The evolution of metabolic dependency in bacteria

Dissertation

To Fulfill the
Requirements for the Degree of
"Doctor of Philosophy" (PhD)

Submitted to the Council of the Faculty
of Biology and Pharmacy
of the Friedrich Schiller University Jena

by
M.Sc. Glen D'Souza
born on February 3rd, 1989 in Neerude, India

Das Promotionsgesuch wurde eingereicht und bewilligt am:

Gutachter:
1) Dr. Christian Kost, Max Planck Institute for Chemical Ecology, Jena, Germany
2) Dr. Akos Kovacs, Institute of Microbiology, Friedrich-Schiller-University, Jena, Germany
3) Prof. Stuart West, Department of Zoology, Oxford University, United Kingdom

Das Promotionskolloquium wurde abgelegt am: 13th June, 2016

To Ella for life's several learnings and the gift of education
To Jerry for the strength, motivation, imagination and realizing every possible need
To Shona for being there, in every possible way, every time

"Somewhere, something incredible is waiting to be known."

— Carl Sagan

Table of contents

Overview of Manuscripts

Overview of Manuscripts

This thesis is based on the following manuscripts.

Author abbreviations: Glen D'Souza (GD), Silvio Waschina (SW), Christoph Kaleta, (CKa), Christian Kost (CKo), Samay Pande (SP), and Katrin Bohl (KB)

1. Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria.

Authors: Glen D'Souza, Silvio Waschina, Samay Pande, Katrin Bohl, Christoph Kaleta, and Christian Kost

Published in: Evolution (2014); Volume 68; Pages: 2559-2570;

DOI: 10.1111/evo.12468.

Presented as: Chapter 1

Author contributions:

Conceived the project	CKo 100%
Designed the experiments	GD 50%, CKo 50%
Performed all experiments	GD 100%
Constructed *A. baylyi* strains	SP 100%
Analysed the data	GD 56%, SW 40%, CKo 4%
Conceived *in silico* analysis	CKo 100%
Designed *in-silico* analyses	SW 20%, CKa 70%, CKo 10%
Performed in-silico analyses	SW 95%, KB 5%
Wrote the manuscript	GD 40%, SW 20%, CKa 10%, CKost (30%)

2. Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes.

Authors: Glen D'Souza, Silvio Waschina, Christoph Kaleta, and Christian Kost

Published in: Evolution (2015); Volume 69; Pages: 1244-1254;

DOI:10.1111/evo.12640.

Presented as: Chapter 2

Author contributions:

Conceived the project	GD 50%, CKo 50%
Designed the experiments	GD 70%, CKo 25%, SW 5%
Performed all experiments	GD 100%
Analysed experimental data	GD 95%, CKo 5%
Conceived *in silico* analysis	GD 40%, CKo 40%, SW 20%
Designed *in-silico* analyses	SW 75%, CKa 25%
Implemented and performed *in-silico* analyses	SW 100%
Analysed *in-silico* predictions	SW 90%, CKa 10%
Wrote the manuscript	GD 70%, SW 15%, CKo% 12.5%, CKa 2.5%

3. Metabolic network architecture and carbon source determine metabolite production costs

Authors: Silvio Waschina, Glen D'Souza, Christian Kost, and Christoph Kaleta

In major revision at The FEBS Journal (as of 14th December 2015)

Presented as: Chapter 3

Author contributions:

Conceived the project	SW 60%, CKo 20%, CKa 20%
Designed the experiments	SW 90%, GD 10%
Performed all experiments	SW 100%
Analysed experimental data	SW 100%
Designed auxotrophic strains	GD 100%
Conceived *in silico* analysis	SW 75%, CKa 20%, CKo 5%
Designed *in-silico* analyses	SW 100%
Implemented and performed *in-silico* analyses	SW 100%
Analysed *in-silico* predictions	SW 76%, GD 8%, CKo 8%, CKa 8%
Wrote the manuscript	SW 85%, GD 5%, CKo 5%, CKa 5%

4. Experimental evolution of metabolic dependency in bacteria

Authors: Glen D'Souza and Christian Kost

Manuscript to be submitted to PLoS Biology

Presented as: Chapter 4

Author contributions:

Conceived the project	GD 50%, CKo 50%
Designed the experiments	GD 50%, CKo 50%
Performed all experiments	GD 100%
Analysed experimental data	GD 100%
Wrote the manuscript	GD 60%, CKo 40%

Glossary of terms

The terms that are listed here occur multiple times throughout the thesis. The definitions denote the context of their usage in the text.

Biosynthetic or metabolic cost is the amount of a certain resource (e.g. carbon source, ATP or protein machinery) that is required to produce a certain metabolite.

Clusters of orthologous genes (COGs) are groups of three or more orthologous genes which are evolutionarily related and are thus considered to be descendants of the same ancestral domain.

Community structure is a broad term encompassing the composition of a community in terms of the diversity of different species present and their abundances.

Community function is the ecological function performed by the community.

Conditionally essential genes are genes which are required for the growth of an organism in a given environment but not required in another environment.

Environmental plasticity is the differential response of a genotype to distinct sets of environments.

Gene family is a set of several genes which have similar biological functions.

Metabolic network is the entire set of reactions that comprise all the metabolic pathways in a cell.

Metabolic flux encompasses the rate at which enzymes convert the substrate into the product to feed downstream reactions in a pathway.

Life history trait is any trait which influences the survival and reproduction of an organism in a given environment.

Muller's Ratchet describes a scenario where many deleterious mutations can accumulate irreversibly in a population of individuals.

Mutualism is an interaction between two species in which each species derives a benefit.

Niche encompasses the entire set of abiotic and biotic conditions in which a species, genotype or individual can persist.

Nutritional or metabolic environment describes the entire set of catabolites and anabolites that are available to a growing population of cells in their immediate environment.

Population bottlenecks are the drastic reductions in population sizes of a resident population.

Summary

Bacteria that have adapted to nutrient rich or stable environments often have reduced genome sizes. A consequence of this reductive evolution in such specialized bacteria is that a loss of genes is accompanied with the loss of biosynthetic functions, which deems these bacteria auxotrophic and thus unable to grow in the absence of an environmental supplementation of the nutrient that they require. Examples of such bacteria with reduced biosynthetic capabilities include Lactic acid bacteria, endosymbiotic bacteria and even seemingly free-living bacterial species, and thus span across a range of diverse environments and lifestyles. What drives the loss of genes and concomitantly the loss of biosynthetic functions in these micro-organisms? Two potential mechanisms have been generally attributed to account for these losses: genetic drift and natural selection. However, experimental studies to uncover the role of either mechanism in driving biosynthetic disarmament in bacteria have rarely been undertaken, and thus the occurrence of auxotrophies begs an explanation. This thesis aims to address this intriguing loss of biosynthetic abilities in bacteria and unravel the evolutionary mechanisms causing these losses.

Biosynthetic gene loss is widespread and selection can explain these losses. As a first step, a quantitative measure of auxotrophies in the microbial world was gained based on the presence or absence of biosynthetic genes from the set of all completely sequenced and well-annotated genomes. An analysis of genomes and the corresponding metabolic networks of 949 (subsequently updated to 1432) bacterial species, revealed that 76% of the surveyed Eubacterial species with both symbiotic and free-living lifestyles lack the genes required to biosynthesize one or multiple amino acids, nucleobases and vitamins or co-factors. Such a loss of biosynthetic functions and the accompanying dependency on the environment for nutrient uptake is intriguing since these functions are essential for an autonomous cellular functioning. To determine if selection can explain these losses, the fitness consequences of losing biosynthetic genes relative to a prototrophic ancestor were measured in synthetically constructed auxotrophic mutants of *Escherichia coli* and *Acinetobacter baylyi* where genes responsible for biosynthesis of amino acids or nucleobases or vitamins were deleted. Almost all of the

auxotrophic mutants tested were generally fitter compared to prototrophic strains, in presence of an environmental supplementation of amino acids. Furthermore, deletion of different biosynthetic genes, which resulted in the same amino acid auxotrophies but theoretically entailed distinct protein costs, resulted in distinct fitness consequences for the auxotrophic bacteria, suggesting that these selective benefits originate in part from a general saving of costs associated with biosynthetic metabolism. These findings strongly suggested that selective advantages can explain the rampant loss of biosynthetic functions in bacteria in nature.

Environmental plasticity and epistasis influence the selective advantages of gene loss. Since many eubacterial genomes lack multiple biosynthetic functions, the fitness effects of harboring multiple auxotrophies were subsequently evaluated. In general, an increase in auxotrophic mutations did not result in increased fitness benefits suggesting epistasis was prevalent and resulted in a trend of diminishing returns. Furthermore, the patterns of epistasis and fitness magnitudes were influenced by the carbon source that was provided. These findings suggest that the metabolic environment and epistasis can influence selection mediated biosynthetic gene loss resulting in distinct auxotrophic combinations being selected over others in a given environmental condition. This argument was in line with the findings that some combinations of auxotrophies occurred more often than expected randomly in sequenced bacterial genomes and suggests that the metabolic environment can play a crucial role in the evolution of auxotrophies.

General trade-offs in metabolism and network architecture influence fitness benefits in auxotrophic bacteria. Different carbon sources can lead to distinct fluxes through a metabolic network and thus distinct metabolic costs of biosynthesis will be incurred depending on the carbon source. This can explain why different carbon sources result in distinct selective benefits for the auxotrophic strains. By employing a flux balance analysis on the basis of the metabolic network of *E. coli* the costs of metabolite biosynthesis that are incurred when using different carbon sources were estimated. This analysis revealed that the architecture of the metabolic network plays an important role in determining the metabolic costs associated with biosynthesis of a metabolite, in this case amino acids. These predictions were then experimentally verified using auxotrophic

strains of *E. coli* and the findings strongly indicate the carbon source used and its position in a metabolic network can dictate the growth benefits in auxotrophic strains. This implies that trade-offs in central metabolism can result in distinct sets of auxotrophs to specialize on distinct carbon resources to optimize their fitness gains upon gene loss.

Rapid adaptive loss of biosynthetic functions in nutrient-rich environments. The observed selective advantages associated with the loss of biosynthetic functions suggest that auxotrophic genotypes should readily evolve in environments where the metabolite is externally supplemented. Subsequently, replicate populations of a prototrophic strain of *Escherichia coli* were serially propagated for 2,000 generations in an environment containing all 20 amino acids. Indeed, in amino acid-rich environments auxotrophs rapidly evolved and their occurrence increased over time. Surprisingly, auxotrophs also evolved when no amino acids were present. In all cases the loss of biosynthetic functions was adaptive and was a result of diverse mutations in both structural and regulatory genes. Interestingly, auxotrophs lost multiple biosynthetic functions but always co-existed with prototrophic strains by negative frequency-dependent selection, suggesting that auxotrophs derived amino acids from prototrophic strains by cross-feeding in addition to the ones that were environmentally supplied. Importantly, these results provide strong evidence that the loss of biosynthetic genes and functions in bacteria observed in nature can arise as a result of adaptive processes when the requisite metabolite is provisioned by the biotic or abiotic environment.

In summary, in this thesis using a combination of synthetic ecology, experimental evolution and *in silico* computations of genomes and metabolic networks, the causes underlying the biosynthetic disarmament and the factors that influence the ensuing dynamics were investigated. Using these findings, I make the case for adaptation as a major force driving the loss of biosynthetic functions in microbes in nature and for potentiating the evolution of metabolic interactions and networks within microbial communities.

Zusammenfassung

Die Anpassung von Bakterien an nährstoffreiche Umgebungen ist oft verbunden mit einer Verkleinerung der Genomgröße. Mit dem Verlust von genetischer Information verlieren die entsprechenden Bakterien auch biosynthetische Fähigkeiten. Die Metabolite, die so nicht mehr synthetisiert werden können, aber dennoch essentiell für Zellwachstum und -teilung sind, müssen dementsprechend aus der Umgebung aufgenommen werden. Mikroorganismen, die solche Metabolite nicht mehr selbstständig produzieren können werden *auxotroph* genannt. Milchsäurebakterien und Endosymbionten sind typische Beispiele für auxotrophe Bakterien, die im Laufe der Evolution einen Großteil ihrer biosynthetischen Fähigkeiten verloren haben. Aber auch scheinbar freilebelende Arten fehlen häufig die entsprechenden Gene um bestimmte wachstumsrelevante Metabolite herzustellen. Was beeinflusst den Verlust von biosynthetischen Funktionen in diesen Mikroorganismen? Häufig wird der Verlust auf zwei evolutionäre Prozesse zurückgeführt: Genetischer Drift und natürliche Selektion. Wie stark der jeweilige Einfluss dieser beiden Prozesse jedoch tatsächlich ist wurde bislang nur marginal experimentell untersucht. Die vorliegende Arbeit beschäftigt sich mit den Fragen wie der verblüffend häufige Verlust von biosynthetischen Funktionen in Bakterien erklärt werden kann und welche evolutionären Mechanismen dabei eine Rolle spielen.

