed above, assumes that leaves are uniformly distributed over a spheroidal canopy of surface area, S_{can}, defined by the semi-axes $h'_{can} = h_{can}/2$ and r_{can}. In that case, the transmission coefficient is given by a classic absorption formula

$$\tau = e^{-\sqrt{\alpha} K(\psi) LAI} \tag{C.27}$$

[51, 173], where α is leaf absorptivity and LAI is the leaf area index, a measure of the total leaf area per unit ground below the canopy, which is given by

$$LAI = \frac{a_l n^N}{\pi r_{can}^2}. \tag{C.28}$$

The extinction coefficient, $K(\psi) = 2P_{can}/S_{can}$, describes the ratio of canopy projected area, P_{can}, to canopy surface area, S_{can}, and is a function of the solar zenith angle ψ. Given a spheroidal canopy, it follows that

$$P_{can} = \pi r_{can}(r_{can}\cos\theta + h'_{can}\sin\theta\tan - \psi), \tag{C.29}$$

and

$$S_{can} = 2\pi \left(r_{can}^2 + \frac{r_{can} \left(h'_{can}\right)^2 \arcsin \left[\frac{\sqrt{\left(h'_{can}\right)^2 - r_{can}^2}}{h'_{can}} \right]}{\sqrt{\left(h'_{can}\right)^2 - r_{can}^2}} \right), \tag{C.30}$$

[120, 173] where $\theta = \arctan\left(-h'_{can}/r_{can} \tan\psi\right)$ is the angle defining the point of ray tangency on the canopy surface and $h'_{can} = h_{can}/2$.

The transmission coefficient, τ, can be calculated by combining the above equations. The total solar radiation captured by the canopy is then given by

$$R_{abs} = \left[1 - \xi_c - (1 - \xi_s)\,\tau\right] P_{can} R_{inc}, \tag{C.31}$$

where R_{inc} (W m^{-2}) is the incoming radiation per unit area (normal to the ground), and ξ_c and ξ_s are described below.

C.4.3 Heat flux areas

Our canopy energy budget is derived by scaling up from the budget for a single leaf whose properties are assumed invariant across trees of different size. In this picture, changes in energy fluxes are predominantly due to allometric relationships for the canopy shape and number of leaves. Leaves are assumed to act in parallel, and thus the total area for sensible heat loss is $a_j = 2a_L$, where a_L is the total one-sided area of all the leaves on the tree, given by $a_L = a_l n^N \propto h^3 \propto M^{3/4}$ [292]. The evaporative flux, on the other hand, occurs over the total area of the stomatal openings in the leaves, $a_f = 2a_L \delta_s a_s$, where δ_s is the density of stomata on a leaf, and a_s is the area of a single stoma.

If L is the thermal radiative flux emanating from the tree, then a good approximation for the effective radiative surface area is the total one-sided leaf area, so $a_g \approx a_L$. Since $S_{can} \sim h^2$, whereas $a_L \sim h^3$, leaves will overlap on the canopy surface for sufficiently large trees so that the above approximation for a_g becomes less accurate. We tested other schemes for a_g, such as using S_{can}, or the minimum of S_{can} and a_L, and found similar results to the ones presented here.

C.4.4 Scaling of the canopy radius and height

In previous work, it has been shown that the canopy radius scales as $r_{can}^3 \propto r_0^2$ [192], or

$$r_{can} \propto n^{N/3} r_N^{2/3} \propto h \tag{C.32}$$

which follow from Eqs. C.5 and C.6. The scaling with height is consistent with empirical data where it has been shown that $r_{can} = \beta_5 h^{\eta_3}$, with $\eta_3 = 1.14$, and $\beta_5 = 35.24$ (cm m$^{-\eta_3}$; for h in meters) (see SI of ref. [83]).

Unlike r_{can}, we were unable to locate data characterizing the scaling of h_{can} with total tree height, h. Following ref. [82], we assume that the canopy scales isometrically so that $h_{can} = \beta_6 h$, where β_6 is a constant. A lower bound for the canopy height can be obtained from a configuration in which all of the branches of the tree emanate directly upwards from the top of the trunk leading to $h_{can} \geq h - l_0$; clearly, an upper bound is given by the total height of the tree: $h_{can} \leq h$. Since, $h \approx l_0 / \left(1 - n^{-1/3}\right)$, we obtain the bound

$$h \geq h_{can} > n^{-1/3} h \approx 0.79\, h \tag{C.33}$$

or

$$1 \geq \beta_6 > n^{-1/3} \approx 0.79 \tag{C.34}$$

for $n = 2$. Note, further, that, from data the diameter of the canopy, $d_{can} = 0.7 h^{1.1}$, indicating that, generally speaking, canopy height exceeds canopy diameter, $h_{can} > d_{can}$, suggesting that the canopy forms an approximate prolate spheroid. For the discussion presented in the text, we took $\beta_6 = 1$. Varying β_6 between 0.8 and 1 slightly shifted the mean of the distribution in Figure 5-1.

C.4.5 Canopy Radiation Coefficients

Above, we discussed the derivation of τ, which is one of the radiation coefficients involved in calculating the total radiation absorbed by a canopy. The other two coefficients correspond to the canopy and soil reflection, respectively, ξ_c and ξ_s. The soil reflectivity can be taken as a constant value which is given in Table 1, but the

canopy reflectivity is expected to change based on the size and scattering properties of the canopy. It has been shown that the canopy reflection, as a function of the leaf area index, LAI and extinction coefficient $K(\psi)$ is well approximated by

$$\xi_c = \frac{\xi_c^* + f e^{-2K(\psi)LAI}}{1 + \xi_c^* f e^{-2K(\psi)LAI}}, \tag{C.35}$$

where

$$f = \frac{\xi_c^* - \xi_s}{\xi_c^* \xi_s - 1} \tag{C.36}$$

and ξ_c^* is the deep canopy reflection coefficient, which is constant [173, 51]. For small LAI, Eq. C.35 is approximately the ground reflectivity ξ_s, and for large LAI, it is ξ_c^*.

Combining equations C.26, C.27, and C.35 it is possible to determine the total absorption coefficient of the canopy and this is the calculation used to produce the curve in Figure 5-4.

C.4.6 Canopy conductances

Each of the above terms depends on the conductances of heat, vapor, and radiation which are described below. Assuming that the canopy surface is exposed to forced laminar convection, the heat conductance in air is given by

$$g_{Ha} = \frac{0.664 \hat{\rho}_a D_H Re_a^{1/2} Pr_a^{1/3}}{d} \quad (\text{mol m}^{-2}\ \text{s}^{-1}) \tag{C.37}$$

where $Re_a = ud/\nu_a$ is the dimensionless Reynolds number for air, $Pr_a = \nu_a / D_H$ is the dimensionless Prandtl number for air, D_H ($\text{m}^2\ \text{s}^{-1}$) is the heat diffusivity, $\nu_a(T_a)$ ($\text{m}^2\ \text{s}^{-1}$) is the kinematic viscosity of air, u ($\text{m}\ \text{s}^{-1}$) is the wind speed, $d = 0.81\left(2\left(a_l/\pi\right)^{1/2}\right)$ (m) is the characteristic dimension of a circular leaf of area a_l (m), and $\hat{\rho}_a$ is the molar density of air given by the Boyle-Charles law as

$$\hat{\rho}_a = 44.6 \frac{p_a}{101.3} \frac{273.15}{\mathbf{T_A}} \quad (\text{mol m}^{-3}) \tag{C.38}$$

for Kelvin temperature $\mathbf{T_A}$ and air pressure $p_a = 101.3e^{-A/8200}$ (kPa) at altitude A (m) [51].

We assume that the canopy is in series with the atmospheric boundary layer, so $g_v = 1/\left(\frac{1}{g_{vL}} + \frac{1}{g_{va}}\right)$, where g_{vL} and g_{va} are, respectively, the vapor conductances for the leaf and atmosphere. Assuming forced laminar convection, as before, the boundary layer conductance is

$$g_{va} = \frac{0.664\hat{\rho}_a D_v Re_a^{1/2} Sc_a^{1/3}}{d} \quad (\text{mol m}^{-2}\ \text{s}^{-1}), \tag{C.39}$$

where $Sc_a = \frac{\nu_a}{D_v}$ is the dimensionless Schmidt number, and D_v ($\text{m}^2\ \text{s}^{-1}$) is the diffusivity for water vapor in air. The leaf conductance is calculated by considering the stomata as acting in parallel. We model a single stoma as a cylinder of depth z_s (m) in which case Fick's law gives

$$g_{vs} = \frac{\hat{\rho}_a D_v}{a_s(0)\int_{z_0}^{z_b} \frac{dz}{a_s(z)}} = \frac{\hat{\rho}_a D_v}{z_s} \quad (\text{mol m}^{-2}\ \text{s}^{-1}) \tag{C.40}$$

for the conductance of a single stoma. Summing over the entire leaf area, the leaf conductance is given by

$$g_{vL} = g_{vs} a_s \delta_s \quad (\text{mol m}^{-2}\ \text{s}^{-1}), \tag{C.41}$$

where a_s (m^2) is the area of the opening of a single stoma, and δ_s (stomata m^{-2}) is the stomatal density on a leaf.

C.4.7 Trait values

For each tree trait we examined the literature in order to gain a sense of the variation of each trait across many species, plant sizes, and environments and we picked values that were representative of that variation. For several of these values we were able to check for agreement with means from the TRY database [67] which is a collection of multiple databases containing global values for various plant traits across numerous

environments and species. For each trait there are several ways to analyze the distribution of trait values produced by the TRY database. The first is to take the straight mean, which we present in Table 1. The second is to recognize that most of these traits appear to be lognormally distributed in which case it is possible to logarithmically transform the data, find the transformed mean, and then take the exponential of that value. Formally this is the median assuming a lognormal distribution and these values are also given in Table 1. These two analyses typically provide a range of possible representative values for each trait, and the values that we used fall within this range.

C.5 Detailed examination of predicted vs. observed tree heights

C.5.1 Relationships between Q_0, Q_e, and Q_p for observed tallest trees

We tested the model by looking at observed maximum tree heights in different environments to check whether they do indeed adhere to the given resource constraints governing the required, evaporative, and available flow rates $Q_p(h_{obs}) \geq Q_e(h_{obs}) \geq Q_0(h_{obs})$. We calculate each of these quantities using observed maximum heights and measured environmental conditions. For trees which have grown to reach the upper bound on height in a given environment we would expect that $Q_p(h_{max}) = Q_e(h_{max}) = Q_0(h_{max})$. Figure C-1 A-C shows the resulting pair-wise relationship between each of these quantities. The derived data generally fall along the green line which represents equality between the two quantities being considered. In each plot the scaling of the two flows is independent of the choice of γ, but the overall intercept will depend on this value.

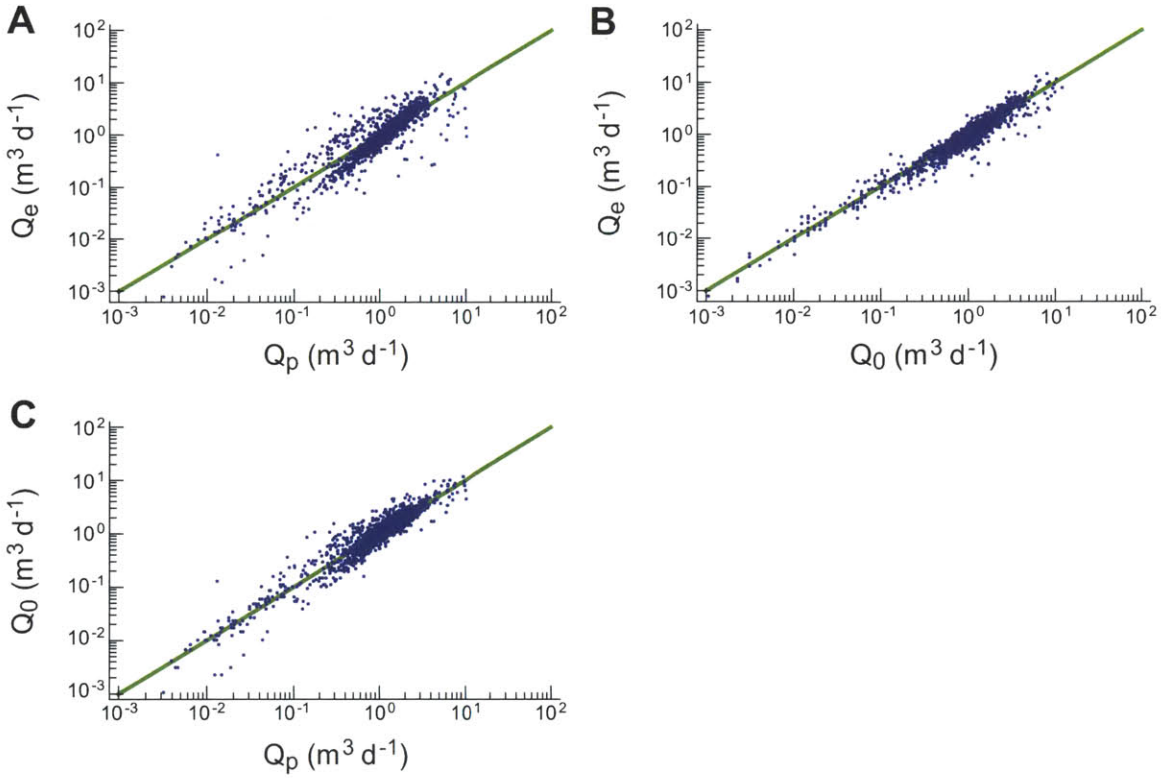


Figure C-1: Comparisons between each of the water fluxes. Each flux is calculated for an observed tallest tree. In each of the scatter plots the green curve is the one-to-one correspondence line. **(A)** The relationship between the available flow of water, Q_p, and the calculated evaporation, Q_E. **(B)** The relationship between the theoretical basal metabolism, Q_0, and Q_E. **(C)** Q_p vs. Q_0.

C.5.2 Systematic deviations within the model predictions

The error between predictions and observations, as shown in Figure 5-1 C of the main text, is slightly bimodal with the secondary mode corresponding to situations where we *under*-predict tree height. Most trees in the under-prediction mode are found in mountainous regions (elevations greater than 1000m) where there is often high spatial variability in precipitation. Thus the interpolated meteorology may not be representative of the environment experienced at the tree location. We tested this hypothesis by comparing precipitation estimates from the Parameter-elevation Regression on Independent Slopes Model (PRISM) [66, 218] to estimates from the North American Regional Reanalysis (NARR)[169, 10]. Large discrepancies between the two databases may indicate either differences in the modeling approach used to assimilate individual station data or high spatial variability in the meteorology. Figure C-2 A shows the distribution of the discrepancies between the two analyses. The two data-sets are typically in good agreement, with the PRISM data generally being a little larger than the NARR data. When we removed locations where the error between the two precipitation estimates was more than one standard deviation away from the mean then the under-prediction mode was diminished (Figure C-2 B) in the distribution of error between observations and predictions of maximum tree height.

C.5.3 Temperature Shifts

A central question in ecology is how ecosystems respond to environmental shifts. Temperature plays a dominant role in controlling evaporative rates in our framework and here we investigate how our predictions for maximum tree height respond to changes in mean annual temperature. We consider a fairly large change in mean annual temperature of ± 2 °C. The maps in Figures C-3 A and B illustrate the resulting percentage change in predicted maximum height. Averaging over the entire continent the maximum tree height decreases by 11% for a $+2$ °C shift while maximum height increases by 13% for a -2 °C shift. For an increase in temperature nearly all predicted

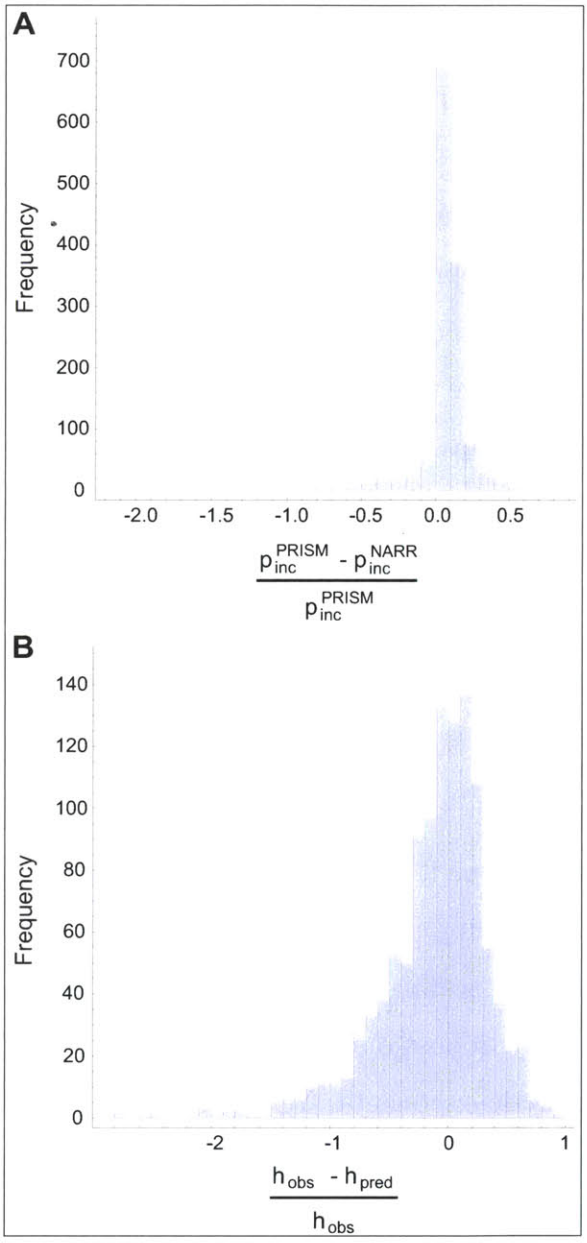


Figure C-2: The dependence of model error on precipitation estimates. **(A)** Histogram of the distribution of the discrepancies between the PRISM and NARR data for rates of precipitation. **(B)** Histogram of the distribution of the discrepancies between predicted and observed tree height. Pairs of trees and station data have been removed when the error between the PRISM and NARR databases is more than 1 standard deviation from the mean resulting in a reduction of the slight bimodality of the error distribution.

maximum heights will decrease, although the percentage change varies regionally, and for a decrease in temperature nearly all predicted heights will increase.

Looking across different environments we find that warmer environments are less sensitive to a ± 2 °C shift than colder environments, hence the strong gradient in percentage change with latitude and/or elevation. The northern and mountainous regions are more sensitive to an absolute shift in temperature. However, it should be noted that a ± 2 °C shift in a colder environment represents a much larger percentage change in mean annual temperature compared with a warmer environment. Another analysis which accounts for this effect is to change temperature by $\pm 10\%$ in every environment. Doing this we find that warmer environments are more sensitive to a constant percentage change than colder environments (Figure C-3 C and D). For the $+10\%$ change in mean annual temperature we find that the average change in maximum tree height is -8% across the entire continental United States. For the -10% change maximum heights change by an average of $+9\%$.

C.5.4 Parameter Sensitivity

It should be noted that the empirical exponents used in this study are fits to data and are accompanied by some amount of error. In addition there are several assumptions or derivations that we have made regarding the scaling of plant features for which there is no data. An important question is how robust our model is to small deviations in the values of these empirical or analytic exponents. Yet understanding the covariation of all of the exponents such that they all are in agreement with data is an interesting and complicated problem (see refs. [217, 216, 215]). For the suite of exponents involved in this paper the mechanisms of covariation are simply not known, and we do not have data for each property. Thus a full-scale, empirically-realistic sensitivity analysis to parameter combinations is not possible. Given these limitations we can however examine the effects of perturbing each exponent independently of the others. Figure C-4 shows changes in the median relative error, $\left|\frac{h_{obs}-h_{pred}}{h_{obs}}\right|$, of our predictions given a percent change in the dominant exponents of our model (including analytic assumptions or results). The dominant exponents are those which

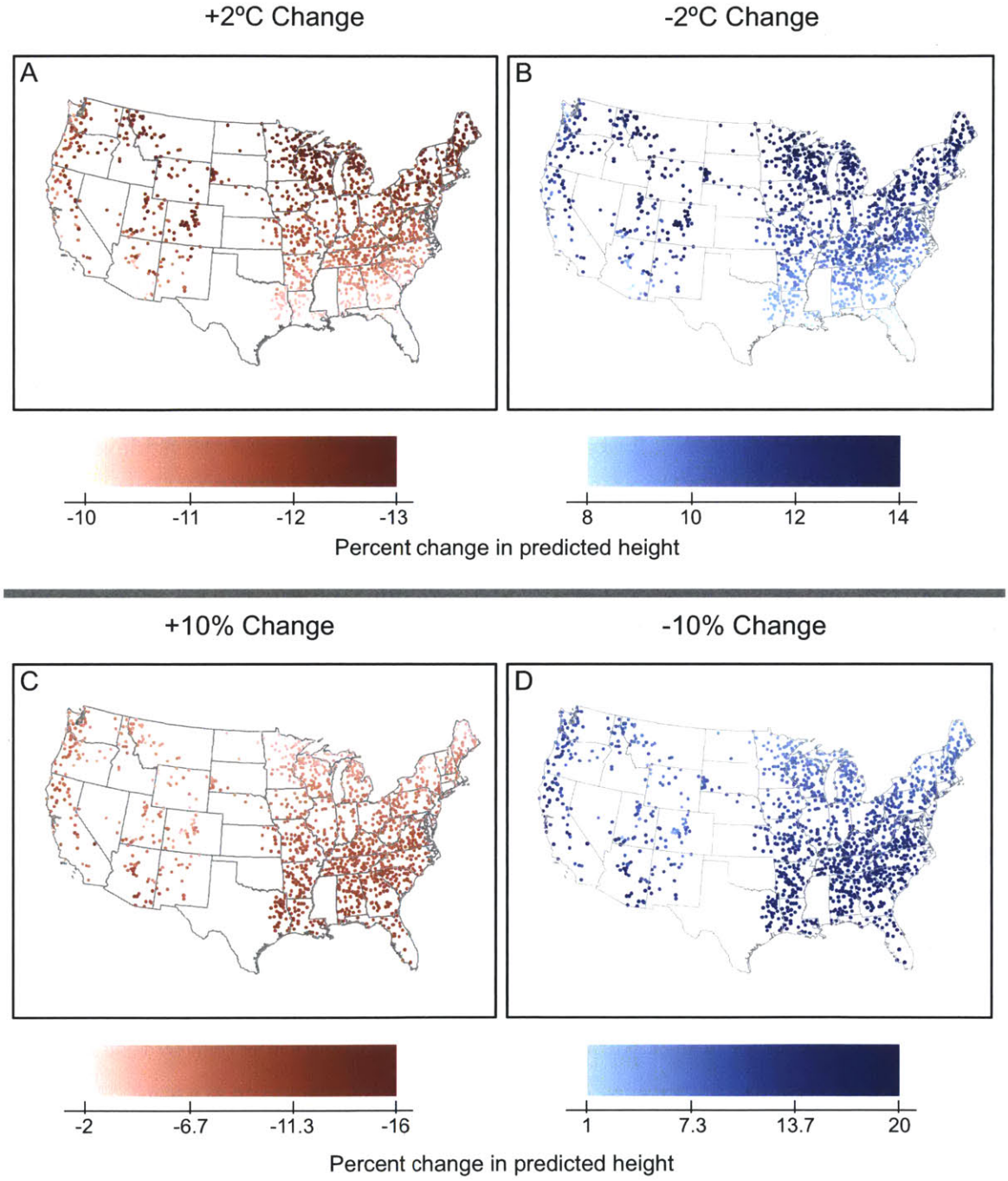


Figure C-3: The resulting percentage change in predicted maximum tree height given a **(A)** +2°C change, **(B)** −2°C change, **(C)** +10% change, and **(D)** −10% change in mean annual temperature.

control water acquisition (the root radius scaling, r_{root}), and those which regulate the canopy energy budget via its geometry and heat flux areas (canopy height, h_{can}, and radius r_{can}, and the thermal, sensible, and latent heat flux areas, a_g, a_j, a_f respectively). Our original assumption is that all heat flux areas scale according to the scaling of leaf area (see above). In Figure C-4 we test changes to this leaf area scaling as well as changes to the scaling of each heat flux area individually. In each plot zero represents the parameter value used in our model, and it can be seen that for each parameter there is a minimum in the median relative error (Figure C-4). This minimum often does not occur at the value used for our original predictions, revealing how our model could be optimized to more accurately predict data. In the future it may be possible to optimize all exponents simultaneously in order to achieve greater predictive power. However, our goal here is to test the zeroth order theory based on the available observations.