Der Verlust von biosynthetischen Genen ist weitverbreitet in Bakterien und kann durch natürliche Selektion erklärt werden. Um die Häufigkeit von Auxotrophien in natürlich vorkommenden Bakterien abzuschätzen wurde untersucht ob bestimmte Gene mit biosynthetischen Funktion in komplett sequenzierten und annotierten Genomen von 949 (später aktualisiert auf 1432) Bakterien vorkommen. Auf diese Weise konnte theoretisch vorhergesagt werden, ob die Stoffwechselwege für die Synthese von Metaboliten, die notwendig für das Zellwachstum sind, in den metabolischen Netzwerken der entsprechenden Arten vorhanden sind. Diese Analyse ergab, dass 76% der untersuchten Eubakterien die Fähigkeit fehlt eine oder mehrere Metabolite (einschließlich Aminosäuren, Nukleotide, Vitamine und Kofaktoren) eigenständig zu synthetisieren. Die untersuchten Metabolite sind essentiell für alle lebenden Zellen und

der Verlust von biosynthetischen Stoffwechselwegen stellt somit eine absolute Abhängigkeit der auxotrophen Bakterien an die Nährstoffzusammensetzung ihrer natürlichen Umgebung dar. Weiterhin wurde untersucht, ob natürliche Selektion den biosynthetischen Funktionsverlust erklären kann. Dafür wurden synthetisch auxotrophe Mutanten von *Escherichia coli* und *Acinetobacter baylyi* erzeugt, indem Gene aus dem Genom herausgeschnitten wurden, die essentiell für die Produktion von einer Aminosäure, einem Nukleotid oder einem Vitamin sind. Anschließend wurde die Fitness dieser auxotrophen Mutanten relativ zu dem jeweiligen prototrophen Wildtyp, der alle Metabolite weiterhin eigenständig produzieren kann, bestimmt. Nahezu alle untersuchten Mutanten zeigten eine höhere Fitness als der prototrophe Wildtypstamm wenn der entsprechende Metabolit, für den der mutierte Genotyp auxotroph ist, dem Wachstumsmedium hinzugefügt wurde. Des Weiteren wurden verschiedene Gene innerhalb des gleichen Biosyntheseweges aus dem Genom des Wildtyps entfernt. Hierbei zeigten sich signifikante Fitnessunterschiede zwischen verschiedenen auxotrophen Mutanten obwohl sie alle auxotroph für denselben Metaboliten sind. Da die Enzyme eines Stoffwechselweges oft unterschiedlich groß sind und in verschiedenen Mengen in der Zelle vorkommen spricht dieses Ergebnis dafür, dass Proteinkosten, welche mit der Produktion der Enzyme verbunden sind, maßgeblich den Fitnessvorteil von auxotrophen Genotypen beeinflussen.

Plastizität der Umwelt sowie Epistasis beeinflussen den Selektionsvorteil von Genverlust. Ausgehend von dem vorherigen Ergebnis, dass viele Bakterienarten wahrscheinlich auxotroph für mehrere Metabolite sind, wurde auch der Fitnesseffekt von multiplen Gendeletionen, die wiederum zu multiplen Auxotrophien eines mutierten Genotyps führen, evaluiert. Im Allgemeinen führte der Verlust von mehreren biosynthetischen Funktionen zu einem Fitnessvorteil der auxotrophen Genotypen, der niedriger ausfiel als erwartet. Daraus lässt sich ableiten, dass häufig epistatische Effekte zwischen den Mutationen auftreten, die unterschiedliche Auxotrophies verursachen, und dabei den Fitnessvorteil von biosynthetischen Funktionsverlusts abnimmt falls der Genotyp bereits auxotroph für einen anderen Metaboliten ist – ein Effekt der *abnehmender Ertrag* (engl. *diminishing returns*) genannt wird. Der Einfluss von Epistasis war wiederum auch davon abhängig, welche Kohlenstoffquelle für das Zellwachstum zur

Verfügung stand. Diese Ergebnisse deuten darauf hin, dass die Art der Kohlenstoffquelle sowie Epistasis den adaptiven Verlust von Genen mit biosynthetischer Funktion beeinflussen, was wiederum in Übereinstimmung mit den genombasierten Auxotrophievorhersagen ist, bei denen bestimmte Kombinationen von Auxotrophies häufiger vorkommen als erwartet.

In nährstoffreichen Umgebungen verliert *Escherichia coli* biosynthetische Fähigkeiten in evolutionär relativ kurzem Zeitraum. Aufgrund der enormen Fitnessvorteile, die mit Auxotrophiemutationen verbunden sind, ist zu erwarten, dass die Häufigkeit von auxotrophe Genotypen sich innerhalb einer Population stark ansteigt wenn die entsprechenden Metabolite aus der Umgebung aufgenommen werden können. Um diese Hypothese zu testen, wurden zwölf unabhängige Populationen des anfänglich prototrophen *E. coli* Stammes für 2.000 Generationen täglich in frisches Wachstumsmedium übertragen, das alle 20 proteinogenen Aminosäuren enthielt. Auxotrophe Genotypen evolvierten schnell und stiegen in ihrer Häufigkeiten im Laufe des Experiments. Interessanterweise evolvierten auch auxotrophe Stämme in dem Kontrollexperiment, bei dem keine Aminosäuren dem Wachtsumsmedium hinzugefügt wurden. Es konnte darüber hinaus auch hier gezeigt werden, dass der Verlust der biosynthetischen Funktionen adaptiv ist und die Folge von diversen Mutationen in Struktur- und Regulationsgenen. Auxotrophe Genotypen verloren mehrere biosynthetische Funktionen aber koexistierten immer zusammen mit prototrophen Stämmen, was daruf hindeutet, dass auxotrophe Genotypen auch Aminosäuren von prototrophen Zellen beziehen, die diese Aminosäuren weiterhin produzieren. Zusammenfassend zeigt dieses Evolutionsexperiment, dass die Evolution von auxotrophen Bakterien in der Natur durch adaptive Prozesse erklärt werden kann wenn die entsprechenden Metabolite aus der biotischen- und/oder abiotischen Umgebungen bezogen werden können.

Die Struktur des metabolischen Netzwerkes beeinflusst den Fitnessvorteil von auxotrophen Bakterien. Chemisch verschiedene Kohlenstoffquellen verursachen unterschiedliche Verteilungen von Reaktionsflüssen durch das metabolische Netzwerk. Somit können wiederum die metabolischen Kosten, die mit der Produktion von

Metaboliten verbunden sind, auch zwischen verschiedenen Kohlenstoffquellen variieren. Um zu testen, ob diese Kostenunterschiede auch die verschiedenen Fitnessvorteile von Auxotrophien, abhängig von der Kohlenstoffquelle, erklären können wurde eine *Fluss-Bilanz-Analyse* (FBA) auf der Grundlage des metabolischen Netzwerks von *E. coli* durchgeführt. Mit dieser Methode konnten die metabolischen Kosten für die Produktion von allen 20 Aminosäuren unter einer Vielzahl von möglichen Kohlenstoffquellen theoretisch vorhergesagt werden. Diese Analyse zeigte, das die Kosten für die Produktion einer bestimmten Aminosäure tatsächlich stark zwischen zwei Kohelnstoffquellen variieren kann. Des Weiteren wurden die vorhergesagten Kostenunterschiede experimentell nachgewiesen und es konnte so gezeigt werden, dass die Position, an welchem eine Kohlenstoffquelle in das metabolische Netzwerk einfließt, deutlich den Wachstumsvorteil von auxotrophen Stämmen bestimmt. Diese Ergebnisse suggerieren, dass abhängig von der verfügbaren Kohlenstoffquelle unterschiedliche auxotrophe Genotypen evolvieren können.

In dieser Arbeit wurde eine Kombination von verschiedenen Methoden aus der synthetischer Ökologie, experimenteller Evolution und metabolischer Modellierung verwendet um die evolutionären Faktoren zu beleuchten, die zur Reduktion biosynthetischer Selbstständigkeit von Bakterien beitragen. Die Resultate lassen darauf schließen, dass vor allem Adaption eine treibende Kraft ist, die den häufigen Verlust von biosynthetischen Fähigkeiten in Mikroorganismen in der Natur erklärt. Außerdem kann die Reduzierung des metabolischen Funktionsspektrums durch Adaption auch die Evolution von metabolischen Interaktionen und interzellulären biochemischen Netzwerken innerhalb von mikrobiellen Gesellschaften begünstigen.

Introduction

Introduction

Microorganisms (including viruses, bacteria, archaea, yeasts and protists) colonize every habitable *niche* on this planet. A general characteristic of microbial life in nature is that these organisms do not occur in isolation but rather exist within complex communities which are essentially assemblages of multiple species that live together [1-3]. A result of such a communal mode of existence is that microorganisms frequently interact with each other due to activities pertaining to their growth and metabolism [4]. These interactions can range from being negative to positive resulting in the constituent members of a microbial community to engage in diverse relationships that span from competition to cooperation [2,3,5,6]. In competitive interactions, one or both of the interacting partners are negatively affected [3,5]. Examples of such interactions are host-parasite relationships like those between bacteria and their lytic phages; or bacterial genotypes that secrete antibiotic or toxins to inhibit the growth of other genotypes [2,3]. In contrast, cooperative relationships or *mutualisms* result in beneficial effects for both partners [2,3,5,7]. For instance, multiple bacterial species can cooperate to exchange metabolic compounds [4,8] or contribute to produce components required to build biofilms which increase resistance to antibiotics [9,10]. Thus, these interactions help microorganisms address their nutrient requirements, provide protection from environmental factors like anti-microbial compounds or predators and allow adaptation to ecological *niches* in ever-changing environments in nature [4,10,11]. Hence, the outcome of these interactions along the competition-cooperation continuum will be a strong determinant of the evolution of an individual species, and consequently also of the evolution of microbial communities [1,3,6,12]. Therefore, the knowledge of how varied interactions between multiple species evolve in nature is critical to our understanding of the general principles that shape microbial life in nature.

Empirical investigations of natural communities of bacteria have suggested that competitive interactions are much more frequent than cooperative relationships [13-15]. However, a large body of recent work supports a contrarian view that cooperative interactions are also prevalent in the microbial world [4,8,16-19]. A major aspect of cooperative relationships in nature is metabolic exchange between different microbial partners [4,5,8] and such interactions influence diverse processes [2,20] such as agriculture, bioremediation, nutrient cycling, biogeochemical cycles, and poly-microbial infections in higher animals [13,21]. Several metabolites such as amino acids, cofactors, sugars, fatty acids, electrons and many other chemical species can be potentially exchanged between bacteria [4,5,22-28] Therefore, these interactions can have marked effects on the *structure* and *function* of microbial communities [29]. The molecules or chemical components mediating these interactions are encoded in the genome of a bacterial cell and since a large portion of the open reading

frames in microbial genomes is dedicated for such interactions [4], it is imperative that the emergence of microbial metabolic interactions will be significantly influenced by the i) evolution of bacterial genomes and ii) the ensuing effects of genome dynamism on the metabolic network of a bacterial cell.

1.1 The evolution of bacterial genomes

The genetic content of an organism is it's principle information carrier from one generation to the next, and determines the evolutionary potential of life-forms in terms of their metabolic capabilities, lifestyles, and ability to occupy different ecological *niches* [30]. As a consequence, organismal genomes differ with regards to size and coding capabilities, which ultimately determines not only organismal structure and function, but also their evolutionary capacity for change and adaptability [31]. The differences in genomes are evident not only amongst the members of the same kingdom, but also between different kingdoms of life. For instance, genomes of prokaryotes and eukaryotes differ in size and coding content by several orders of magnitude. However, within the prokaryotic domain itself, there are stark differences in genome sizes. For instance at the time of writing this thesis, the smallest known bacterial genome was that of *Candidatus* Trembalaya princeps and *Candidatus* Hodgkinia cicadicola (Figure 1) [32], sized 139 kb and 144 kb respectively; while the largest genome (14.7 mb) was that of *Sorangium cellulosum* [33].

The massive difference in genome sizes signifies the immense state of flux that bacterial genomes are undergoing. These dynamics and variation in gene repertoires in bacteria is attributable to both, gene gain or loss. Genomes can expand via gene duplication or horizontal gene transfer, or in contrast, contract as consequence of a loss of genetic regions [34-36]. Differences in genetic content often result in bacteria having an immense variety of gene repertoires, not only between distinct species, but also within seemingly clonal isolates of the same species [37]. The latter observation is best described in the *pan-genome* concept [36,38], the basis of which is that only a fraction of prokaryotic genes are present in all genomes of a certain species i.e. the *core genome* , whereas the rest are disparately distributed between genomes of the same species i.e. the *dispensable genome* [38,39]. Thus, dynamism across the entire gene loss and gain spectrum is critical for the evolution of bacteria with respect to genes which are essential for bacterial growth and survival.

However, in a recent study, Puigbo and coworkers (2014) established that the rate of loss of *gene families* predominates over gain in bacteria [36]. Based on their results of a comparative genome analyses, the authors of this study observed that prokaryotic genome evolution is characterized by long phases where genomes contract and relatively short time periods of gene gain [36]. Thus, gene loss seems to be a dominant force in the evolution of microbial

genomes and a number of studies support this argument [40-43]. For instance, in a recent study, Bolotin and Hershberg (2015) observed that gene loss is prevalent within seemingly clonal species of pathogenic bacteria like *Mycobacterium leprae* [42]. They posit that gene loss generates the widespread genetic and phenotypic variation which is observed in many clinical isolates of this species [42].