For the canopy height and radius and the scaling of the total leaf area, small changes to the exponents result in small changes to the median relative error. However, the median relative error is very sensitive to the scaling of the root radius which directly affects moisture gathering, and to differences in the scaling of the heat flux areas compared to the overall leaf area. The thermal and sensible heat flux areas contain multiple minima in the relative error and these two parameters have competing effects with one another illustrating the complicated connection between exponents and the importance of parameter covariation for a true assessment of sensitivity. Each heat flux exponent represents a physiological response to the environment, and given the complicated connection of these exponents it is of future interest to test whether our framework can be used to predict changes in exponents across environments as observed by [217].

C.5.5 Improving predictions using parameter optimization

From Figure 5-1 C it is apparent that, despite overall good agreement with the model predictions, there are significant outliers representing deviations between observations and our predictions. This is hardly surprising, given the many potential sources of

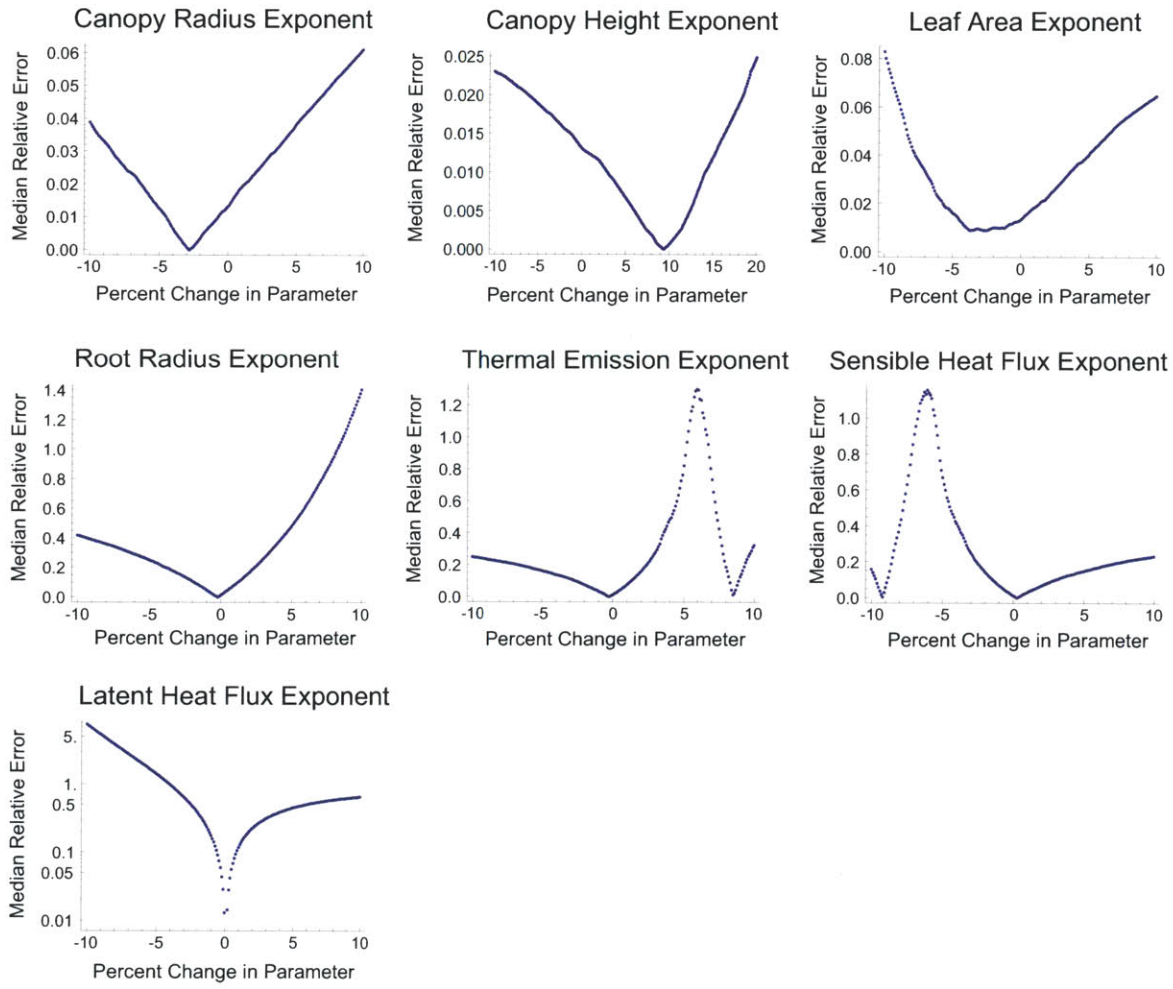


Figure C-4: Sensitivity of the model to parameter values. The change in the median relative error between observations and predictions, $\left|\frac{h_{obs}-h_{pred}}{h_{obs}}\right|$, as a result of a percentage change in the given scaling exponent. The zero percentage change represents the empirical or analytic values used for the predictions in the main text.

error such as a possible mismatch between meteorological measurements and the actual conditions experienced by the observed trees (see discussion regarding Figure C-2), and processes such as logging, mortality, and fire which are not included in the scope of our analysis. Additionally, the measured tallest trees may not have reached the environmentally-determined upper bound. As we have seen in Figure 5-5, our model anticipates the observed environmental dependence of optimal plant traits and this implies that using only a single set of plant traits across all environments may also be a source of error. Finally, some of our assumptions regarding unmeasured scaling relationships may be inaccurate and contribute to error within the predictions. Our goal was to capture the central tendencies, or average behavior, given a set of known (both observed and theoretically motivated) scaling relationships and average plant traits. However, here we show how our predictions can be improved by finding additional scaling relationships which optimize the model.

In the main text we solved for the stomatal density which maximized height at a given location and set of environmental conditions (Figure 5-5). Looking across all of the tree sites we found a relationship between stomatal density and temperature. Similarly, we can solve for the stomatal density, δ_s, which gives $Q_P(h_{obs}) = Q_E(h_{obs}, \delta_s, \{m\})$ for every observed tallest tree (given the local meteorological conditions $\{m\}$). Doing this we find that stomatal density should scale with tree height as $\delta_s \propto h^{-.71}$ (assuming a constant stomatal area). Incorporating this scaling into our model then improves our predictions as illustrated in Figure C-5 B. In our model we also use a constant value of the root absorption efficiency γ and we can also optimize the scaling of this trait alone where we find that $\gamma \propto h^{-.78}$ also reduces the error in the predictions (Figure C-5 C). This demonstrates the potential utility of the model in anticipating scaling laws for which measurements do not already exist. However, it should be noted that these two analyses optimize scaling with respect to only a single tree feature one at a time, and, for comparison with future data, it is likely necessary to co-optimize all exponents of relevance. This analysis is an example to illustrate that adding additional scalings to the model can improve predictability and reduce the variance of the deviations; however finding the realistic set of exponents

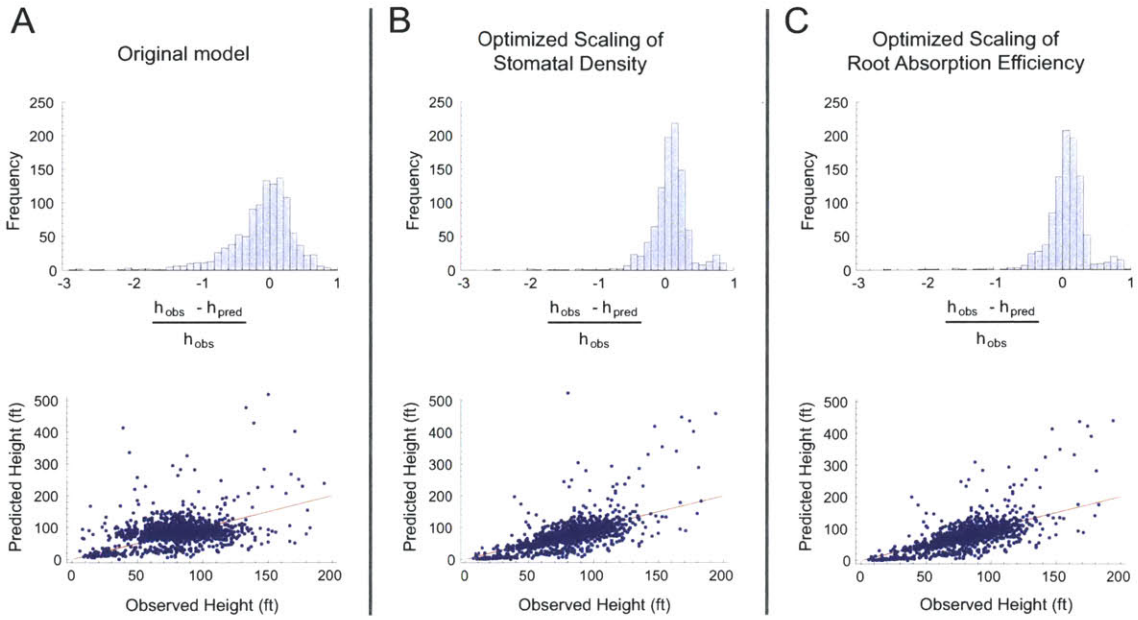


Figure C-5: Optimized scaling and model error. The change in the model predictions given an optimization in the scaling of either **(B)** stomatal density or **(C)** root absorption efficiency compared to **(A)** the original model. The red curve represents the one-to-one line. The variance of the error $\left(\frac{h_{obs}-h_{pred}}{h_{obs}}\right)$ is reduced from .22 in **(A)** to .10 in **(B)** and **(C)**. For all three analyses tree sites have been removed when the error between the PRISM and NARR precipitation estimates is more than 1 standard deviation from the mean error similar to the analysis summarized by Fig. S2. In each histogram error values less than -3.0 were omitted accounting for 19 values in **(A)** and 3 values in **(B)** and **(C)**.

subject to multiple limitations is a subject of ongoing research and requires a more advanced optimization study

Table C.1: Tree trait values, scaling parameters, and other model constants

Parameter	Symbol	Values		Ref.	TRY database		
		Theor.	Empirical		Mean	Median	n
Branching Parameter	n	2	—	[292]	—	—	—
Branching Exponent	a	1	—	[292]	—	—	—
Terminal Branch (Petiole) Radius	r_N	—	0.4 mm	[292]	—	—	—
Terminal Branch (Petiole) Length	l_N	—	4 cm	[292]	—	—	—
Proportionality Constant For Metabolism	β_1	—	0.257 liter day^{-1} cm$^{-\eta_1}$	[80]	—	—	—
Exponent for Metabolism	η_1	2	1.788	[292, 80]	—	—	—
Root to Stem Mass Proportionality	β_3	—	0.423	[189, 191]	—	—	—
Root Absorption Efficiency	γ	1/3	—	this study	—	—	—
Average Leaf Area	a_l	—	13 cm^2	[237, 88, 2, 32, 39, 103, 260, 128, 214, 205]	62.10 cm^2	11.76 cm^2	77000
Depth of a Stoma	z_s	—	10 μm	[51]	—	—	—
Leaf Stomatal Density * (averaged over both leaf sides)	δ_s	—	220 stomata mm^{-2}	[21, 130, 1, 141, 212, 107]	120 mm^{-2}	101 mm^{-2}	363
Area of a Stoma*	a_s	—	235.1 μm^2 $= \pi\,(17.9/2)^2$	[1, 107, 51]	491 μm^2	459 μm^2	2159
Stomatal Area per Leaf Area* (Stomatal Density × Area of a Stoma)	$\delta_s a_s$	—	.051	[21, 130, 1, 141, 212, 107, 51]	.059	.047	—
Leaf Absorptivity (full spectrum)	α	—	0.50	[51, 173]	—	—	—
Soil Reflection Coefficient (full spectrum)	ξ_s	—	0.30	[51, 173]	—	—	—
Deep Canopy Reflection Coefficient (full spectrum)	ξ_c^*	—	0.22	[51, 173]	—	—	—
Proportionality Constant for Canopy Radius	β_5	—	35.24 cm m$^{-\eta_3}$	SI of [83]	—	—	—
Exponent for Canopy Radius	η_3	1	1.14	SI of [83]	—	—	—
Leaf Emmisivity	ϵ	—	0.95	[51, 173]	—	—	—

*It should be noted that our model relies on the bulk property of "Stomatal Area per Leaf Area" for calculations requiring an evaporative flux area a_f. Thus the sub-parameters a_s and δ_s need not agree with the TRY database as long as the bulk property $\delta_s a_s$ does.

Bibliography

[1] MD Abrams and ME Kubiske. Leaf structural characteristics of 31 hardwood and conifer tree species in central Wisconsin: influence of light regime and shade-tolerance rank. *Forest Ecology and Management*, 31(4):245–253, 1990.

[2] DD Ackerly and PB Reich. Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. *American Journal of Botany*, 86(9):1272–1281, 1999.

[3] G Alagappan and RM Cowan. Effect of temperature and dissolved oxygen on the growth kinetics of Pseudomonas putida F1 growing on benzene and toluene. *Chemosphere*, 54(8):1255–1265, 2004.

[4] AE Albani, S. Bengtson, DE Canfield, A. Bekker, R. Macchiarelli, A. Mazurier, EU Hammarlund, P. Boulvais, JJ Dupuy, C. Fontaine, et al. Large colonial organisms with coordinated growth in oxygenated environments 2.1 Gyr ago. *Nature*, 466(7302):100, 2010.

[5] R Albert and HG Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *Journal of Theoretical Biology*, 223(1):1–18, 2003.

[6] F Alcantara and M Monk. Signal propagation during aggregation in the slime mould Dictyostelium discoideum. *Microbiology*, 85(2):321, 1974.

[7] B Allen. *Studies in the mathematics of evolution and biodiversity.* PhD thesis, Boston University, 2010.

[8] W Alt. Biased random walk models for chemotaxis and related diffusion approximations. *Journal of mathematical biology*, 9(2):147–177, 1980.

[9] J André and B Godelle. The evolution of mutation rate in finite asexual populations. *Genetics*, 172(1):611–626, 2006.

[10] NCEP/NCAR North American American Regional Reanalysis Archive. http://dss.ucar.edu/pub/narr/. Online, March 2009. Database accessed on 20 March 2009.

[11] GJE Baart, M Willemsen, E Khatami, A De Haan, B Zomer, E Coen Beuvery, J Tramper, and DE Martens. Modeling neisseria meningitidis b metabolism at different specific growth rates. *Biotechnology and bioengineering*, 101(5):1022–1035, Dec 2008.

[12] ME Baird, PR Oke, IM Suthers, and JH Middleton. A plankton population model with biomechanical descriptions of biological processes in an idealised 2D ocean basin. *Journal of Marine Systems*, 50(3-4):199–222, 2004.

[13] R Bakke, MG Trulear, JA Robinson, and WG Characklis. Activity of pseudomonas aeruginosa in biofilms: steady state. *Biotechnology and bioengineering*, 26(12):1418–1424, 2004.

[14] JR Banavar, J Damuth, A Maritan, and A Rinaldo. Supply–demand balance and metabolic scaling. *Proceedings of the National Academy of Sciences*, 99(16):10506–10509, 2002.

[15] J.R. Banavar, J. Damuth, A. Maritan, and A. Rinaldo. Supply–demand balance and metabolic scaling. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10506, 2002.

[16] JR Banavar, ME Moses, JH Brown, J Damuth, A Rinaldo, RM Sibly, and A Maritan. A general basis for quarter-power scaling in animals. *Proceedings of the National Academy of Sciences*, 107(36):15816–15820, 2010.

[17] AL Barabási and R Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[18] JE Barrick and RE Lenski. Genome-wide mutational diversity in an evolving population of Escherichia coli. In *Cold Spring Harbor symposia on quantitative biology*, volume 74, page 119. NIH Public Access, 2009.

[19] JE Barrick, DS Yu, SH Yoon, H Jeong, TK Oh, D Schneider, RE Lenski, and JF Kim. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, 461(7268):1243–1247, 2009.

[20] NJ Basson. Competition for glucose between candida albicans and oral bacteria grown in mixed culture in a chemostat. *Journal of medical microbiology*, 49(11):969–975, 2000.

[21] DJ Beerling and WG Chaloner. The Impact of Atmospheric CO2 and Temperature Changes on Stomatal Density: Observation from Quercus robur Lammas Leaves. *Annals of Botany*, 71(3):231–235, 1993.

[22] J Beringer, FS Chapin III, CC Thompson, and AD McGuire. Surface energy exchanges along a tundra-forest transition and feedbacks to climate. *Agricultural and Forest Meteorology*, 131(3-4):143–161, 2005.

[23] JA Bertout, SA Patel, and MC Simon. The impact of o2 availability on human cancer. *Nature Reviews Cancer*, 8(12):967–975, 2008.

[24] L Bettencourt, J Lobo, D Helbing, C Kühnert, and GB West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301, 2007.

[25] H Beyenal, SN Chen, and Z Lewandowski. The double substrate growth kinetics of pseudomonas aeruginosa. *Enzyme and microbial technology*, 32(1):92–98, 2003.

[26] I Bjedov, O Tenaillon, B Gérard, V Souza, E Denamur, M Radman, F Taddei, and I Matic. Stress-induced mutagenesis in bacteria. *Science*, 300(5624):1404–1409, 2003.

[27] L. M Blank and U Sauer. Tca cycle activity in saccharomyces cerevisiae is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology*, 150(4):1085–1093, Apr 2004.

[28] B. Blonder, C. Violle, L.P. Bentley, and B.J. Enquist. Venation networks and the origin of the leaf economics spectrum. *Ecology Letters*, 14(2):91–100, 2011.

[29] L. Boe, M. Danielsen, S. Knudsen, JB Petersen, J. Maymann, and PR Jensen. The frequency of mutators in populations of Escherichia coli. *Mutation research*, 448(1):47–55, 2000.

[30] G.B. Bonan, S. Levis, L. Kergoat, and K.W. Oleson. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, 16(2):1021, 2002.

[31] G.B. Bonan, S. Levis, S. Sitch, M. Vertenstein, and K.W. Oleson. A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics. *Global Change Biology*, 9(11):1543–1566, 2003.

[32] WJ Bond and J Midgley. Allometry and sexual differences in leaf size. *American Naturalist*, 131(6):901–910, 1988.

[33] M.E. Boraas. Population dynamics of food-limited rotifers in two-stage chemostat culture. *Limnology and Oceanography*, 28(3):546–563, 1983.

[34] D.B. Botkin, J.F. Janak, and J.R. Wallis. Some ecological consequences of a computer model of forest growth. *The Journal of Ecology*, 60:849–872, 1972.

[35] D.P Boudreaux and V.R Srinivasan. A continuous culture study of growth of bacillus cereus t. *Journal of General Microbiology*, 122(1):129–136, 1981.

[36] B. A Bradley. A nima-related kinase, cnk2p, regulates both flagellar length and cell size in chlamydomonas. *Journal of Cell Science*, 118(15):3317–3326, Aug 2005.

[37] G. Bratbak and I. Dundas. Bacterial dry matter content and biomass estimations. *Applied and Environmental Microbiology*, 48(4):755, 1984.

[38] D. Bray. *Cell movements: from molecules to motility.* Routledge, 2001.

[39] C Brouat, M Gibernau, L Amsellem, and D McKey. Corner's rules revisited: ontogenetic and interspecific patterns in leaf-stem allometry. *New Phytologist*, 139(3):459–470, 1998.

[40] CM Brown and AH Rose. Effects of temperature on composition and cell volume of candida utilis. *Journal of bacteriology*, 97(1):261–272, 1969.

[41] DE Brown and DJ Halsted. The effect of acid ph on the growth kinetics of trichoderma viride. *Biotechnology and bioengineering*, 17(8):1199–1210, 1975.

[42] E.W. Brown, J.E. LeClerc, B. Li, W.L. Payne, and T.A. Cebula. Phylogenetic evidence for horizontal transfer of mutS alleles among naturally occurring Escherichia coli strains. *Journal of Bacteriology*, 183(5):1631, 2001.

[43] J.H. Brown, J.F. Gillooly, A.P. Allen, V.M. Savage, and G.B. West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.

[44] J.H. Brown, J.F. Gillooly, A.P. Allen, V.M. Savage, and G.B. West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.

[45] J.H. Brown and G.B. West. *Scaling in biology.* Oxford University Press, USA, 2000.

[46] E.O. Budrene and H.C. Berg. Complex patterns formed by motile cells of Escherichia coli. *Nature*, 349(6310):630–633, 1991.

[47] H. Bugmann. A review of forest gap models. *Climatic Change*, 51(3):259–305, 2001.

[48] A.T. Bull. The renaissance of continuous culture in the post-genomics age. *Journal of industrial microbiology & biotechnology*, 37(10):993–1021, 2010.

[49] B.A Bulthuis, G.M Koningstein, A.H Stouthamer, and H.W van Verseveld. A comparison between aerobic growth of bacillus licheniformis in continuous culture and partial-recycling fermentor, with contributions to the discussion on maintenance energy demand. *Arch Microbiol*, 152(5):499–507, 1989.

[50] DK Button and JC Garver. Continuous culture of torulopsis utilis: a kinetic study of oxygen limited growth. *Journal of General Microbiology*, 45(2):195–204, 1966.

[51] G S Campbell and J M Norman. *Introduction to environmental biophysics, second edition.* Springer Verlag, 1998.

[52] L.H. Caporale. *The implicit genome.* Oxford University Press, USA, 2006.