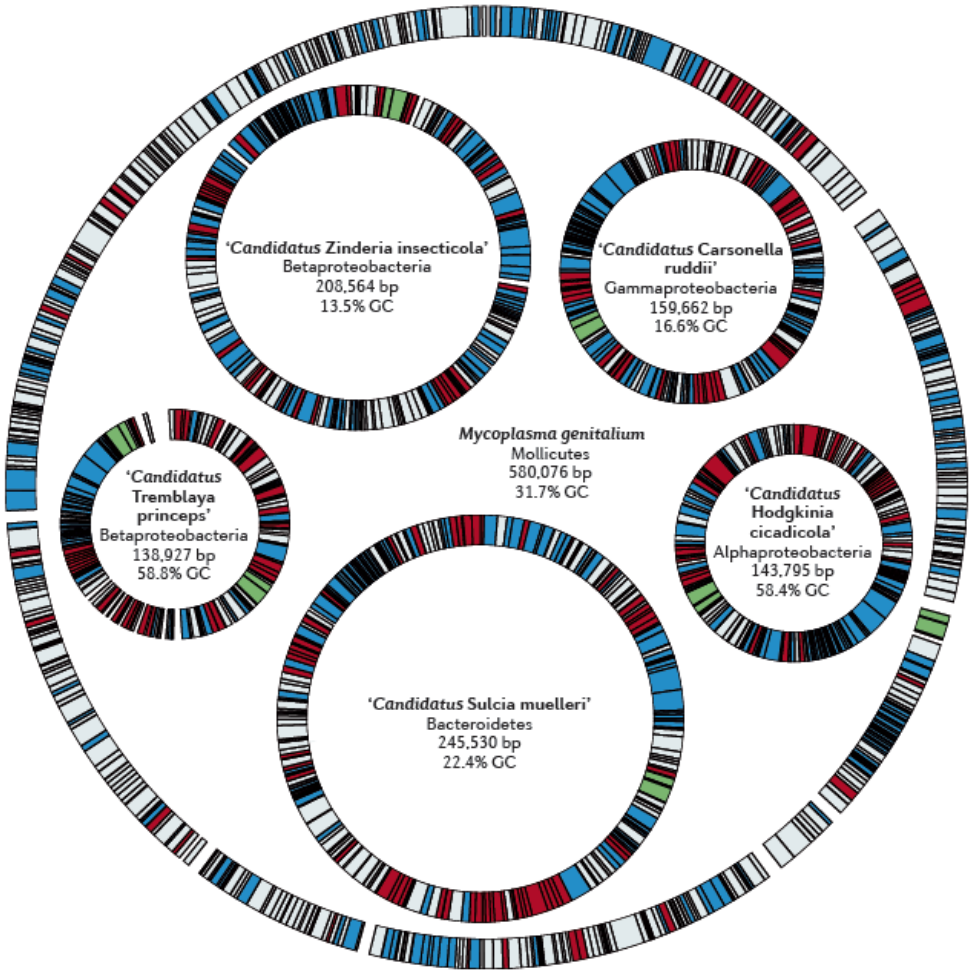
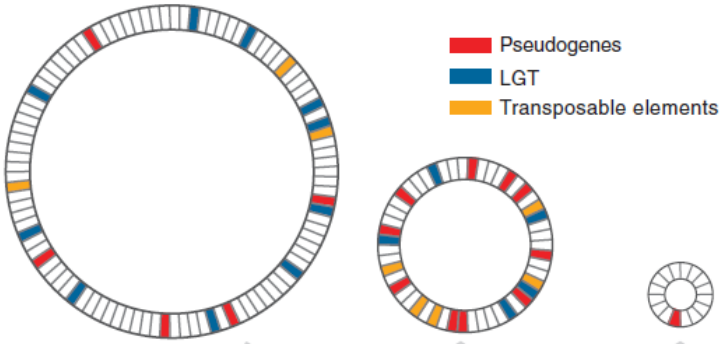


Figure 1: The diversity in sizes of bacterial genomes. Circular maps indicate genes involved in information processing (blue), vitamin and amino acid biosynthesis (maroon), ribosomal RNA (green), and other processes (light grey), while breaks indicate non-coding regions. Figure from McCutcheon and Moran (2012) [32].

1.1.1 Gene loss in bacteria

Loss of genetic content is a common feature of many bacterial genomes, spanning the entire range of microbial lifestyles and environments they dwell in [44]. A characteristic feature of bacterial genomes is that they reduce in size with the level of specialization to a particular environment or association with another organism like an eukaryotic host (Figure 2) [45-47]. Genome reduction is especially prevalent in bacteria that are in symbiotic or parasitic

associations with higher organisms [32,47,48]. For instance bacterial endosymbionts of insects from genera like *Sulcia*, *Buchnera*, *Wigglesworthia*, *Hogdkinia*, or even the above-mentioned *Cd.* Trembalaya have highly contracted genomes; that can be much smaller than the suggested minimal genome size for *autonomous* bacterial growth and survival i.e. 400 kb [49]. Similarly, bacteria with a pathogenic lifestyle such as those belonging to genera like *Rickettsia* or *Mycoplasma* have also lost a large amount of genetic content (Fig. 1 and 2) and possess highly reduced genomes [48,50].



	Free-living	Recent or facultative pathogen	Obligate symbiont or pathogen
Genome size	Large (5-10 MB)	Intermediate (2-5 MB)	Small (0.5-1.5 MB)
Number of pseudogenes	Few	Many	Rare
Incidence of LGT	Frequent	Frequent to rare	Rare to none
Selfish genetic elements	Few	Common	Rare
Genome organization	Stable or unstable	Unstable	Stable
Effective population size	Large	Small	Small

Figure 2: Genome size evolution across different bacterial lifestyles. The extent of genome reduction and the genomic features are highly related to the nature of association with a host, i.e. free-living, facultative, or obligate. Figure from Ochman and Davalos [45].

Loss of genetic content has also been observed in bacteria belonging to the genera *Pelagibacter* [51] and *Prochlorococcus* [52], which are widespread in the Earth's oceans, and saprophytic microorganisms like lactic acid bacteria such as *Lactococcus* or *Lactobacillus* species that are specialized to dairy products [53,54]. Even certain clonal isolates of pathogenic bacteria like *Mycobacterium tuberculosis* or *Yersinia pestis* lose genes, resulting in difference in genetic repertoires within species [42]. Hence, it is evident that gene loss is prevalent in the microbial world. An interesting question thus is: what kinds of genes are lost?

1.1.2 Genes lost as a consequence of reductive genome evolution

Several studies have explored what kinds of genetic content are lost in bacteria in the course of reductive genome evolution. The genomic regions that are lost constitute genes that

encode for a variety of distinct cellular functions in bacteria. Merhej and coworkers (2009) analyzed several bacterial genomes and compared their gene composition depending on the lifestyles of each bacterium and found that bacteria in facultative and obligate associations with eukaryotic hosts had significantly less *orthologous gene content* (COGs) than free living bacteria. 41% of COGs that were lost in the course of adaptation to a host dependent lifestyle mainly encoded for metabolism, while the other COGs which were lost are those that encode for cellular information processes and signaling [55]. By studying genomes of parasitic and mutualistic bacteria, Renesto and coworkers [56], and Shigenobu and coworkers [57] observed that these bacteria lacked a great number of genes for amino acid, nucleotide and cofactor biosynthesis. Bolotin *et al.* [58] analyzed the complete genomes of strains of the lactic acid bacterium *Streptococcus thermophiles* and found that genes related to virulence, energy metabolism, transport, and biosynthesis of amino acids, nucleotides and cofactors were commonly absent or truncated i.e. *pseudogenized*. Similarly van de Guchte *et al.* (2006) [54] analyzed the complete genome of *Lactobacillus bulgaricus* and observed that *pseudogenization* and gene loss was also prominent with respect to the biosynthesis of amino acids, and metabolism.

In summary, a general trend towards reductive genome evolution in bacteria and the frequent loss of metabolic gene repertoires from genomes of bacteria entails that they develop dependencies on the environment they dwell in. A major implication as a result of the loss of conditionally essential genes from the genome of a bacterial cell is the alteration of the *metabolic network*. This is because certain reactions encoded by the lost genes will now be non-functional. What are the changes that occur in metabolic processes in bacteria that lose essential biosynthetic genes?

1.2 The reductive evolution of bacterial metabolism

Metabolic processes in microbial cells serve two main purposes: catabolism and anabolism [59,60]. In catabolism, energetic resources are broken down into readily usable forms of energy that the organism can use whereas in anabolism, cellular biomolecules are synthesized from chemical sources [60,61]. These biomolecules comprise the proteinogenic amino acids, nucleotides, lipids and cofactors. A consequence of the immense variation in gene repertoires is that bacteria can exhibit a wide spectrum of metabolic capabilities. Expectedly, a trend towards a general contraction of genomes and a specialization to nutrient-rich and stable environments will likely also result in a general reduction of the *metabolic breadth* of bacteria. For instance the size of the *metabolic network* of the free-living bacterium *Escherichia coli* consists of 904 genes which are involved in 941 biochemical reactions [62]. In contrast, the *metabolic network* of an obligate endosymbiont *Buchnera asp. APS*, a close relative of *E.coli* [63] and having a highly reduced genome, is comprised of 196 gene products which take part in

263 reactions. This constitutes only 21% of genes and 27% of the reactions in the *E. coli* [64]. Furthermore, Parter and colleagues have shown that the size of the metabolic network tends to decrease depending on the lifestyle of a bacterium i.e. bacteria in obligate associations with eukaryotic hosts have smaller metabolic networks as compared to free-living bacteria (Fig. 3) [65]. As depicted in Fig. 3, metabolic network size in bacteria tends to decrease as microorganisms transition from conditions that would favor the prevalence of free-living bacteria such as those in soil (terrestrial: T) or across diverse host ranges (multiple: M) to conditions that that are have specialized (S) or that would favor facultative (F) or obligate associations (O) with eukaryotic hosts.

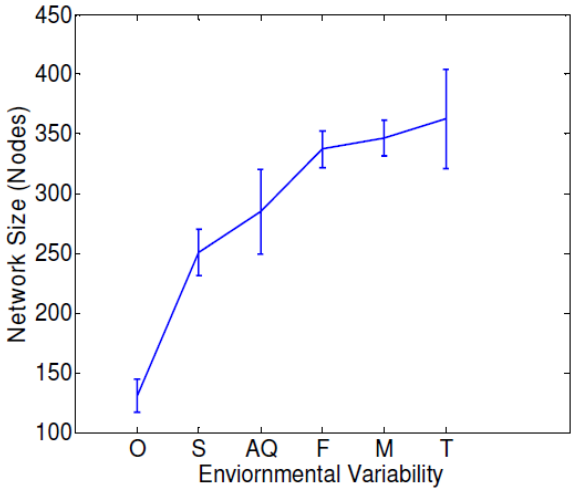


Figure 3: Size of the metabolic network increases with increasing environmental variability. The abbreviations on the X axis denote each environment type: O-Obligate, S-Specialized, AQ-Aquatic, F-Facultative, M-Multiple, T-Terrestrial. Figure from Parter *et al*., (2007) [65]

The widespread reduction in bacterial metabolic capabilities is intriguing for a number of reasons. First, loss of catabolic functions will severely reduce the metabolic breadth of bacteria. These losses however, do not impose any dependencies on a bacterial cell since they probably can utilize other catabolic sources and rather reflect adaptations to the environment. Second, the loss of anabolic abilities, especially the gene repertoire to biosynthesize essential metabolites will hinge organismal growth and survival on an environmental uptake of these nutrients. Endosymbiotic, facultative and specialized bacteria are all devoid of some form of biosynthetic abilities [32,58,66,67]. Since most molecular components of a bacterial cell are either composed of amino acids and nucleotides or require cofactors for their activity (e.g. enzymes), an unrestricted supply of these compounds is the primary requisite for the optimal functioning of bacterial cells.

However, metabolic changes associated with adaptation to certain environments or lifestyles in an organism can bring about changes in relationships between coexisting organisms [68].

For instance, endosymbionts like *Buchnera*, *Sulcia* or *Baumannia* form intricate metabolic complementarities with their eukaryotic hosts and/or with co-occurring bacterial endosymbionts [60,66,69]. In the case of the sharpshooter insect, the coevolution of the endosymbionts *Sulcia muelleri* and *Baumannia cicadellinicola* has resulted in the partitioning of their biosynthetic pathway such that the sum of their metabolic pathways complement the nutritional needs of each other and the host [69]. Similarly, coevolving strains of lactic acid suffice their anabolic needs by *cross-feeding* essential nutrients [8,54]. Pathogens like *Rickettsia* and *Mycoplasma* get these compounds from the eukaryotic hosts they infect [56]. Therefore, a general consequence of losses of metabolism associated reactions can result in previously non-existing associations or dependencies between microorganisms coexisting in a certain environment.

As I stated earlier the losses of anabolic capabilities is especially intriguing. Why must a *prototrophic* bacterial genotype lose autonomous anabolic capabilities and instead become completely dependent on the environment for the uptake of these metabolic compounds for cellular functioning? This question is the central focus of this thesis. On the outset, the loss of biosynthetic functions is likely a general adaptation to the specialized environment. However, the understanding of the evolution of metabolic dependencies in bacteria or their prevalence in nature has rarely been tested. Our knowledge of the loss of biosynthetic functions is limited to numerous comparative genomic studies which show that bacteria lack genes encoding these functions. Such approaches examine genetic segments and changes at the molecular level rather than investigating the fitness consequences of the loss of a function. In the next section, the mechanisms that have been proposed to be causal reasons for the loss of genes and the corresponding functions in bacteria will be discussed.

1.3 Evolutionary mechanisms driving the loss of genes and functions

Two prominent mechanistic explanations exist on the subject of gene loss in bacteria: random genetic drift and selection. A great deal of theoretical and empirical effort has been directed towards understanding the relative contributions of either mechanism in shaping microbial genome evolution and a set of criteria has been proposed under which drift or selection will be strong agents of evolutionary change. Therefore, it is very likely that these criteria will also have significant effects on the loss of biosynthetic functions in bacteria with distinct lifestyles.

1.3.1 Random genetic drift

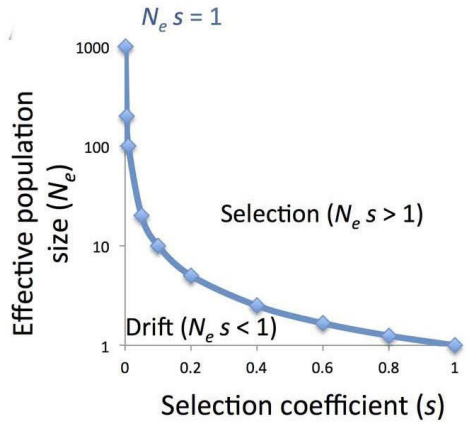
Genetic drift - the stochastic changes of allele frequencies in a finite population- has been suggested as the primary process to explain the extreme genome reduction in endosymbiotic bacteria [32,70,71]. A consequence of their obligate associations with an eukaryotic host is

that the activities of endosymbiotic bacteria are confined to the intracellular environment in which they live [46]. The number of bacterial cells within the host are thus subject to drastic reductions in size i.e. *bottlenecks* at periodic intervals when transmitted vertically to the offspring [71-73]. A reduction in population sizes (N_e; Box 1) can greatly affect the impact of genetic drift [41,73]. When N_e is low, mutations with deleterious effects can persist and even become fixed in the populations [73]. This is because a low N_e reduces the chances of deleterious mutations being eliminated by selection [46]. Furthermore, the intracellular environment of the host significantly restricts the opportunity for recombination to purge deleterious mutations [71,73]. Thus, selection fails to retain even essential genes, resulting in the irreversible accumulation of multiple deleterious mutations or an inactivation of multiple genes, in the same genome, in a process termed as *Muller's Ratchet* [46,71]. This process combined with the fact that mutational outcomes in bacteria are strongly biased towards deletions as opposed to insertions [74,75] indicate that genome reduction in bacterial endosymbionts can occur due to *genetic drift*. Experimental evolution studies with the bacterium *Salmonella typhimurium* do provide support for this hypothesis [76]. When this bacterium was subjected to repeated single cell *bottlenecks* for a long period of time, a general reduction in genome size along with a loss of essential genes was observed [76].

Most studies exploring the role of genetic drift usually employ comparative genomic approaches. For instance, theoretical studies with *Buchnera,* an endosymbiont of aphids, suggest that proteins in this bacterium are less stable due to the accrual of deleterious mutations over time in protein coding genes [71,77]. These stabilities are buffered by an increase in the activity of chaperones. Mutations in these chaperones were favored by natural selection in order to optimize their chaperoning capabilities [46]. Another study of 42 closely related bacteria found that the d_N/d_S *ratios* (Box 1) of bacteria with reduced genomes were consistently higher as compared to bacteria with larger genomes, suggesting that selection tries to counter-act the negative effects arising from the deleterious mutations accruing as a result of drift [74]. However, such studies most often do not account for the presence of certain fitness increasing mutations or changes and rather indicate that deleterious mutations consistently accrue in small genome bacteria. Furthermore, experimental investigations that actually scrutinize the fitness consequences of gene loss or those that validate the link between drift and loss of biosynthetic functions have rarely been conducted. Furthermore, the loss of genes and functions results in bacteria developing metabolic dependencies on the environment. In many instances these dependencies come in the form of intricate associations with eukaryotic hosts [72,78,79] or coevolving bacteria [32,66,69] and result in intricate metabolic complementarities between these organisms. Explaining such tightly integrated patterns of metabolic dependencies is difficult, especially in the context of random

genetic drift. Thus, it is highly likely that selection could have also contributed to the loss of these functions.

Box 1: Effective population size (N_e) and d_N/d_S



The parameter N_e is called the effective size of the population. It is the size of an idealized and theoretical population that has the same value of a certain genetic property (for e.g. the level of genetic drift it undergoes) as that of a complicated population in question [41,73,80]. It is used in classical population genetic models to describe the level of genetic drift that a population undergoes [41,46]. Accordingly, the fixation probability of a mutant allelic variant in populations (Box figure 1: From Martinez-Cano et al., (2015) [46]) depends on the product of N_e and s ($N_e s$), where s is the coefficient of selection [41,46]. When $N_e s > 1$, the fate of the mutant allele is determined by selection whereas when $N_e s < 1$, genetic drift determines the fixation probability of this allele [41,46,73,80]. Therefore, in populations where N_e is low, $N_e s$ will be low and thus the strong effects of drift will render selection inefficient [41,73]. Conversely in populations with a large N_e, $N_e s$ will be large and effects of selection will be much stronger. However, N_e cannot always be used to compare distantly related organisms since most measures of N_e are limited by the fact that it cannot discern between divergent ecotypes of the same species [41]. The solution used in many studies is to thus use the d_N/d_S ratio which is the rate of non-synonymous substitutions over the rate of synonymous substitutions in genes [41]. Thus, the ratio can be calculated using a comparison of sequence data of orthologous genes from various species and then used to gauge the strength of selection [41,81]. The basic idea is that selection of synonymous sites is negligible [41]. Thus, the ratio d_N/d_S is expected to be higher than 1 if natural selection promotes changes in the protein sequence; whereas a ratio lesser than 1 is expected only if natural selection suppresses protein changes [81].

1.3.2 Natural selection

Genome reduction can be thought to be a consequence of adaptive process under three distinct circumstances. The first hypothesis is based on the idea that selection will drive the loss of superfluous genes that impose a fitness cost to a bacterium. In certain environments, particular genetic regions that are no longer required and thus offer no or very less adaptive value can be irreversibly lost from the genome, along with the concomitant loss of encoded functions. This process is termed as '*genome-streamlining*' and could potentially result in the transfer of metabolic burden that was otherwise dedicated for functions which are now under relaxed selection to other important cellular processes [75,82]. The basic crux of this hypothesis is that natural selection favors genome reduction in free-living bacteria as a way to cellular economization [46,82]. This is a plausible theory to explain the genome reduction of free-living bacteria such as *Prochlorococcus,* which live in generally nutrient-poor, but stable environments [37]. This trend is also shared by *Candiatus* Pelagibacter ubique, which inhabits

similar nutrient-limited environments [41]. These prokaryotes have relatively large population sizes in contrast to endosymbionts [46]. Since high N_e renders the effect of drift to be negligible (Box 1) [41,46,83] the effect of selection is very strong causing fitness increasing mutations like deletions to be fixed in the population.

The second hypothesis can be thought to of as an extension of the first one discussed above. It is based on the idea that selection can favor gene loss in predictable or stable environments, where *robustness* with regards to the ability to respond to changing environments is no longer required [46,84]. A source of robustness in bacteria is the presence of multiple genes which are redundant in the function they perform. For instance, *E. coli* has multiple transketolase enzymes which are 99 % identical to each other but function in different metabolic pathways [84]. Thus, if one of these enzymes is lost, another redundant function can fulfill the role of the lost enzyme [84]. Endosymbiotic bacteria inhabit intracellular *niches* that are relatively stable. However, these bacteria have lesser redundancy in functions compared to free-living prokaryotes [84]. For instance, bacteria with reduced genomes have more protein families but each family has very few members and the thus it is hypothesized that the protein family diversity in endosymbiotic or host associated bacteria probably arises as a result of host specific adaptation since the probability of losing a gene is higher if multiple copies or paralogs are present [84]. Thus, selection could have favored the loss of genes which are redundant in cellular functions and have multiple copies in bacteria inhabiting stable and predictable environments [46,84].

The third hypothesis is based on the idea that selection favors the loss of biosynthetic functions in bacteria when the lost function can be supplemented by the environment. The pioneering studies of Zamenhoff and Eichhorn (1967) using amino acid auxotrophic strains of *Bacillus subtilis* was one of the first experimental evidences to suggest that loss of biosynthetic capability when the function was no longer needed and instead provided in the environment, could be adaptive [85]. The authors made this argument since in their experiments, auxotrophic strains of the bacterium where selectively favored over prototrophic strains in the presence of the required amino acid i.e. histidine in the environment [85]. These findings were further supported by another study by Dykhuizen (1978) who also found that auxotrophic mutants of *E. coli* which required tryptophan are selectively favored over prototrophic strains in presence of this amino acid in the environment [86]. Why would auxotrophs be selected over prototrophs in presence of the metabolite? Zamenhof and Eichhorn (1967) posit that when the metabolite is present in the environment bacterial cells can shut down their endogenous machinery to produce the metabolite and thus save the costs associated with production of the metabolite [85]. The biosynthesis of macromolecules in cells entails *metabolic costs* like the energy for chemical transformations to produce the metabolite [87,88] and the costs associated with the

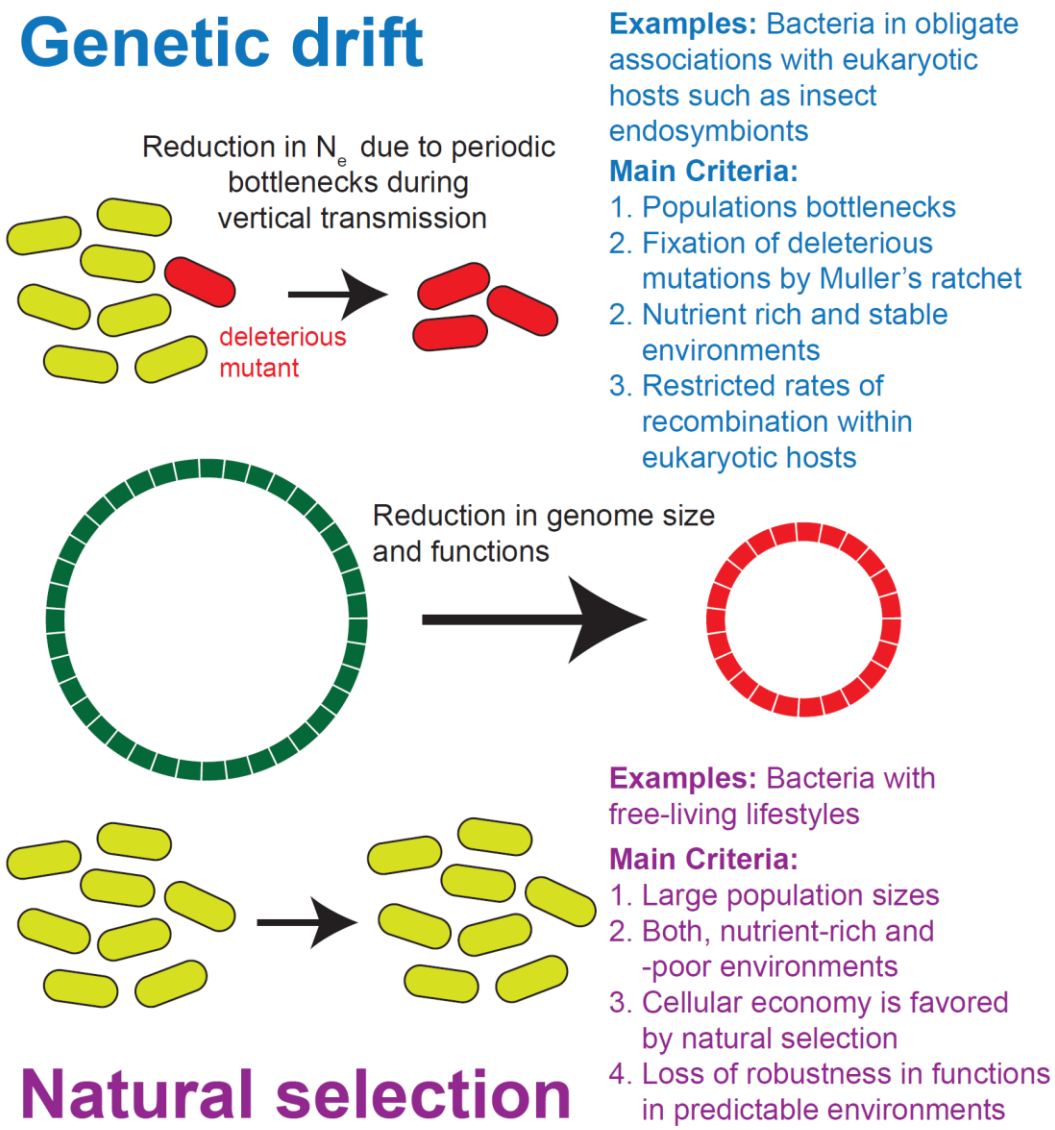


Figure 4: Summary of the main mechanisms proposed to drive the loss of genes and functions in bacteria. The main criteria that determine if genetic drift or natural selection strongly influence the loss of biosynthetic genes in bacteria with distinct lifestyles are listed on the right [41,46,71,82,84,85].

production of the protein machinery to transcribe and translate genetically encoded functions [89,90]. The savings of these costs can result in selective advantages that can drive the loss of biosynthetic functions in microorganisms. Thus, based on this hypothesis, microorganisms that dwell in nutrient-rich environments should readily lose metabolic genes and functions because of the advantages related to cellular economy that are on offer.