[53] S. Cayley, BA Lewis, HJ Guttman, and MT Record Jr. Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *Journal of molecular biology*, 222(2):281, 1991.

[54] L. Ceccarelli. *Shaping science with rhetoric: The cases of Dobzhansky, Schrödinger, and Wilson.* University of Chicago Press, 2001.

[55] L. Chao and E.C. Cox. Competition between high and low mutating strains of *Escherichia coli*. *Evolution*, 37(1):125–134, 1983.

[56] J. Charney, P.H. Stone, and W.J. Quirk. Drought in the Sahara-A biogeophysical feedback mechanism. *Science*, 187:434–435, 1975.

[57] F Chen and M.R Johns. Heterotrophic growth of chlamydomonas reinhardtii on acetate in chemostat culture. *Process Biochemistry*, 31(6):601–604, 1996.

[58] F. Chen, W. Liu, A. Eisenstark, R. Johnston, G. Liu, and S. Liu. Multiple genetic switches spontaneously modulating bacterial mutability. *BMC Evolutionary Biology*, 10(1):277, 2010.

[59] F. Chen, W. Liu, Z. Liu, Q. Zou, Y. Wang, Y. Li, J. Zhou, A. Eisenstark, R.N. Johnston, G. Liu, B. Yang, and S. Liu. mutl as a genetic switch of bacterial mutability: turned on or off through repeat copy number changes. *FEMS Microbiology Letters*, 312(2):126–132, 2010.

[60] F Chen, Q Xia, and Lu-Kwang Ju. Competition between oxygen and nitrate respirations in continuous culture of pseudomonas aeruginosa performing aerobic denitrification. *Biotechnology and bioengineering*, 93:1069–1078, Jan 2006.

[61] K.D. Coates, C.D. Canham, M. Beaudet, D.L. Sachs, and C. Messier. Use of a spatially explicit individual-tree model (SORTIE/BC) to explore the implications of patchiness in structurally complex forests. *Forest Ecology and Management*, 186(1-3):297–310, 2003.

[62] M Cocaign-Bousquet, A Guyonvarch, and N.D Lindley. Growth rate-dependent modulation of carbon flux through central metabolism and the kinetic consequences for glucose-limited chemostat cultures of corynebacterium glutamicum. *Applied and environmental microbiology*, 62(2):429–436, 1996.

[63] E.A. Codling, M.J. Plank, and S. Benhamou. Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834, 2008.

[64] National Renewable Energy Laboratory Continental United States Low Resolution Solar Grid. http://www.nrel.gov/gis/data_analysis.html. Online, June 2005. Database accessed on 10 June 2005.

[65] T.F. Cooper. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biology*, 5(9):e225, 08 2007.

[66] C Daly, R P Neilson, and D L Phillips. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of applied meteorology*, 33(2):140–158, 1994.

[67] TRY database. www.try-db.org. Online, July 2010. Database accessed on 26 July 2010.

[68] D. De Beer, P. Stoodley, F. Roe, and Z. Lewandowski. Effects of biofilm structures on oxygen distribution and mass transport. *Biotechnology and bioengineering*, 43(11):1131–1138, 1994.

[69] J.P. DeLong, J.G. Okie, M.E. Moses, R.M. Sibly, and J.H. Brown. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences*, 107(29):12941, 2010.

[70] E. Denamur, G. Lecointre, P. Darlu, O. Tenaillon, C. Acquaviva, C. Sayada, I. Sunjevaric, R. Rothstein, J. Elion, F. Taddei, et al. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell*, 103(5):711–721, 2000.

[71] E. Denamur and I. Matic. Evolution of mutation rates in bacteria. *Molecular Microbiology*, 60(4):820–827, 2006.

[72] M.M. Desai and D.S. Fisher. The balance between mutators and nonmutators in asexual populations. *Genetics*, 188(4):997–1014, 2011.

[73] L.E.P. Dietrich, T.K. Teal, A. Price-Whelan, and D.K. Newman. Redox-active antibiotics control gene expression and community behavior in divergent bacteria. *Science Signalling*, 321(5893):1203, 2008.

[74] T.G. Dobzhansky and T. Dobzhansky. *Genetics and the Origin of Species*. Columbia Univ Pr, 1937.

[75] JW Drake, B Charlesworth, D Charlesworth, and JF Crow. Rates of spontaneous mutation. *Genetics*, 148:1667–1686, 1998.

[76] D.J. Earl and M.W. Deem. Evolvability is a selectable trait. *Proceedings of the National Academy of Sciences of the United States of America*, 101(32):11531–11536, 2004.

[77] J.S. Edwards, R.U. Ibarra, and B.O. Palsson. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature biotechnology*, 19(2):125–130, 2001.

[78] M. Eigen, J. McCaskill, and P. Schuster. The molecular quasispecies. *Adv. Chem. Phys*, 75:149–263, 1989.

[79] M. Eigen and P. Schuster. A principle of natural self-organization. *Naturwissenschaften*, 64(11):541–565, 1977.

[80] B J Enquist, J H Brown, and G B West. Allometric scaling of plant energetics and population density. *Nature*, 395(6698):163–165, 1998.

[81] B J Enquist and K J Niklas. Invariant scaling relations across tree-dominated communities. *Nature*, 410(6829):655–660, 2001.

[82] B J Enquist, G B West, and J H Brown. A general quantitative theory of forest structure and dynamics. *Proceedings of the National Academy of Sciences USA*, 106(17):7040–7045, 2009.

[83] B J Enquist, G B West, and J H Brown. Extensions and evaluations of a general quantitative theory of forest structure and dynamics. *Proceedings of the National Academy of Sciences USA*, 106(17):7046–7051, 2009.

[84] R. Erban and H.G. Othmer. From signal transduction to spatial pattern formation in E. coli: a paradigm for multiscale modeling in biology. *SIAM J. Appl. Math*, 65:361–391, 2004.

[85] P.G. Falkowski, Z. Dubinsky, and K. Wyman. Growth-irradiance relationships in phytoplankton. *Limnology and Oceanography*, pages 311–321, 1985.

[86] P.G. Falkowski, T. Fenchel, and E.F. Delong. The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320(5879):1034, 2008.

[87] P.G Falkowski and T.G Owens. Light–shade adaptation: two strategies in marine phytoplankton. *Plant physiology*, 66(4):592–595, 1980.

[88] D S Falster and M Westoby. Leaf size and angle vary widely across species: what consequences for light interception? *New Phytologist*, 158(3):509–525, 2003.

[89] D.S. Falster and M. Westoby. Plant height and evolutionary games. *Trends in Ecology & Evolution*, 18(7):337–343, 2003.

[90] D.S. Falster and M. Westoby. Alternative height strategies among 45 dicot rain forest species from tropical Queensland, Australia. *Journal of Ecology*, 93(3):521–535, 2005.

[91] D.S. Falster and M. Westoby. Tradeoffs between height growth rate, stem persistence and maximum height among plant species in a post-fire succession. *Oikos*, 111(1):57–66, 2005.

[92] I.S. Farmer and C.W. Jones. The energetics of Escherichia coli during aerobic growth in continuous culture. *European Journal of Biochemistry*, 67(1):115–122, 1976.

[93] J. Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78(2):737–756, 1974.

[94] T Fenchel. Intrinsic rate of natural increase: the relationship with body size. *Oecologia*, 14(4):317–326, 1974.

[95] T Fenchel. Ecology of heterotrophic microflagellates. 2. bioenergetics and growth. *Mar. Ecol. Prog. Ser*, 8:225–231, 1982.

[96] Z Finkel, A Irwin, and O Schofield. Resource limitation alters the 3/4 size scaling of metabolic rates in phytoplankton. *Marine Ecology Progress Series*, Jan 2004.

[97] J.A. Foley, I.C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, and A. Haxeltine. An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles*, 10(4):603–628, 1996.

[98] P.L. Foster. Adaptive mutation in Escherichia coli. *Journal of bacteriology*, 186(15):4846, 2004.

[99] J Frankena, G.M Koningstein, H.W van Verseveld, and A.H Stouthamer. Effect of different limitations in chemostat cultures on growth and production of exocellular protease by bacillus licheniformis. *Applied microbiology and biotechnology*, 24(2):106–112, 1986.

[100] J Frankena, H.W Verseveld, and A.H Stouthamer. A continuous culture study of the bioenergetic aspects of growth and production of exocellular protease in bacillus licheniformis. *Applied microbiology and biotechnology*, 22(3):169–176, 1985.

[101] J Frankena, HW Van Verseveld, and AH Stouthamer. Substrate and energy costs of the production of exocellular enzymes by bacillus licheniformis. *Biotechnology and bioengineering*, 32(6):803–812, 1988.

[102] R. S. Galhardo, P. J. Hastings, and S. M. Rosenberg. Mutation as a stress response and the regulation of evolvability. *Critical Reviews in Biochemistry and Molecular Biology*, 42(5):399–435, 2007.

[103] D M Gates and R B Schmerl. *Perspectives of biophysical ecology*. Springer, 1975.

[104] P.J. Gerrish, A. Colato, A.S. Perelson, and P.D. Sniegowski. Complete genetic linkage can subvert natural selection. *Proceedings of the National Academy of Sciences*, 104(15):6266–6271, 2007.

[105] J.F. Gillooly, J.H. Brown, G.B. West, V.M. Savage, and E.L. Charnov. Effects of size and temperature on metabolic rate. *Science*, 293(5538):2248, 2001.

[106] J.F. Gillooly, E.L. Charnov, G.B. West, V.M. Savage, and J.H. Brown. Effects of size and temperature on developmental time. *Nature*, 417(6884):70–73, 2002.

[107] I Gindel. Stomatal number and size as related to soil moisture in tree xerophytes in Israel. *Ecology*, 50(2):263–267, 1969.

[108] A. Giraud, I. Matic, O. Tenaillon, A. Clara, M. Radman, M. Fons, and F. Taddei. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science*, 291(5513):2606, 2001.

[109] A. Giraud, M. Radman, I. Matic, and F. Taddei. The rise and fall of mutator bacteria. *Current Opinion in Microbiology*, 4(5):582–585, 2001.

[110] M. Godin, F.F. Delgado, S. Son, W.H. Grover, A.K. Bryan, A. Tzur, P. Jorgensen, K. Payer, A.D. Grossman, M.W. Kirschner, et al. Using buoyant mass to measure the growth of single cells. *Nature Methods*, 7(5):387–390, 2010.

[111] J.C Goldman and E.J Carpenter. A kinetic approach to the effect of temperature on algal growth. *Limnology and Oceanography*, 19(5):756–766, 1974.

[112] J. Gong, W. Liu, G Liu, F. Chen, J. Li, G Xu, L. Wang, R.N. Johnston, A. Eisenstark, and S. Liu. Spontaneous conversion between mutl and 6 bp[delta]mutl in salmonella typhimurium lt7: Association with genome diversification and possible roles in bacterial adaptation. *Genomics*, 90(4):542–549, 2007.

[113] H.J Gons and L.R Mur. Energy requirements for growth and maintenance of scenedesmus protuberans fritsch in light-limited continuous cultures. *Archives of microbiology*, 125(1):9–17, 1980.

[114] BR Grant and I.M Turner. Light-stimulated nitrate and nitrite assimilation in several species of algae. *Comparative Biochemistry and Physiology*, 29(3):995–1004, 1969.

[115] JBS Haldane. A mathematical theory of natural and artificial selection. V. Selection and mutation. *Proc. Camb. Phil. Soc.*, 23:838–844, 1927.

[116] A. Haxeltine and I.C. Prentice. BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochemical Cycles*, 10(4):693–709, 1996.

[117] A. Haxeltine, I.C. Prentice, and I.D. Creswell. A coupled carbon and water flux model to predict vegetation structure. *Journal of Vegetation Science*, 7(5):651–666, 1996.