The importance of natural selection in driving genome size reduction in bacteria also comes from multiple experimental evolution experiments [91,92]. Lee and Marx (2012) found that non-essential accessory genes were frequently lost from almost 80 % of the evolving populations of the bacterium *Methylobacterium extorquens* AM1 [91]. Gene loss was

accompanied with an increase in fitness suggesting that selection favors the loss of superfluous genes during adaption to specialized environments [91]. In another study, Koskiniemi (2012) tested the fitness consequences [92] of losing genetic segments in the bacterium *Salmonella enterica* and found that fitness increasing deletions were rapidly fixed in the experimental populations.

Thus, a large body of evidence strongly suggests that selection can favor the loss of biosynthetic functions in bacteria. However as evident from the mechanistic explanations, several factors can influence the potential contributions of genetic drift and selection in driving the evolution of dependency.

1.4 Factors determining the loss of biosynthetic genes in bacteria

As with most organisms, the diverse nature of bacterial lifestyles and their ecological *niches* makes their evolutionary progress contingent on a variety of factors such as the selective pressures exerted by the changes in their population dynamics [41,71,93], environment [65,94], their metabolism [19,60,84,95], interactions between different genetic loci [96-99], and interactions with other organisms [26,46,100]. These factors may influence, either in isolation or in combination, gene loss in bacteria by way of natural selection or drift. Thus it is paramount to understand the effects exerted by each factor to cause natural selection or drift to drive the loss of functions in bacteria.

1.4.1 The population biology of bacteria

As discussed earlier N_e strongly influences the efficacy of selection or drift [72,73] and thus should have a decisive effect on genome size (Box 1) [41]. Cao and colleagues (2014) have observed that population sizes of bacteria are also important factors governing the fitness consequences of gene deletions in *E. coli* [101]. In this study, the variation in fitness for 28 deletion mutant strains of *E. coli* ranged from 5 to 174 % between high and low population densities [101]. Natural bacterial populations are variable in terms of their population sizes depending on the environments they inhabit. For instance, the number of *Bacteriodes* cells in the intestinal content of humans generally range from $10^{11} – 10^{12}$ cells g^{-1}, while those of *E. coli* in the same environment are in the range of 10^{7}-10^{8} cells g^{-1} [102]. In contrast, populations of endosymbionts have been suggested to be very small in size [71], although experimental evidence from the European beewolf indicates that the populations of symbiotic bacterium *Candidatus* Streptomyces philanthi consist of around 10^{7} cells [103]. However, insect endosymbionts undergo repeated bouts of bottlenecks. For instance, in the beewolf system, bottlenecks reduce endosymbiont population size by 10^{5} leading to the bacterial load in the new progeny being around 970 cells [103].

Very large population sizes reduce the effects of genetic drift, while increasing the influence of selection to act on natural populations of bacteria [46]. Conversely as mentioned earlier in endosymbionts, small population sizes should lead to drift overriding selection [71]. Interestingly, Wahl and coworkers (2002) mathematically modelled the effect of bottlenecks in the exponentially expanding experimental bacterial populations and observed that fitness gains of adaptive mutations can still be high at low bottleneck sizes, even though the rate of adaptation was low at these bottleneck sizes [104]. Thus, it is plausible that adaptive mutations can also occur at low population sizes. However, experimental studies have rarely tested the effect of population bottlenecks on influencing appearance or outcome of biosynthetic mutations (such as those conferring the loss of biosynthetic genes) and if they can indeed be adaptive.

1.4.2 The metabolic environment and network

Nutrition is likely the most important factor governing the evolution of bacteria as environmental supply of utilizable resources directly determines the fitness of a bacterial species. The diverse ecological interactions including cooperation or competition in microbial communities are a direct result of variation in metabolism [4,105]. Bacterial cells need to optimally tune their growth strategies in a particular environment by using *metabolic resources* in the environment and allocating them to cellular processes like catabolism and anabolism in addition to the general cellular ribosomal machinery [106,107]. These growth strategies will in turn determine if cells can achieve the highest growth rates in the prevailing environment [106]. The growth rate to a large extent will determine the fitness of a bacterial cell and as a consequence mutant strains that are able to respond better to a certain environment by regulating their cellular composition will have high growth rates and outnumber other genotypes [108]. Therefore, it is quite clear that environmental conditions impose strong selective pressures on the evolution of microorganisms. To understand how these selective pressures manifest in cells it is essential to understand the interplay between selection, biochemistry and structure of the *metabolic network* in a bacterial cell [107].

Any change in the constituent reactions, e.g. loss of enzymes, of a *metabolic network* can alter its structure and have marked effects on its evolution [60]. The enzymes in a metabolic pathway are controlled by a series of regulatory mechanisms which precisely control flux in metabolic networks [89]. These regulatory mechanisms serve to optimize costs such that they are minimized and also to allow cells to respond quickly to environmental changes [89]. Since costs incurred in metabolism, especially biosynthesis, play an important role in determining cellular growth [88,90,109], they will have important consequences for the fitness of bacterial cells. Furthermore, selection tends to favor bacteria that make the most energetic gains from a given resource [110]. Thus, the *metabolic network* is likely an important factor governing the

adaptation to an environment especially in the context of a reduction in the biosynthetic capabilities of bacteria.

As mentioned earlier, selection can favor the loss of biosynthetic functions in bacteria in presence of anabolites in the metabolic environment. Thus it can be hypothesized that the presence of ready-to-use anabolites in the external environment can be considered as incentives to stop autonomous production of metabolites whereas the type of catabolic resource determines the exact value of the incentive, which is in turn influenced by the metabolic network of a bacterium.

1.4.3 Metabolic interactions

An important factor governing the adaptive evolution of metabolic dependencies in nature can be the presence of organisms that provide the required metabolite. A majority of interactions within members of microbial communities involve metabolic exchange which is likely a consequence of a general leakage of metabolites from bacterial cells due to the inherent nature of their metabolic activities [4,5,100,111]. A result of this leakage of functions in the environment is that not only can the growth of dependent genotypes be supported by genotypes that produce these metabolites, but also potentiate the loss of functions by providing the required selective impetus.

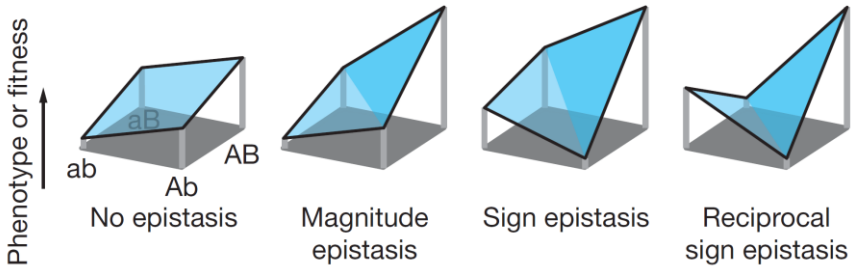
The 'Black Queen Hypothesis' (BQH) proposed by Morris and coworkers (2012) describes such a process of the origin of dependencies through *leaky* benefits termed as *Black Queen (BQ) functions* [51]. Such *BQ* functions are both essential for the growth of the constituents of a bacterial community and at the same time is freely or partially available to all members of the community. The basic premise of the BQH is built on two conditions. First, these *BQ functions* are publicly available. Second, bacteria that use this freely available *BQ function* can benefit from losing genes associated with their own autonomous production of this *BQ function*. These mutants, which are also termed as 'beneficiaries' can stably coexist with strains that retain the function i.e. 'helpers' in the community. Thus the BQH describes an adaptive mechanism under which bacteria may lose essential functions [26,100]. The BQH was first used to describe the loss of functions and origin of dependency relating to hydrogen peroxide (HOOH) detoxification in a natural community of marine bacteria, where *Prochlorococcu*s species (beneficiaries) depend on other helper bacteria [112]. These bacteria lack the genes for catalase-peroxidase, an enzyme that breaks down HOOH. However, the 'helper' bacteria that are often found in the natural community have retained this function, which is also leaky. Thus, as a result of detoxifying HOOH for themselves, the 'helpers' also reduce HOOH in the extracellular environment allowing the growth of the 'beneficiaries' i.e *Prochlorococcu*s [51]. However, the BQH can be easily applicable to other *leaky* functions such as metabolic by-products, enzymes like invertase or iron chelating molecules [100,113].

1.4.4 Genetic interactions: Epistasis

Genome reduced bacteria often lose large stretches genetic content [32,72,76,92] or multiple coding genes [42,46,58,114] and frequently multiple and diverse biosynthetic genes are lost [54,58,66,69]. Moreover, bacterial genomes harboring deletions in biosynthetic genes may also contain mutations in non-biosynthetic genes. A result of harboring multiple gene mutations in the same genotype can be such that the phenotypic impact, e.g. fitness effect, of one mutation is dependent on the effect of the other mutation [115]. This dependence is termed as *epistasis* (Box 3) and is evolutionarily important since *epistatic* effects can alter the course of evolution for adaptively evolving populations (Box 3) [97,99,115-117].

Box 2: Epistasis and its effects on fitness

Epistasis implies that the phenotypic effect of one mutation depends on the presence of other mutations in genetic background in which it occurs [97,115,118]. Epistasis can have multiple consequences on phenotypic fitness (Box Figure 2: From Poelwijk *et al*., (2007) [118]). Consider two alleles 'a' and 'b' which theoretically can mutate to allelic forms 'A' and 'B', respectively.



In the case when there is no epistasis, mutation from 'a' to 'A' yields the same fitness effect for in different genotypes ('b' or 'B'), while for magnitude epistasis the fitness effect can differ in magnitude, but not in sign. The magnitude can be more than expected from the product of the effect of individual mutations in isolation i.e. positive epistasis or less than that is expected from the product of the individual effects i.e. negative epistasis. In the case of sign epistasis, the sign of the fitness effect changes. A change in sign of the fitness effect can also occur for both mutations, which is termed as reciprocal sign epistasis [118-120]. Mutations exhibiting magnitude epistasis or no epistasis are always favored (or disfavored), regardless of the genetic background in which they appear [118]. In contrast, mutations exhibiting sign epistasis may be rejected by natural selection, even if they are eventually required to increase fitness [118].

Several studies have shown that *epistasis* is prevalent in the *metabolic networks* of many organisms [115,121,122]. For instance, a general trend observed in epistatic patterns of genes involved in metabolism in *Saccharomyces cerevisiae* and *E. coli* is that the type of *epistatic* interaction strongly depends on the essentiality and redundancy of the interacting genes [122]. Non-essential genes with overlapping functions often display *negative epistasis* whereas, essential genes without overlapping functions frequently show *positive epistasis* [122]. This finding is especially interesting for the study of the loss of biosynthetic genes in bacteria which can be thought of as being *conditionally essential* and thus may display *positive* or *negative*

epistasis, both of which have been shown to be favored by natural selection [118]. Thus, gaining knowledge of the magnitude of effects that is exerted by *epistasis* on reductive genome evolution and loss of biosynthetic functions in bacteria is necessary. However, experimental studies to investigate these effects have rarely been undertaken.

1.5 Approaches to determine evolutionary mechanisms of reductive evolution of biosynthetic functions

The ideal way to determine the reasons for tremendous diversity in genomes and metabolism of microbial communities is to observe their evolutionary progression in terms of various parameters, for e.g. population dynamics, growth and fitness. However, studies involving natural populations of bacteria will suffer from certain consistent drawbacks. Firstly, the natural populations and their genomes in their current state are simply snapshots of multiple, ongoing evolutionary processes and will offer limited information with regards to what drove the genomes to their current state. Secondly, since these populations are only a snapshot, the lack of a prototrophic or metabolically autonomous ancestor for each species will make a comparative analysis difficult. Thirdly, most of the bacteria having highly reduced genomes, because of their close associations with their environments and complex nutritional requirements, are limited by their cultivability on standard laboratory growth media or even tractable to genetic approaches. In order to circumvent these issues many alternate approaches employing *in silico* analyses, and/or studies with model organisms combined with laboratory approaches can be employed to address fundamental questions in microbial evolutionary ecology.

1.5.1 Synthetic ecology

Synthetic biology has been traditionally employed to better understand biological systems by constructing their simplified or synthetic forms, thus eliminating the complexities associated with nature, to determine functioning of biological systems [123,124]. An important component of synthetic ecology is that it employs a bottom-up approach to address the limitations that are associated with studying natural communities in order to design synthetic microbial consortia that could serve as simplified models of their natural counterparts, while affording enhanced tractability and controllability [125]. Thus, by experimentally measuring observable properties of a synthetically engineered system, such as interactions between genotypes in a consortium and stability of the consortia, fundamental principles underlying the basis of evolution in microbial communities can be understood [125].

For instance, synthetic ecology can be used to implement defined genetic or environmental perturbations, which can potentially result in new inter-species interactions through metabolite exchange or alter the stability of a community [125]. Synthetic ecology has been

recently employed by two distinct studies to understand the evolutionary ecology of cooperation in microorganisms. Shou and coworkers (2007) engineered a complementary metabolic interaction based on the exchange of lysine and adenine between two yeast strains [126]. Similarly Pande and colleagues (2013) introduced cooperative interactions between different *E.coli* strains based on the complementary exchange of amino acids. Using such purposefully designed interactions within microbial consortia, the evolutionary stability and dynamics between the interacting types in these communities were determined [24].

Importantly, employing such an approach allows the experimental measurement of fitness consequence of the genetic or environmental or genetic perturbation since the evolutionary success of an organism is based on its fitness in a given environment. Thus, if a trait is adaptive it should have positive fitness consequences for genotypes bearing this trait. For instance, Khan and colleagues measured the fitness of genotypes that differed in the presence of five beneficial mutations that first appeared in an *E. coli* population evolving to minimal medium [117]. By reconstructing these mutations into ancestral genetic backgrounds in different combinations they could disentangle the specific fitness effects of each mutation when alone or present with other mutations [117]. Thus, by adopting such a bottom-up approach and a rational experimental design, fitness and growth patterns of genotypes of a model organism lacking one or a certain set of biosynthetic genes can be compared to the fitness and growth of a completely *prototrophic* genotype in a consortium. Thus, the fitness consequence of being *auxotrophic* or prototrophic can be experimentally quantified, knowledge of which is essential to understand if the loss of function can be favored by natural selection. Furthermore, by manipulating the environmental conditions to modulate both, the supplementation of the required anabolite to auxotrophs and the nature of catabolic resource, fitness consequences at a range of concentrations of the anabolite and effects relating to *environmental plasticity* can be quantitatively elucidated.

1.5.2 Experimental evolution

Experimental evolution is the study of evolutionary processes that occur in experimental populations in response to a set of conditions that are imposed by the experimenter [127]. The basic essence of such approaches is to propagate replicate population of an organism to a particular environment for a defined number of *generations*, to gain an understanding of how organisms adapt to the environment. Inferences are then made by comparing the evolved populations to their ancestral relatives with regards to multiple distinct *life history* parameters like fitness [128]. Thus, experiments can be tailored to unravel the influence of drift, mutations, gene flow, and selection on gene variation and on heritable phenotypic traits [129]. The use of long-term evolutionary studies also make it possible to directly investigate an important question in evolutionary biology: what is the genetic basis of

adaptation? [130] Thus by attempting to link phenotypes, genotypes and fitness to each other allows unravelling of causes and consequences underlying adaptation in experimental populations.

Employing such experimental approaches using model bacterial species along-with experimental evolution offers considerable advantages over other approaches. First, due to their relatively short generation times, long term experiments over large evolutionary time scales can be performed. For instance, the most famous long term experimental evolution has been running for ~60,000 generations of *E. coli* growth, spanning the last 27 years (as of June 2015) [131]. Second, representative population samples from several time points over the time course of the experiment can be preserved as frozen stocks. This allows the analysis of the 'fossil record' of several populations of the evolving populations for insights into fitness and genomics of adaptation over different evolutionary time points. Third, replicate populations are initiated from the same isogenic ancestor, and thus undergo the same selection regimes allowing the experimenter to measure the effect of chance in adaptive processes. Fourth, the use of microbial populations allows for easy imposition or changes of different selection pressures, such as the type of metabolic environment. Last, the use of model organisms with a completely sequenced and well-annotated genome allows the use of whole genome sequencing approaches in the derived populations to identify the causal molecular mechanisms that underlie adaptive changes in the evolving populations or the basis of a phenotype [129]. Thus, an experimental evolution approach can offer significant insights into how selective pressures shape microbial populations in their natural environments.

Interestingly adaptive trait loss has been observed in experimental evolution experiments. Cooper and Lenski (2000) found that adaptive evolution in a particular metabolic environment caused the decay of unused catabolic functions, which are otherwise required in alternate environments [132]. Trait loss has also been observed in the case of *Bacillus subtilis* populations which were evolved in rich medium repeatedly leading to evolution of cells which had lost the ability to sporulate, a function which was no longer required in this environment [133]. Thus, experimental evolution allows for a careful design to test the effect of both adaptive and non-adaptive processes on the concomitant evolution of genomes and biosynthetic functions along with the ability to manipulate selection regimes.

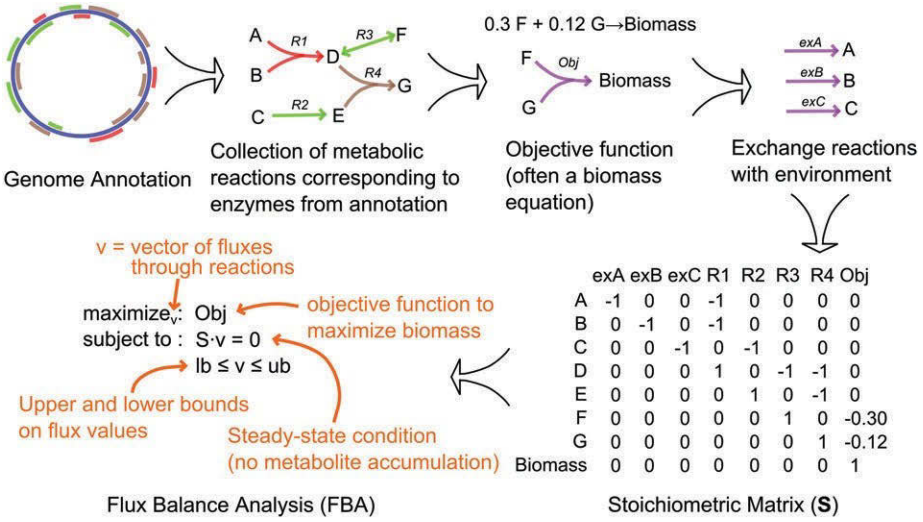
1.5.3 *in silico* approaches

An increasingly large pool of genomic and metagenome datasets has provided a rich resource to test the generality of initial observations linking bacterial lifestyle to genome features [21,134]. Using this genomic information, the metabolic potential of different species which have a completely sequenced genome can be reconstructed by analyzing the information

provided in the annotated genome combined with information from curated databases and literature [135]. Such *genome-scale metabolic network reconstructions* (Box 3) have been successfully applied to understand the metabolic capabilities and engineer several bacteria [21,134,136,137] as well as to predict metabolic interactions in the context of microbial communities [138,139]. Furthermore, flux balance analysis is a critical component to understand the quantitative nature of metabolic fluxes through such *metabolic networks*.

Box 3: Genome-scale metabolic network reconstructions

Genome-scale metabolic network reconstructions involve building a draft reconstruction of a metabolic network of an organism from the information derived from analysis of its genome annotation using databases to find homologous proteins with known enzymatic activity. Subsequently, a metabolic objective is defined, which for microbes is often assumed to be a biomass equation and it is generally assumed that bacterial cells are optimized to have high growth rates. Exchange reactions are then defined which will allow metabolites to enter the reconstructed metabolic network. Next, the reactions constituting this network are then compiled into a stoichiometric matrix (S). Flux balance analysis is then used to determine flux distributions through the network that optimize the metabolic objective subject to *steady-state constraints* and *fluxes*. Figure from Biggs *et al*., (2015) [21,134]



	exA	exB	exC	R1	R2	R3	R4	Obj
A	-1	0	0	-1	0	0	0	0
B	0	-1	0	-1	0	0	0	0
C	0	0	-1	0	-1	0	0	0
D	0	0	0	1	0	-1	-1	0
E	0	0	0	0	1	0	-1	0
F	0	0	0	0	0	1	0	-0.30
G	0	0	0	0	0	0	1	-0.12
Biomass	0	0	0	0	0	0	0	1

FBA calculates the flow of metabolites through a reconstructed metabolic network (Box 3). Thus it is possible to predict the growth rate of an organism or the rate of production of a certain metabolite [140]. The analysis is performed under *steady-state* conditions and it only requires information about the stoichiometry of reactions involved in metabolic pathways and on the metabolic demands [141]. An important assumption is that the cell performs optimally with respect to a metabolic function, such as maximization of biomass production or minimization of nutrient utilization, on the premise that selection pressures during evolution guide the systems towards optimality [142]. FBA has been used to analyze and predict the metabolic genotype–phenotype relation, the effect of gene deletions, and also

analysis of genome scale models of metabolism in various organisms [141]. Furthermore, in conjunction with experimental kinetic data, it is possible to verify the predictions of FBA models and thus build a framework to elucidate links between growth yield, network robustness, and gene essentiality [143,144]. Thus, the use of such *in silico* approaches could be extremely beneficial in determining the metabolic adaptations that underlie bacterial growth in a certain metabolic environment.

1.6 Aims and outline of the thesis

The ubiquity and importance of bacterial communities in nature is well known. In contrast, very little information is known with regards to the importance of the interactions between constituent members in the formation of these assemblages [100]. The loss of metabolic functions from microbial genomes can give rise to previously absent dependencies between species and thus forge metabolic interactions between members of microbial communities. Also, the seemingly widespread trend towards a reduction in genetic content and metabolic functions further highlight the possible importance of the loss of functions in governing the evolution of microbial communities. However, the evolutionary mechanisms that drive these losses are unclear. Is the loss of biosynthetic functions in bacteria and the ensuing metabolic dependencies an evolutionary adaptation to the environment? Answering this question is of paramount importance to understand the evolution of metabolic interactions, and communities. In order to address the above question, it is essential to it break it down into two distinct parts to better understand the evolution of dependencies in bacteria.

The first part aims to understand if selection can explain the loss of biosynthetic functions and attempts to answer these questions: (i) Is the loss of biosynthetic genes and functions common among diverse groups of bacteria? (ii) When bacteria lose biosynthetic functions, what is the effect on the fitness on such bacterial strains? iii) How do genetic interactions between biosynthetic mutations i.e. *epistasis* affect fitness of the strains that have lost multiple biosynthetic functions? (iv) Can distinct metabolic environments influence the fitness consequences of the loss of biosynthetic functions? And (v) how do these metabolic environments and the underlying metabolic network influence the metabolic costs incurred by a bacterial cell? The answers to this question were explored using a combination of computational and experimental approaches. First, an *in silico* estimation of the prevalence of biosynthetic gene loss in nature was addressed (chapter 1). Second, the consequences of biosynthetic gene loss was determined using mutant strains of two model bacterial genotypes *E. coli* and *Acinetobacter baylyi* that lacked biosynthetic genes in a variety of nutritional environments (chapter 1). Third, multiple biosynthetic mutations were introduced into the same genome to unravel the effect of epistasis (chapter 2). Finally, a computational analysis of metabolic costs in different metabolic environments and a comparison with growth strategies of bacteria was attempted (chapter 3).

In the second part, the issue of natural selection was directly tackled and it was asked whether natural selection favors the loss of metabolic autonomy in bacteria in nutrient-rich environments, thus making them dependent on the environment. Isogenic populations of *E. coli* were evolved for 2000 generations in nutrient-rich or poor environments to determine the propensity of bacteria to lose biosynthetic genes and functions in either environment (chapter 4). The evolved genotypes were characterized in terms of their fitness, the presence

or absence of biosynthetic functions, and the genomic basis for the observed evolutionary changes (chapter 4).

In the general discussion section, I discuss the major conclusions from the findings of this thesis and their implications on the adaptive evolution of bacterial interactions, and communities.

Chapter 1

Chapter 1

Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria

Published in: Evolution (2015); Volume 68; Pages: 2559-2570; DOI: 10.1111/evo.12468

Only minor typographical modifications have been made to the published article in this chapter.

Authors: Glen D'Souza, Silvio Waschina, Samay Pande, Katrin Bohl, Christoph Kaleta, and Christian Kost

Bacteria that have adapted to nutrient-rich and stable environments are typically characterized by reduced genomes. The loss of biosynthetic genes frequently renders these lineages auxotroph, hinging their survival on an environmental uptake of certain metabolites. The evolutionary forces that drive this genome degradation, however, remain elusive. Our analysis of 949 metabolic networks revealed auxotrophies are likely highly prevalent in both symbiotic and free-living bacteria. To unravel whether selective advantages can account for the rampant loss of anabolic genes, we systematically determined the fitness consequences that result from deleting conditionally essential biosynthetic genes from the genomes of *Escherichia coli* and *Acinetobacter baylyi* in the presence of the focal nutrient. Pairwise competition experiments with each of 20 mutants auxotrophic for different amino acids, vitamins, and nucleobases against the prototrophic wild type unveiled a pronounced, concentration-dependent growth advantage of around 13% for virtually all mutants tested. Individually deleting different genes from the same biosynthesis pathway entailed gene-specific fitness consequences and loss of the same biosynthetic genes from the genomes of *E. coli* and *A. baylyi* differentially affected the fitness of the resulting auxotrophic mutants. Taken together, our findings suggest adaptive benefits could drive the loss of conditionally essential biosynthetic genes.