[118] JJ Heijnen and JA Roels. A macroscopic model describing yield and maintenance relationships in aerobic fermentation processes. *Biotechnology and Bioengineering*, 23(4):739–763, 1981.

[119] M.A. Herman and D.R. Soll. A comparison of volume growth during bud and mycelium formation in Candida albicans: a single cell analysis. *Microbiology*, 130(9):2219, 1984.

[120] D Hilbert and S Cohn-Vossen. *Geometry and the Imagination*. American Mathematical Society, 1999.

[121] C. Hou, W. Zuo, M.E. Moses, W.H. Woodruff, J.H. Brown, and G.B. West. Energy uptake and allocation during ontogeny. *Science*, 322(5902):736, 2008.

[122] Forest Inventory and United States Department of Agriculture Forest Service Analysis. http://fiatools.fs.fed.us/fiadb-downloads/datamart.html. Online, December 2007. Database accessed on 20 December 2007.

[123] Eiji Ishikawa and Hiroshi Abe. Lycopene accumulation and cyclic carotenoid deficiency in heterotrophic chlorella treated with nicotine. *Journal of Industrial Microbiology and Biotechnology*, 31(12):585–589, Dec 2004.

[124] S. Janson, B. Bergman, E.J. Carpenter, S.J. Giovannoni, and K. Vergin. Genetic analysis of natural populations of the marine diazotrophic cyanobacterium trichodesmium. *FEMS microbiology ecology*, 30(1):57–65, 1999.

[125] T. Johnson. Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *Genetics*, 151(4):1621, 1999.

[126] GC Johnston, CW Ehrhardt, A Lorincz, and BL Carter. Regulation of cell size in the yeast saccharomyces cerevisiae. *Journal of bacteriology*, 137(1):1–5, 1979.

[127] E Kalnay, M Kanamitsu, R Kistler, W Collins, D Deaven, L Gandin, M Iredell, S Saha, G White, J Woollen, et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–472, 1996.

[128] J F Karlik and A H McKay. Leaf area index, leaf mass density, and allometric relationships derived from harvest of blue oaks in California oak savanna. In *Standiford RB, McCreary DD, Purcell KL (tech. coords.). Proc Fifth Symp Oak Woodlands: Oaks in Californias Changing Landscape, Oct*, pages 22–5, 2001.

[129] E.F. Keller. *Making sense of life: Explaining biological development with models, metaphors, and machines*. Harvard Univ Pr, 2003.

[130] CK Kelly and DJ Beerling. Plant life form, stomatal density and taxonomic relatedness: a reanalysis of Salisbury (1927). *Functional Ecology*, 9(3):422–431, 1995.

[131] C.P. Kempes, S. Dutkiewicz, and M.J. Follows. Growth, metabolic partitioning, and the size of microorganisms. *Proceedings of the National Academy of Sciences*, 109(2):495–500, 2012.

[132] C.P. Kempes, G.B. West, K. Crowell, and M. Girvan. Predicting maximum tree heights and other traits from allometric scaling and resource limitations. *PLoS One*, 6(6):e20551, 2011.

[133] T.T. Kibota and M. Lynch. Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. *Nature*, 381(6584):694–696, 1996.

[134] M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713–719, 1962.

[135] D.A. King. The adaptive significance of tree height. *American Naturalist*, 135:809–828, 1990.

[136] M. Kirschner and J. Gerhart. Evolvability. *Proceedings of the National Academy of Sciences of the USA*, 95(15):8420–08427, 1998.

[137] AMJ Kliphuis, AJ Klok, DE Martens, PP Lamers, M Janssen, and RH Wijffels. Metabolic modeling of chlamydomonas reinhardtii: energy requirements for photoautotrophic growth and maintenance. *Journal of applied phycology*, pages 1–14, Apr 2011.

[138] R. Knutti, MR Allen, P. Friedlingstein, JM Gregory, GC Hegerl, GA Meehl, M. Meinshausen, JM Murphy, GK Plattner, SCB Raper, et al. A review of uncertainties in global temperature projections over the twenty-first century. *Journal of Climate*, 21(11):2651–2663, 2008.

[139] G.W. Koch, S.C. Sillett, G.M. Jennings, and S.D. Davis. The limits to tree height. *Nature*, 428(6985):851–854, 2004.

[140] S. Kooijman. *Dynamic energy and mass budgets in biological systems*. Cambridge Univ Pr, 2000.

[141] C Körner, P Bannister, and A F Mark. Altitudinal variation in stomatal conductance, nitrogen content and leaf anatomy in different plant life forms in New Zealand. *Oecologia*, 69(4):577–588, 1986.

[142] CQ Lan, G Oddone, DA Mills, and DE Block. Kinetics of lactococcus lactis growth and metabolite formation under aerobic and anaerobic conditions in the presence or absence of hemin. *Biotechnology and bioengineering*, 95(6):1070–1080, Dec 2006.

[143] N. Lane and W. Martin. The energetics of genome complexity. *Nature*, 467(7318):929–934, 2010.

[144] J.E. LeClerc, B. Li, W.L. Payne, and T.A. Cebula. High mutation frequencies among Escherichia coli and Salmonella pathogens. *Science*, 274(5290):1208, 1996.

[145] Y.K.U.N. Lee and S. Pirt. Energetics of photosynthetic algal growth: influence of intermittent illumination in short (40 s) cycles. *Microbiology*, 124(1):43, 1981.

[146] EG Leigh. Natural selection and mutability. *American Naturalist*, 104(937):301–305, 1970.

[147] EG Leigh. The evolution of mutation rates. *Genetics*, 73:1–18, 1973.

[148] R.E. Lenski, M.R. Rose, S.C. Simpson, and S.C. Tadler. Long-term experimental evolution in escherichia coli. i. adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6):1315–1341, 1991.

[149] R.E. Lenski and M. Travisano. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences of the United States of America*, 91(15):6808, 1994.

[150] DW Levine and CL Cooney. Isolation and characterization of a thermotolerant methanol-utilizing yeast. *Applied and Environmental Microbiology*, 26(6):982–990, 1973.

[151] H. Levine, I. Aranson, L. Tsimring, and T.V. Truong. Positive genetic feedback governs cAMP spiral wave formation in Dictyostelium. *Proceedings of the National Academy of Sciences of the United States of America*, 93(13):6382, 1996.

[152] G. Levinson and GA Gutman. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4(3):203, 1987.

[153] Z Lewandowski, G Walser, and W.G. Characklis. Reaction kinetics in biofilms. *Biotechnology and Bioengineering*, 38:877–882, 2012.

[154] C. Li, J.K. Ichikawa, J.J. Ravetto, H.C. Kuo, J.C. Fu, and S. Clarke. A new gene involved in stationary-phase survival located at 59 minutes on the escherichia coli chromosome. *Journal of bacteriology*, 176(19):6015–6022, 1994.

[155] E. Litchman, CA Klausmeier, and K. Yoshiyama. Contrasting size evolution in marine and freshwater diatoms. *Proceedings of the National Academy of Sciences*, 106(8):2665, 2009.

[156] GR Liu, K Edwards, A Eisenstark, YM Fu, WQ Liu, KE Sanderson, RN Johnston, and SL Liu. Genomic diversification among archival strains of salmonella enterica serovar typhimurium lt7. *Journal of Bacteriology*, 185(7):2131–2142, 2003.

[157] M. Loferer-Krossbacher, J. Klima, and R. Psenner. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and Environmental Microbiology*, 64(2):688, 1998.

[158] M Lopes, N Gomes, C Gonçalves, M.A.Z Coelho, M Mota, and I Belo. Yarrowia lipolytica lipase production enhanced by increased air pressure. *Letters in Applied Microbiology*, 46(2):255–260, Dec 2008.

[159] T. Løvdal, E.F. Skjoldal, M. Heldal, S. Norland, and T.F. Thingstad. Changes in morphology and elemental composition of Vibrio splendidus along a gradient from carbon-limited to phosphate-limited growth. *Microbial Ecology*, 55(1):152–161, 2008.

[160] P. Lu, C. Vogel, R. Wang, X. Yao, and E.M. Marcotte. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature biotechnology*, 25(1):117–124, 2006.

[161] R.M. Macnab. Genetics and biogenesis of bacterial flagella. *Annual review of genetics*, 26(1):131–158, 1992.

[162] EF Mao, L. Lane, J. Lee, and JH Miller. Proliferation of mutators in a cell population. *Journal of Bacteriology*, 179(2):417, 1997.

[163] ME Martínez, JM Jiménez, and F El Yousfi. Influence of phosphorus concentration and temperature on growth and phosphorus uptake by the microalga scenedesmus obliquus. *Bioresource technology*, 67(3):233–240, 1999.

[164] I. Matic, M. Radman, F. Taddei, B. Picard, C. Doit, E. Bingen, E. Denamur, et al. Highly variable mutation rates in commensal and pathogenic Escherichia coli. *Science*, 277(5333):1833, 1997.

[165] T.A. McMahon and R.E. Kronauer. Tree structures: Deducing the principle of mechanical design. *Journal of theoretical biology*, 59(2):443–466, 1976.

[166] D. Medvigy, SC Wofsy, JW Munger, DY Hollinger, and PR Moorcroft. Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *Journal of Geophysical Research*, 114(G1):G01002, 2009.

[167] D. Medvigy, S.C. Wofsy, J.W. Munger, and P.R. Moorcroft. Responses of terrestrial ecosystems and carbon budgets to current and future environmental variability. *Proceedings of the National Academy of Sciences USA*, 107(18):8275–8280, 2010.

[168] J Meghrous, E Huot, M Quittelier, and H Petitdemange. Regulation of nisin biosynthesis by continuous cultures and by resting cell of lactococcus lactis subsp. lactis. *Research in microbiology*, 143(9):879–890, 1992.

[169] F Mesinger, G DiMego, E Kalnay, K Mitchell, P C Shafran, W Ebisuzaki, D Jović, J Woollen, E Rogers, E H Berbery, et al. North American regional reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360, 2006.

[170] K.J. Mitchell. Dynamics and Simulated Yield of Douglas-fir. *Forest Science*, 21(Supplement 17):1–39, 1975.

[171] J.M. Mitchison. *The biology of the cell cycle*. Cambridge University Press, 1971.

[172] J. Monod. The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394, 1949.

[173] J L Monteith and M H Unsworth. *Principles of environmental physics, second edition*. Academic Press, 2008.

[174] PR Moorcroft. Recent advances in ecosystem-atmosphere interactions: an ecological perspective. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1521):1215–1227, 2003.

[175] P.R. Moorcroft. How close are we to a predictive science of the biosphere? *Trends in Ecology & Evolution*, 21(7):400–407, 2006.

[176] PR Moorcroft, GC Hurtt, and SW Pacala. A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs*, 71(4):557–586, 2001.

[177] S. Mori, K. Yamaji, A. Ishida, S.G. Prokushkin, O.V. Masyagina, A. Hagihara, ATM Hoque, R. Suwa, A. Osawa, T. Nishizono, et al. Mixed-power scaling of whole-plant respiration from seedlings to giant trees. *Proceedings of the National Academy of Sciences USA*, 107(4):1447–1451, 2010.

[178] M Morita, Y Watanabe, and H Saiki. High photosynthetic productivity of green microalga chlorella sorokiniana. *Applied biochemistry and biotechnology*, 87(3):203–218, 2000.

[179] M.E. Moses, C. Hou, W.H. Woodruff, G.B. West, J.C. Nekola, W. Zuo, and J.H. Brown. Revisiting a model of ontogenetic growth: estimating model parameters from theory and data. *The American Naturalist*, 171(5):632–645, 2008.