1. Introduction

Although it has been known for a long time that factors such as deletions, duplications, and horizontal gene transfer can drastically shape the size and information content of bacterial genomes, one of the most surprising insights that resulted from sequencing multiple isolates of the same, seemingly identical species was the enormous plasticity that characterized all genomes analysed so far [145-147]. For example, a comparison of sixty-one publically available *Escherichia coli* and *Shigella* spp. genome sequences revealed that only 6% of the predicted gene families were represented in every genome (i.e. the '*core genome*'), while all others were present only in a subset of strains (i.e. the '*accessory*' or '*pan-genome*') [148]. Interestingly, even the gene repertoire that constituted the core genome lacked genes that were otherwise deemed essential for the growth of *E. coli* [149-151]. These observations raise the question what major forces drive the loss of genes that essentially contribute to cellular fitness.

Genome reduction is a typifying feature of bacteria that occur in nutrient-rich or constant environments such as lactic acid bacteria [54], endosymbionts [69], or pathogens [47], respectively. Under these conditions, coding regions that provide little or no adaptive value in a given environment may be lost [47,133]. This so-called '*genome streamlining*' is thought to reduce the metabolic burden for basic cellular processes and could thus provide the resulting genotype with selective advantages over other genotypes that still bear these costs [52]. Adaptive benefits as a consequence of losing essential biosynthetic functions may arise when the corresponding metabolite is sufficiently present in the bacterial growth environment or provided by a co-occurring organism [51]. The latter scenario likely explains why amino acid biosynthesis pathways are sometimes partitioned between a eukaryotic host and its prokaryotic endosymbiont [63] or between multiple co-symbionts [69].

[48,50,74]. When bacteria transition from a free-living to a symbiotic lifestyle such as the bacterial endosymbionts of insects [47,69,152], repeated bottlenecks of relatively small populations may result in a weakened selection even for required genes, thus resulting in an elimination of dispensable genes [79]. Indeed, experimentally evolving *Salmonella enterica* by subjecting it to regular population bottlenecks resulted in a reduction of genome size and a concomitant loss of essential genes [76]. Similar mechanisms might act on obligate bacterial endosymbionts, thus explaining their typically extremely reduced genomes that retain few essential biosynthetic genes. However, it is generally difficult to infer from genomic analyses whether drift or selection was the main force to explain genome degradation. Hence, alternative approaches are necessary to determine the drivers of bacterial gene loss.

In vitro approaches are ideal for this purpose, because experiments can be purposefully designed and environmental conditions rigorously controlled. Long-term evolution experiments, in which different bacterial strains were serially propagated and thus allowed to

adapt to the respective environments have shown that large genomic deletions are indeed prevalent under these conditions [91,92,153]. Moreover, fitness advantages accompanied some of these deletions, suggesting selection rather than drift drove the loss of these genes [91,92,117]. Interestingly, in a study with the bacterium *Methylobacterium extorquens* AM1 [91], the observed fitness advantage did not seem to result from a general shortening of the genome, but was rather due to the loss of specific genes. However, determining whether the deletion of a metabolic gene has a negative, neutral or beneficial effect on the fitness of the resulting mutant is a non-trivial task. Problems that arise when naturally evolved mutants are being analysed are first that very often multiple genes are lost simultaneously [92], thus making it difficult to link an observed change in fitness to the loss of a particular gene. Second, multiple auxotrophies often hamper the culturability and hence experimental amenability of a given strain (e.g. bacterial endosymbionts). Third, genetic interactions among different mutations that arose independently impede the determination of fitness consequences of a single mutation. These problems can be circumvented by analysing genetically well-characterized natural or engineered mutants. Indeed, *Bacillus subtilis* [85]and *Escherichia coli* [86] mutants impaired in tryptophan biosynthesis revealed significant fitness advantages in the presence of the amino acid relative to prototrophic cells. However, whether the loss of essential metabolic genes always results in selective advantages when the required metabolite is present in the environment as well as which causal mechanisms explain this observation remain obscure.

Here we combine *in silico* analyses with systematic laboratory experiments of genetically engineered mutants to address the following questions: i) How widespread is the loss of conditionally essential metabolic genes amongst bacteria in nature? ii) What fitness consequences result from the loss of a metabolic gene? iii) Does the fitness effect depend on a) the gene analysed, b) the concentration of the corresponding metabolite in the growth environment, or c) the position of the catalytic enzyme within a metabolic pathway?, and iv) Do different species differ in their fitness consequences upon gene loss?

2. Materials and methods

2.1 Competitive fitness assays

To predict putative auxotrophies in different bacterial species, the metabolic networks of 949 bacteria were examined for the presence of metabolic routes leading to the formation of amino acids, vitamins, or nucleobases. As a first step, all biosynthetic pathways known in bacteria to be involved in the formation of each of 20 amino acids, 3 vitamins, and 2 nucleobases (Table S1) were collected from the manually curated metabolic pathway database MetaCyc [154]. The pathways were consolidated (Fig. S3) to identify alternative biosynthetic routes and pathway dependencies (e.g. a pathway that provides the precursor

metabolite). In a second step, the existence of the individual pathway reactions in 949 bacterial species was inferred using the MicroScope genome annotation and analysis platform [155]. Briefly, the MicroScope platform is a collection of microbial metabolic networks, which consist of a subset of those reactions from the MetaCyc database, for which a genome segment (including plasmids) was identified or predicted, that is part of a gene for an enzyme that can catalyse the corresponding reaction. In a third step, an organism was predicted to be auxotrophic for a given metabolite if all possible metabolite-forming routes (Fig. S3) lacked more than 50% of the pathway's reactions as indicated by the absence of the corresponding annotated genes from the organism's genome sequence. This 50% cut-off was chosen, to increase robustness of the predictions against sequencing errors (i.e. missing annotations) and errors during the process of the metabolic network reconstruction. Genes annotated as pseudogenes were excluded from the analysis, since pseudogenes are often a transitional stage of the gene from a functional gene towards complete gene loss [156]. Therefore, all reactions that depended on pseudogenes were classified as 'not present'. The observed results thus represent a conservative estimate of the frequency of auxotrophies in bacteria with a sequenced genome.

All bacterial strains were categorised as '*free-living*', '*gut-inhabiting*', or '*endosymbiotic*' based on the genome meta-information stored in the Genomes Online Database [157].

The total mass of each protein in Mega Dalton (i.e. mass of the individual protein multiplied with the abundance of protein copies per cell) that was involved in the biosynthesis of Arg, His, and Trp was obtained from [89].

2.2 Culture conditions

All cultures were incubated at 30 °C under shaking conditions and experiments were performed in MMAB minimal medium [158] without biotin and using fructose (5 g l^{-1}) instead of malate as carbon source. Growth kinetic studies were performed in 96-microwell plates (Nunc, Denmark) with a culture volume of 0.2 ml. Competitive fitness experiments were performed in 96-deepwell plates (Eppendorf, Germany) with a culture volume of 1 ml.

2.3 Construction of strains

Single gene deletions in *E. coli* that would lead to auxotrophy for a single amino acid, nucleobase, or vitamin were identified using the KEGG pathway- [159] and the Ecocyc database [160]. All deletions were transferred from existing strains [149] using P1 phage-mediated transduction [161] into *E. coli* BW 25113 [149]. To distinguish different strains in competition experiments, the arabinose utilization locus (Ara^+) of *E. coli* strain REL 607 [162] was introduced into all auxotrophs by P1 transduction. Potential genetic targets to

construct auxotrophs for Arg, His, Leu, and Trp in *A. baylyi* were identified using the KEGG pathway database [159] and deletion mutants were constructed as described [163] (see Supporting Methods for details). Conditional lethality of these mutations in MMAB medium was verified in previous studies [149,150,164] as well as by inoculating 10^5 colony-forming units (CFUs) ml^{-1} of these strains into 1 ml MMAB medium. After 24 h, their optical density (OD) was determined spectrophotometrically at 600 nm using a Tecan Infinite F200 Pro platereader (Tecan Group Ltd., Switzerland) and the mutation was deemed conditionally essential when the auxotroph's growth did not exceed the OD_{600nm} of uninoculated minimal medium. In contrast, when the mutant was able to grow (i.e. exceed the OD_{600nm} of uninoculated minimal medium) the strain was excluded from further analysis and the next gene upstream the biosynthetic pathway was deleted until a mutant was found that satisfied the criterion of conditional essentiality. Gene deletions were in all cases confirmed by sequencing the corresponding genomic regions.

2.4 Growth kinetic and fitness assays

For all experiments, auxotrophs were precultured at 30 °C in MMAB medium supplemented with 200 μM of the required nutrient. Growth kinetics of auxotrophic strains and a matching prototrophic WT were recorded in MMAB medium supplemented with the focal nutrient at the respective concentration. The pH of the medium did not change significantly over the course of the experiments. The medium was inoculated with ~10^5 CFUs ml^{-1} of an over-night culture (i.e. 16 h). Growth kinetic experiments were performed in a Tecan Infinite Pro 200 plate reader (Tecan Group Ltd., Switzerland). Growth was measured as absorbance at 600 nm (i.e. OD) every 8 minutes for 24 hours with 3 min of shaking between measurements. The maximum population density (i.e. OD) reached was calculated using the Magellan 7.1 software (Tecan Group Ltd., Switzerland). The relative maximum OD was calculated by dividing the OD of the auxotroph by the OD of the WT grown at the same metabolite concentration. Monoculture experiments of every auxotroph and its cognate WT control were replicated four times for each metabolite concentration tested.

For competitive fitness assays, ~10^5 CFUs ml^{-1} of either WT or auxotrophs were inoculated into 1 ml MMAB medium with the requisite nutrient concentration and cell numbers were determined at 0 h and 24 h by plating. *E. coli* auxotrophs were differentiated from WT using the arabinose utilisation marker (Ara^+/ Ara^-) as described [162] and *A. baylyi*, strains were differentiated using an antibiotic marker (kanamycin). The ara marker was swapped between competitors. None of the two markers used incurred detectable fitness costs (paired samples t-test: $P>0.05$, $n=8$). Competitive fitness of auxotrophs versus WT was determined by calculating the Malthusian parameter (M) of both genotypes: $M = (\ln (N_f/ N_i)/ 24)$, where N_i is initial number of CFUs at 0 h and N_f is the final CFU count after 24 h [162]. Relative

fitness was calculated as the ratio of Malthusian parameters. Coculture experiments were replicated eight times (i.e. comparison of a) WT and deletion mutants within the same biosynthetic pathway, and b) WT and auxo-trophic *A. baylyi* mutants) or four times for each metabolite concentration tested (all others).

The two methods used to quantify bacterial productivity were quantitatively comparable, as indicated by a significant correlation between CFU plate counts and OD readings (Spearman rank correlation: $\varrho = 0.76$, $P = 4.4 * 10^{-26}$, n= 128).

2.5 Statistical analysis

Frequency distributions of auxotrophic bacteria with different lifestyles were compared with a Pearson's Chi-squared test with Yates' continuity correction and the distributions of the number of auxotrophies per organism with the Wilcoxon rank sum test with continuity correction. The Levene's test was used to assess homogeneity of variances and variances were assumed to be inhomogeneous when $P>0.05$.

Statistical differences in the growth parameters (i.e. OD, relative fitness) of WT and auxotrophs were determined by independent sample t-tests (monoculture growth experiments) or paired sample t-tests (co-culture competition experiments). Brown-Forsythe tests followed by either Tamhane's T2 (non-homogenous variances) or LSD (homogenous variances) post-hoc tests were used to infer statistical differences in the relative fitness of mutants lacking different genes of the same biosynthetic pathway. The False Discovery Rate (FDR) procedure of Benjamini et al. (2006) was applied to correct P values after multiple testing [165]. The concentrations required by auxotrophs to exceed WT fitness were compared with Mann-Whitney U-tests. Two-sample t-tests were used to detect fitness costs of genetic markers. The relationship between monoculture OD and plate counts as well as between protein mass invested and the relative position within the three biosynthetic pathways was investigated by applying Spearman's rank correlations. All statistical analyses were performed using the R software (R Development Core Team 2013, version 2.15.3) [166] and the SPSS package (version 17.0, IBM, USA).