[180] A. Moya, R. Gil, A. Latorre, J. Peretó, G.B. Pilar, et al. Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiology Reviews*, 33(1):225–235, 2009.

[181] H.C. Muller-Landau, R.S. Condit, K.E. Harms, C.O. Marks, S.C. Thomas, S. Bunyavejchewin, G. Chuyong, L. Co, S. Davies, R. Foster, et al. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecology Letters*, 9(5):589–602, 2006.

[182] J.D. Murray. *Mathematical Biology: an introduction*. Springer Verlag, 2002.

[183] NOAA Satellite and Information Service NCDC. http://www.ncdc.noaa.gov/oa/climate/stationlocator.html. Online, June 2005. Database accessed on 21 June 2005.

[184] NOAA Earth System Research Laboratory NCEP/NCAR Reanalysis Project. http://www.cdc.noaa.gov/data/reanalysis/reanalysis.html. Online, March 2009. Database accessed on 20 March 2009.

[185] F.C. Neidhardt. *Escherichia coli and Salmonella: Cellular and Molecular Biology. Vol. 1.* ASM Press, 1996.

[186] M. Newman. *Networks: an introduction.* Oxford University Press, Inc., 2010.

[187] K J Niklas. *Plant allometry: the scaling of form and process.* University of Chicago Press, 1994.

[188] K J Niklas. Predicting the height of fossil plant remains: an allometric approach to an old problem. *American Journal of Botany*, 81(10):1235–1242, 1994.

[189] K J Niklas. Plant allometry: is there a grand unifying theory? *Biological reviews*, 79(04):871–889, 2004.

[190] K J Niklas and B J Enquist. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proceedings of the National Academy of Sciences USA*, 98(5):2922–2927, 2001.

[191] K J Niklas and H C Spatz. Growth and hydraulic (not mechanical) constraints govern the scaling of tree height and mass. *Proceedings of the National Academy of Sciences USA*, 101(44):15661–15663, 2004.

[192] K J Niklas and H C Spatz. Allometric theory and the mechanical stability of large trees: proof and conjecture. *American journal of botany*, 93(6):824–828, 2006.

[193] K.J. Niklas. *Plant allometry: the scaling of form and process.* University of Chicago Press, 1994.

[194] K.J. Niklas. Maximum plant height and the biophysical factors that limit it. *Tree physiology*, 27(3):433–440, 2007.

[195] S. Norland, M. Heldal, and O. Tumyr. On the relation between dry matter and volume of bacteria. *Microbial Ecology*, 13(2):95–101, 1987.

[196] M. Nowak and P. Schuster. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet. *J. theor. Biol*, 137:375–395, 1989.

[197] M.A. Nowak. *Evolutionary dynamics: exploring the equations of life.* Belknap Press, 2006.

[198] H Ohara, K Hiyama, and T Yoshida. Kinetics of growth and lactic acid production in continuous and batch culture. *Appl Microbiol Biotechnol*, 37(5):544–548, 1992.

[199] A. Oliver, R. Cantón, P. Campo, F. Baquero, and J. Blázquez. High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*, 288(5469):1251, 2000.

[200] A. Oliver, R. Cantón, P. Campo, F. Baquero, and J. Blázquez. High frequency of hypermutable pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*, 288(5469):1251, 2000.

[201] R.J. Olson, C. Watras, and S.W. Chisholm. Patterns of individual cell growth in marine centric diatoms. *Microbiology*, 132(5):1197, 1986.

[202] H.A. Orr. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52(4):935–949, 1998.

[203] H.A. Orr. The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4):1519, 2003.

[204] S.W. Pacala, C.D. Canham, J. Saponara, J.A. Silander Jr, R.K. Kobe, and E. Ribbens. Forest models defined by field measurements: estimation, error analysis and dynamics. *Ecological Monographs*, 66(1):1–43, 1996.

[205] DF Parkhurst and OL Loucks. Optimal leaf size in relation to environment. *The Journal of Ecology*, 60(2):505–537, 1972.

[206] F. Partensky and L. Garczarek. Prochlorococcus: advantages and limits of minimalism. *Annual Review of Marine Science*, 2:305–331, 2010.

[207] L. Perfeito, L. Fernandes, C. Mota, and I. Gordo. Adaptive mutations in bacteria: high rate and small effects. *Science*, 317(5839):813, 2007.

[208] A.P. Petroff, M.S. Sim, A. Maslov, M. Krupenin, D.H. Rothman, and T. Bosak. Biophysical basis for the geometry of conical stromatolites. *Proceedings of the National Academy of Sciences*, 107(22):9956, 2010.

[209] C. Picioreanu, M.C.M. Van Loosdrecht, and J.J. Heijnen. Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. *Biotechnology and bioengineering*, 58(1):101–116, 1998.

[210] P Pilat and A Prokop. The effect of methanol, formaldehyde, and formic acid on growth of candida boidinii 11 bh. *Biotechnology and bioengineering*, 17(12):1717–1728, 1975.

[211] SJ Pirt. The maintenance energy of bacteria in growing cultures. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 163(991):224–231, 1965.

[212] I Poole, J D B Weyers, T Lawson, and J A Raven. Variations in stomatal density and index: Implications for palaeoclimatic reconstructions. *Plant, Cell & Environment*, 19(6):705–712, 2006.

[213] I.C. Prentice, W. Cramer, S.P. Harrison, R. Leemans, R.A. Monserud, and A.M. Solomon. Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 19(2):117–134, 1992.

[214] H Pretzsch and T Mette. Linking stand-leval self-thinning allometry to the tree-leval leaf biomass allometry. *Trees*, 22:611–622, 2008.

[215] C.A. Price, B.J. Enquist, and V.M. Savage. A general model for allometric covariation in botanical form and function. *Proceedings of the National Academy of Sciences USA*, 104(32):13204–13209, 2007.

[216] C.A. Price, J.F. Gilooly, A.P. Allen, J.S. Weitz, and K.J. Niklas. The metabolic theory of ecology: prospects and challenges for plant biology. *New Phytologist*, 188:696–710, 2010.

[217] C.A. Price, K. Ogle, E.P. White, and J.S. Weitz. Evaluating scaling models in biology using hierarchical Bayesian approaches. *Ecology letters*, 12(7):641–651, 2009.

[218] Oregon State University PRISM Climate Group. http://www.ncdc.noaa.gov/oa/climate/stationlocator.html. Online, December 2007. Database accessed on 16 December 2007.

[219] D. Purves and S. Pacala. Predictive models of forest dynamics. *Science*, 320(5882):1452–1453, 2008.

[220] D.W. Purves, J.W. Lichstein, and S.W. Pacala. Crown plasticity and competition for canopy space: a new spatially implicit model parameterized for 250 North American tree species. *PloS One*, 2(9):e870, 2007.

[221] D.W. Purves, J.W. Lichstein, N. Strigul, and S.W. Pacala. Predicting and understanding forest dynamics using a simple tractable model. *Proceedings of the National Academy of Sciences USA*, 105(44):17018–17022, 2008.

[222] M. Radman, I. Matic, and F. Taddei. Evolution of evolvability. *Annals of the New York Academy of Sciences*, 870(1):146–155, 1999.

[223] I. Ramos, L.E.P. Dietrich, A. Price-Whelan, and D.K. Newman. Phenazines affect biofilm formation by¡ i¿ pseudomonas aeruginosa¡/i¿ in similar ways at various scales. *Research in microbiology*, 161(3):187–191, 2010.

[224] H.D Ratcliffe, J.W Drozd, and A.T Bull. Growth energetics of rhizobium leguminosarum in chemostat culture. *Journal of General Microbiology*, 129(6):1697–1706, 1983.

[225] DA Ratkowsky, RK Lowry, TA McMeekin, AN Stokes, and RE Chandler. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *Journal of Bacteriology*, 154(3):1222–1226, 1983.

[226] C. Rayssiguier, D.S. Thaler, and M. Radman. The barrier to recombination between escherichia coli and salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature*, 342(6248):396–401, 1989.

[227] A Razvi, Z Zhang, and C.Q Lan. Effects of glucose and nitrogen source concentration on batch fermentation kinetics of lactococcus lactis under hemin-stimulated respirative condition. *Biotechnology progress*, 24(4):852–858, 2008.

[228] D.A. Recinos, M.D. Sekedat, A. Hernandez, T.S. Cohen, H. Sakhtah, A.S. Prince, A. Price-Whelan, and L.E.P. Dietrich. Redundant phenazine operons in pseudomonas aeruginosa exhibit environment-dependent expression and differential roles in pathogenicity. *Proceedings of the National Academy of Sciences*, 109(47):19420–19425, 2012.

[229] J.L. Reed and B.Ø. Palsson. Thirteen years of building constraint-based in silico models of escherichia coli. *Journal of Bacteriology*, 185(9):2692–2699, 2003.

[230] G.T. Reeves, R. Kalifa, D.E. Klein, M.A. Lemmon, and S.Y. Shvartsman. Computational analysis of EGFR inhibition by Argos. *Developmental biology*, 284(2):523–535, 2005.

[231] G. Reshes, R. Tsukanov, S. Vanounou, I. Fishov, and M. Feingold. Timing the start of division in E. coli: a single-cell study. *Biophysical Journal*, 96(3S1):631–631, 2009.

[232] G. Reshes, S. Vanounou, I. Fishov, and M. Feingold. Timing the start of division in E. coli: a single-cell study. *Physical Biology*, 5:046001, 2008.

[233] M Rieger, O Käppeli, and A Fiechter. The role of limited respiration in the incomplete oxidation of glucose by saccharomyces cerevisiae. *Microbiology*, 129(3):653–661, 1983.

[234] E.P. C. Rocha, I. Matic, and F. Taddei. Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? *Nucleic Acids Research*, 30(9):1886–1894, 2002.

[235] D Rosen, M Edelman, E Galun, and D Danon. Biogenesis of mitochondria in trichoderma viride: Structural changes in mitochondria and other spore constituents during conidium maturation and germination. *Journal of General Microbiology*, 83(1):31–49, 1974.

[236] JR Roth, E. Kugelberg, AB Reams, E. Kofoid, and DI Andersson. Origin of mutations under selection: the adaptive mutation controversy. *Annual Review of Microbiology*, 60:477–501, 2006.

[237] D L Royer, P Wilf, D A Janesko, E A Kowalski, and D L Dilcher. Correlations of climate and plant ecology to leaf size and shape: potential proxies for the fossil record. *American Journal of Botany*, 92(7):1141–1151, 2005.

[238] S.W. Running and J.C. Coughlan. A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modeling*, 42(2):125–154, 1988.

[239] S.W. Running, R.R. Nemani, D.L. Peterson, L.E. Band, D.F. Potts, L.L. Pierce, and M.A. Spanner. Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation. *Ecology*, 70(4):1090–1101, 1989.

[240] E.S. Russell. *Form and function: A contribution to the history of animal morphology.* EP Dutton, 1917.

[241] M G Ryan and B J Yoder. Hydraulic limits to tree height and tree growth. *Bioscience*, 47(4):235–242, 1997.

[242] E Salas, H Ozier-Lafontaine, and P Nygren. A fractal root model applied for estimating the root biomass and architecture in two tropical legume tree species. *Annals of Forest Science*, 61(4):337–345, 2004.

[243] H Sansawa and H Endo. Production of intracellular phytochemicals in chlorella under heterotrophic conditions. *Journal of bioscience and bioengineering*, 98(6):437–444, 2004.

[244] L.A Sarubbo, M.C Marçal, M.L.C Neves, M.P.C Silva, L.F Porto, and G.M Campos-Takaki. Bioemulsifier production in batch culture using glucose as carbon source by candida lipolytica. *Applied biochemistry and biotechnology*, 95(1):59–67, 2001.

[245] U. Sauer, V. Hatzimanikatis, H.P. Hohmann, M. Manneberg, AP Van Loon, and J.E. Bailey. Physiology and metabolic fluxes of wild-type and riboflavin-producing Bacillus subtilis. *Applied and environmental microbiology*, 62(10):3687, 1996.

[246] V.M. Savage, L.P. Bentley, B.J. Enquist, J.S. Sperry, D.D. Smith, P.B. Reich, and E.I. von Allmen. Hydraulic trade-offs and space filling enable better predictions of vascular structure and function in plants. *Proceedings of the National Academy of Sciences*, 107(52):22722–22727, 2010.

[247] V.M. Savage, J.F. Gillooly, J.H. Brown, G.B. West, and E.L. Charnov. Effects of body size and temperature on population growth. *The American Naturalist*, 163(3):429–441, 2004.

[248] S. Sawai, P.A. Thomason, and E.C. Cox. An autoregulatory circuit for long-range self-organization in Dictyostelium cell populations. *Nature*, 433(7023):323–326, 2005.

[249] K Schmidt-Nielsen. *Scaling, why is animal size so important?* Cambridge Univ Pr, 1984.

[250] E. Schrödinger. *What is life?: the physical aspect of the living cell; with, Mind and matter; & Autobiographical sketches.* Cambridge Univ Pr, 1992.

[251] K.L. Schulze and R.S. Lipe. Relationship between substrate concentration, growth rate, and respiration rate of escherichia coli in continuous culture. *Archives of Microbiology*, 48(1):1–20, 1964.

[252] A.C. Shaver, P.G. Dombrowski, J.Y. Sweeney, T. Treis, R.M. Zappala, and P.D. Sniegowski. Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations. *Genetics*, 162(2):557–566, 2002.

[253] A.C. Shaver and P.D. Sniegowski. Spontaneously arising mutL mutators in evolving Escherichia coli populations are the result of changes in repeat length. *Journal of bacteriology*, 185(20):6076, 2003.

[254] MG Shepherd and PA Sullivan. The production and growth characteristics of yeast and mycelial forms of Candida albicans in continuous culture. *Microbiology*, 93(2):361, 1976.

[255] G.S. Shreve and T.M. Vogel. Comparison of substrate utilization and growth kinetics between immobilized and suspended pseudomonas cells. *Biotechnology and bioengineering*, 41(3):370–379, 1993.

[256] ML Shuler, S. Leung, and CC Dick. A mathematical model for the growth of a single bacterial cell. *Annals of the New York Academy of Sciences*, 326(1):35–52, 1979.

[257] BJ Shuter. Size dependence of phosphorus and nitrogen subsistence quotas in unicellular microorganisms. *Limnology and Oceanography*, 23:1248–1255, 1978.

[258] M.C. Simon and B. Keith. The role of oxygen availability in embryonic development and stem cell function. *Nature Reviews Molecular Cell Biology*, 9(4):285–296, 2008.

[259] W B Smith. Forest inventory and analysis: a national inventory and monitoring program. *Environmental Pollution*, 116:S233–S242, 2002.

[260] W K Smith. Temperatures of desert plants: another perspective on the adaptability of leaf size. *Science*, 201(4356):614–616, 1978.

[261] P.D. Sniegowski and P.J. Gerrish. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544):1255–1263, 2010.

[262] P.D. Sniegowski, P.J. Gerrish, T. Johnson, and A. Shaver. The evolution of mutation rates: separating causes from consequences. *Bioessays*, 22(12), 2000.

[263] PD Sniegowski, PJ Gerrish, and RE Lenski. Evolution of high mutation rates in experimental populations of Escherichia coli. *Nature*, 387(6634):703–705, 1997.

[264] L.J. Stal. Physiological ecology of cyanobacteria in microbial mats and other communities. *New Phytologist*, 131(1):1–32, 1995.

[265] G Stanhill. Some results of helicopter measurements of the albedo of different land surfaces. *Solar Energy*, 13(1):59–66, 1970.

[266] P.S. Stewart. Diffusion in biofilms. *Journal of Bacteriology*, 185(5):1485–1491, 2003.

[267] S.E. Strand and A.J. McDonnell. Mathematical analysis of oxygen and nitrate consumption in deep microbial films. *Water Research*, 19(3):345–352, 1985.

[268] N. Strigul, D. Pristinski, D. Purves, J. Dushoff, and S. Pacala. Scaling from trees to forests: tractable macroscopic equations for forest dynamics. *Ecological Monographs*, 78(4):523–545, 2008.

[269] IJ Sud and M Schaechter. Dependence of the content of cell envelopes on the growth rate of bacillus megaterium. *Journal of bacteriology*, 88(6):1612–1617, 1964.

[270] RJ Summers, DP Boudreaux, and VR Srinivasan. Continuous cultivation for apparent optimization of defined media for cellulomonas sp. and bacillus cereus. *Applied and environmental microbiology*, 38(1):66–71, 1979.

[271] W.G Sunda and S.A Huntsman. Interrelated influence of iron, light and cell size on marine phytoplankton growth. *Nature*, 390(6658):389–392, 1997.

[272] F. Taddei, M. Radman, J. Maynard-Smith, B. Toupance, PH Gouyon, and B. Godelle. Role of mutator alleles in adaptive evolution. *Nature*, 387(6634):700–702, 1997.

[273] S. Tännler, S. Decasper, and U. Sauer. Maintenance metabolism and carbon fluxes in bacillus species. *Microbial cell factories*, 7(1):19, 2008.

[274] O. Tenaillon, F. Taddei, M. Radman, and I. Matic. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Research in Microbiology*, 152(1):11–16, 2001.

[275] O. Tenaillon, B. Toupance, H. Le Nagard, F. Taddei, and B. Godelle. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics*, 152(2):485–493, 1999.

[276] C.J. Tomlin and J.D. Axelrod. Biology by numbers: mathematical modelling in developmental biology. *Nature reviews genetics*, 8(5):331–340, 2007.

[277] Q.H. Tran and G. Unden. Changes in the proton potential and the cellular energetics of Escherichia coli during growth by aerobic and anaerobic respiration or by fermentation. *European Journal of Biochemistry*, 251(1-2):538–543, 1998.

[278] A.M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37, 1952.

[279] J.J. Tyson. Modeling the cell division cycle: cdc2 and cyclin interactions. *Proceedings of the National Academy of Sciences*, 88(16):7328, 1991.

[280] M Veenhuis, I Keizer, and W Harder. Characterization of peroxisomes in glucose-grown hansenula polymorpha and their development after the transfer of cells into methanol-containing media. *Archives of Microbiology*, 120(2):167–175, 1979.

[281] J. Vidal. Physioecology of zooplankton. I. Effects of phytoplankton concentration, temperature, and body size on the growth rate of Calanus pacificus and Pseudocalanus sp. *Marine Biology*, 56(2):111–134, 1980.

[282] L. Von Bertalanffy. Quantitative laws in metabolism and growth. *The Quarterly Review of Biology*, 32(3):217–231, 1957.

[283] G. Von Dassow, E. Meir, E.M. Munro, and G.M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000.

[284] U. Von Stockar, T. Maskow, J. Liu, I.W. Marison, and R. Patino. Thermodynamics of microbial growth and metabolism: an analysis of the current situation. *Journal of Biotechnology*, 121(4):517–533, 2006.

[285] W De Vries, W Kapteijn, EG Van Der Beek, and AH Stouthamer. Molar growth yields and fermentation balances of lactobacillus casei l3 in batch cultures and in continuous cultures. *Microbiology*, 63(3):333–345, 1970.

[286] T Walther and H Reinsch. Mathematical modeling of regulatory mechanisms in yeast colony development. *Journal of theoretical biology*, 229(3):327–338, 2004.

[287] R.H. Waring and SW Running. *Forest ecosystems: analysis at multiple scales.* Academic Press, 2007.

[288] AD Warth. Relationship between the heat resistance of spores and the optimum and maximum growth temperatures of bacillus species. *Journal of bacteriology*, 134(3):699–705, 1978.

[289] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[290] H.G Weger. Ferric and cupric reductase activities in the green alga chlamydomonas reinhardtii: experiments using iron-limited chemostats. *Planta*, 207(3):377–384, 1999.

[291] G.B. West, J.H. Brown, and B.J. Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122, 1997.

[292] G.B. West, J.H. Brown, and B.J. Enquist. A general model for the structure and allometry of plant vascular systems. *Nature*, 400(6745):664–667, 1999.

[293] G.B. West, J.H. Brown, and B.J. Enquist. A general model for ontogenetic growth. *Nature*, 413(6856):628–631, 2001.

[294] G.B. West, B.J. Enquist, and J.H. Brown. A general quantitative theory of forest structure and dynamics. *Proceedings of the National Academy of Sciences*, 106(17):7040–7045, 2009.

[295] G.B. West, W.H. Woodruff, and J.H. Brown. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proceedings of the National Academy of Sciences*, 99(90001):2473–2478, 2002.

[296] M. Westoby. The self-thinning rule. *Advances in ecological research*, 14:167–226, 1984.

[297] C.R. White and R.S. Seymour. Mammalian basal metabolic rate is proportional to body mass2/3. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7):4046, 2003.

[298] I. Wiegand, A.K. Marr, E. Breidenstein, K.N. Schurek, P. Taylor, and R.E.W. Hancock. Mutator genes giving rise to decreased antibiotic susceptibility in pseudomonas aeruginosa. *Antimicrobial agents and chemotherapy*, pages AAC–00233, 2008.

[299] FM Williams. A model of cell growth dynamics. *Journal of theoretical biology*, 15(2):190–207, 1967.

[300] RJ Woods, JE Barrick, TF Cooper, U Shrestha, MR Kauth, and RE Lenski. Second-order selection for evolvability in a large escherichia coli population. *Science*, 331(6023):1433–1436, 2011.

[301] D. Worlitzsch, R. Tarran, M. Ulrich, U. Schwab, A. Cekici, K.C. Meyer, P. Birrer, G. Bellon, J. Berger, T. Weiss, et al. Effects of reduced mucus oxygen concentration in airway pseudomonas infections of cystic fibrosis patients. *Journal of Clinical Investigation*, 109(3):317–325, 2002.

[302] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress on genetics*, 1(6):356–366, 1932.

[303] C.S. Wylie, C.M. Ghim, D. Kessler, and H. Levine. The fixation probability of rare mutators in finite asexual populations. *Genetics*, 181(4):1595, 2009.

[304] J.B. Xavier and K.R. Foster. Cooperation and conflict in microbial biofilms. *Proceedings of the National Academy of Sciences*, 104(3):876, 2007.

[305] J.B. Xavier, C. Picioreanu, and M.C.M. van Loosdrecht. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environmental Microbiology*, 7(8):1085–1103, 2005.

[306] K. Yoda, T. Kira, H. Ogawa, and K. Hozumi. Self-thinning in overcrowded pure stands under cultivated and natural conditions. *Journal of Biology of Osaka City University*, 14:107–129, 1963.

[307] T Yoshida, N. G Hairston, and S. P Ellner. Evolutionary trade-off between defence against grazing and competitive ability in a simple unicellular alga, chlorella vulgaris. *Proceedings of the Royal Society B: Biological Sciences*, 271(1551):1947–1953, Sep 2004.

[308] A. Zaslaver, S. Kaplan, A. Bren, A. Jinich, A. Mayo, E. Dekel, U. Alon, and S. Itzkovitz. Invariant distribution of promoter activities in Escherichia coli. *PLoS Comput Biol*, 5(10), 2009.