3. Results

3.1 Loss of conditionally essential biosynthetic functions is common in bacteria

To determine how common the loss of conditionally essential biosynthetic functions is among natural bacterial isolates, we investigated the frequency with which auxotrophies occurred in each of 949 sequenced eubacterial genomes. The set of genomes analysed covered a phylogenetically diverse spectrum of bacterial phyla (Fig. S1), yet was biased in its composition towards bacteria of biotechnological or medical relevance. Taking advantage of

genome sequences, pathway information, and genome annotation, we focussed our analysis on all 20 proteinogenic amino acids, two nucleosides, as well as three vitamins (Table S1). A majority of Eubacteria (i.e. 76%) were predicted to be auxotrophic for between one and 25 different metabolites that are needed for growth and metabolism (Fig. 1A). The most commonly predicted compounds that could not be synthesized by the organisms analysed were biotin (36%), phenylalanine (36%), and asparagine (37%) (Fig. 1C). In contrast, very few bacteria (i.e. 7%) were auxotrophic for proline and isoleucine. Notably, three fourth of all strains predicted to be auxotrophic had lost more than 85% of the genes involved in the biosynthesis of tryptophan, histidine, leucine, pyrimidines, and purines (Fig. S4), which are the longest linear pathways analysed (Fig. S3). This finding suggests that auxotrophic strains tend to lose entire pathways once a biosynthetic function has been lost.

When putative auxotrophy frequencies were determined for the phylum Enterobacteriaceae (i.e. 116 organisms), 13% of all strains in this subset were predicted to be auxotrophic (Fig. 1A). Here, the most commonly found auxotrophy was tyrosine (Fig. 1D), which could not be synthesized by 9% of the tested Enterobacteriaceae. None of the enterobacterial genomes analysed had lost the ability to produce methionine or threonine.

Mapping all detected auxotrophies onto the lifestyles of the 949 Eubacterial species analysed (Pagani et al. 2012) indicated that 85% free-living, 64% gut-inhabiting, and 91% endosymbiotic bacteria were predicted to be auxotrophic for at least one metabolite (Fig. 1B). Bacteria of the intestinal microflora were less frequently auxotrophic than free-living bacteria and endosymbionts (Chi-squared test with Yates correction: $\chi^2 = 19$, $P = 1.5*10^{-5}$, n=111 and 246 and $\chi^2 = 13$, $P = 3.3*10^{-4}$, n=111 and 57, respectively; Fig. 1B). Furthermore, auxotrophic endosymbionts were predicted to be auxotrophic for 20 metabolites per organism (median), which is significantly more than was predicted for auxotrophic free-living and gut-inhabiting bacteria (both groups: median of 2 auxotrophies per organism) (Mann-Whitney U-test with continuity correction: $W = 9541$, $P<2.2*10^{-16}$, n=52 and 209 and $W = 3033.5$, $P = 8.8*10^{-10}$, n=52 and 71, respectively; Fig. 1B). The phylogenetic distribution of lifestyles among the 949 analysed organisms strikingly matched the phylogenetic distribution of all known bacteria with a completely sequenced genome (Fig. S2).

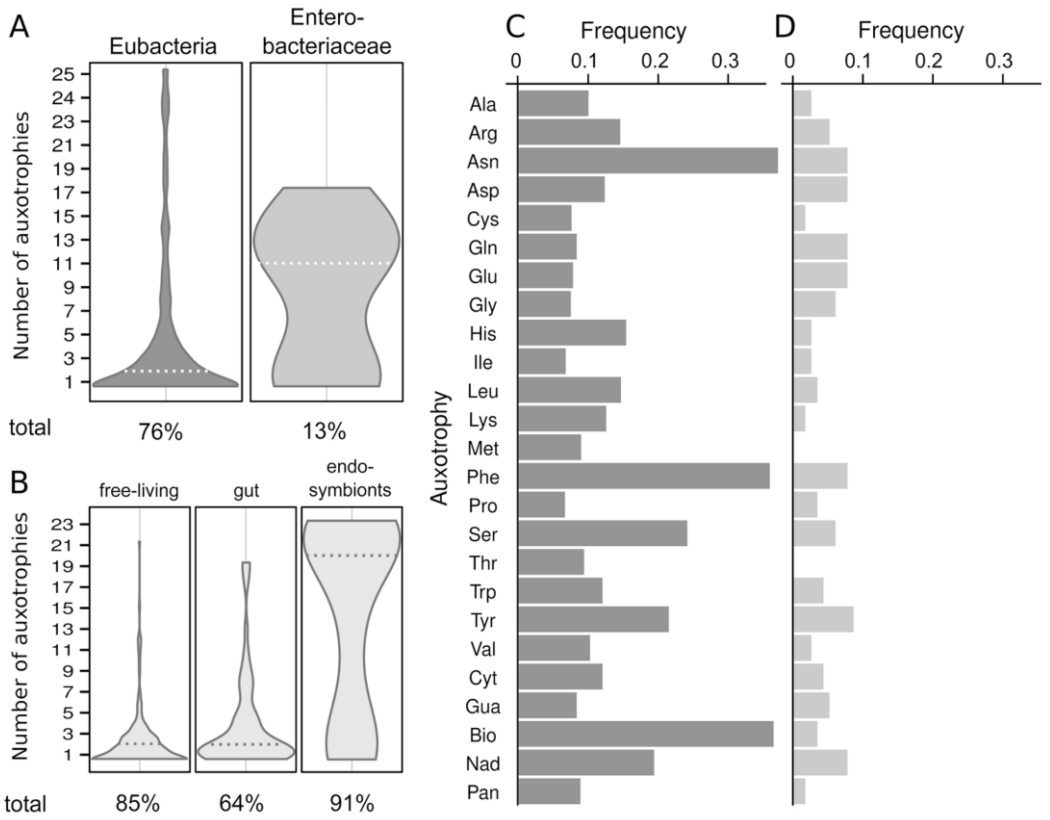


Figure 1. Distribution of metabolic auxotrophies in bacteria. Loss of a given biosynthetic function was predicted *in-silico* using 949 eubacterial, genome-annotated taxa (Vallenet et al. 2009; Caspi et al. 2012). (A) Distribution and median (dashed line) of the number of predicted auxotrophies per auxotrophic organism for Eubacteria (dark grey, n=949) and Enterobacteriaceae (light grey, n=116) for all 25 metabolites analysed. Percentages indicate the fractions of predicted auxotrophic organisms. Violin plots are scaled to the same maximum width. (B) Distribution and median (dashed line) of the number of auxotrophies for all auxotrophic Eubacteria depending on their lifestyle. Percentages indicate the fractions of predicted auxotrophic organisms within each lifestyle group. The lifestyle group sizes are: free-living (n=246), intestinal microflora-associated (n=111), and endosymbiotic organisms (n=57). Violin plots are scaled to the same maximum width. Frequencies of auxotrophies within (C) Eubacteria (n=949), and (D) *Enterbacteriaceae* (n=116). See Table S1 for abbreviations of metabolite names.

Taken together, our *in silico* analysis of eubacterial genomes predicted a surprisingly pervasive loss of multiple conditionally essential metabolic functions including the biosynthesis of amino acids, nucleosides, and vitamins. Furthermore, the distribution and frequency of auxotrophies was strongly dependent on the lifestyle of the bacterial species analysed.

3.2 Auxotrophy-causing mutations are beneficial when the focal metabolite is present in the environment

The growth of *E. coli* WT in monocultures was significantly enhanced by the supplementation of five compounds (i.e. His, Met, Phe, Trp, and Nad; FDR-corrected independent sample t-tests: $P \leq 0.05$, n=4; Fig. S5A,C) although different concentrations of each metabolite were required to achieve this effect. In contrast, growth was unaffected by the addition of eight metabolites (i.e. Ile, Leu, Lys, Pro, Tyr Cyt, Bio, and Pan; FDR-corrected independent sample t-tests: $P > 0.05$, n=4; Fig. S5) and even inhibited by three of the sixteen metabolites tested (i.e. Arg, Thr, and Gua; FDR-corrected independent sample t-tests: $P \leq 0.05$, n=4; Fig. S5A,B).

When each of these metabolites was supplied in increasing concentrations to the corresponding auxotrophs, growth was strongly dependent on the concentration of the respective nutrient (Fig. S6). Half of all gene deletions tested resulted in a maximum population density (i.e. OD) that was significantly increased over WT levels (FDR-corrected independent sample t-tests: $P \leq 0.05$, n=4; Fig. 2A; Fig. S6) at some concentration of the focal metabolite. Exceptions were the auxotrophs for His, Lys, Phe, Tyr, Cyt, and Nad, whose maximum population density did not exceed WT levels (FDR-corrected independent sample t-tests: $P > 0.05$, n=4, Fig. 2A; Fig. S6) as well as the Pro- and Thr auxotrophs that did not even reach WT levels under the range of concentrations tested (FDR-corrected independent sample t-tests: $P < 0.05$, n=4; Fig. 2A; Fig. S6A). The growth advantage over WT of the Arg- and Gua auxotrophs was probably attributable to a significant inhibitory effect of the metabolites added on the growth of the WT (FDR-corrected independent sample t-tests: $P \leq 0.05$, n=4; Fig. S5A,B, Fig. S6A,B), rather than an enhanced growth of the auxotrophic strains. Notably, growth of vitamin auxotrophs exceeded WT levels at much lower concentrations (0.2 - 0.5 μM) than was the case for nucleobase- and amino acid auxotrophs (25 - 200 μM; Mann-Whitney U-test: $P = 1.04*10^{-9}$, n=12 and 28; Fig. 2A, Fig. S6).

To verify whether the observed fitness advantages also manifest when an auxotrophic mutant directly competes against its prototrophic ancestor, pairwise competition experiments were performed, in which each of 16 auxotrophs were directly competed against the prototrophic WT in environments that contained different concentrations of the focal metabolites. Under these conditions, all auxotrophs except the Cyt and Nad auxotrophs reached fitness values that significantly exceeded WT levels (FDR-corrected paired sample t-tests: $P \leq 0.05$, n=4; Fig. 2B; Fig. S6). This included also auxotrophs, whose fitness did not increase over WT levels in monocultures (i.e. auxotrophs for His, Lys, Phe, Pro, Thr, Tyr). Similar to monocultures, vitamin auxotrophs achieved their maximum relative fitness at much lower concentrations (0.05 - 0.5 μM) than amino acid- and nucleobase-deficient strains (25 – 200 μM) required to exceed WT growth (Mann-Whitney U-test: $P = 3.58*10^{-10}$, n=16

and 52; Fig. 2B; Fig. S6). Biotin, one of the compounds for which biosynthesis genes were most frequently lost in natural bacterial isolates (Fig. 1) was required in the lowest concentrations of all metabolites analysed in both mono- and co-culture experiments (Fig. 2A,B; Fig. S6).

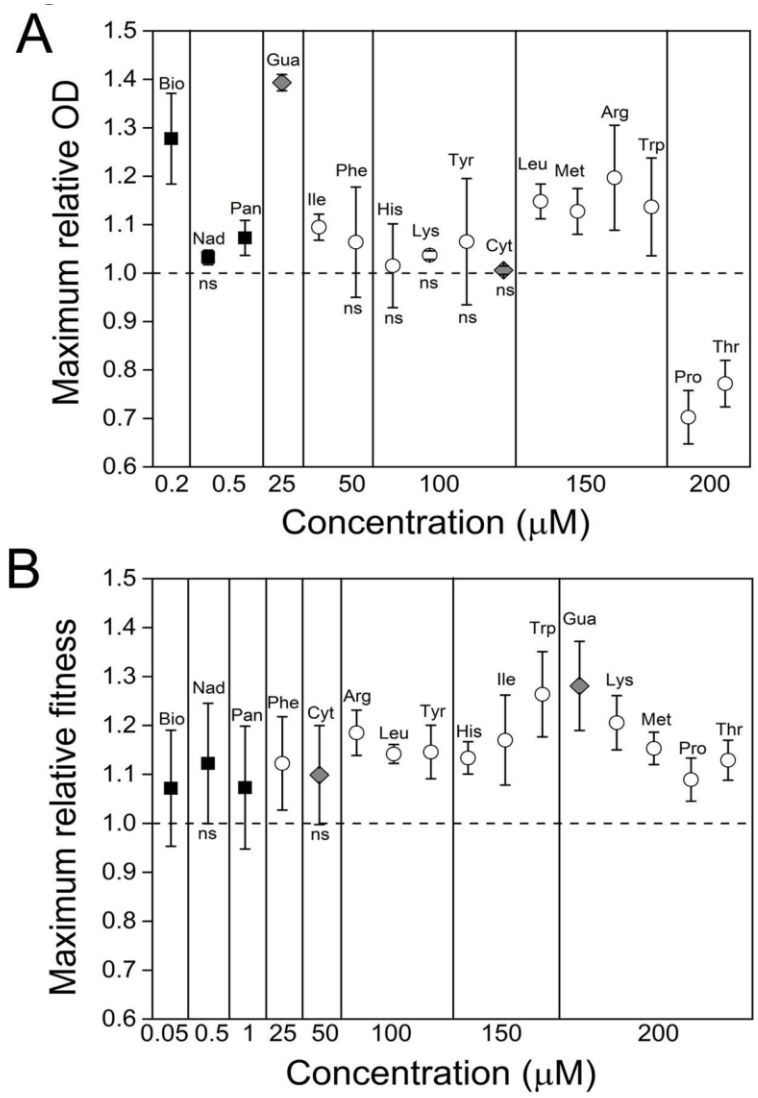


Figure 2. Maximum productivity and competitive fitness of *Escherichia coli* auxotrophs relative to WT. (A) Maximum OD in monoculture, and (B) maximum fitness in coculture of the amino acid- (circles), vitamin- (squares), and nucleobase auxotrophs (diamonds) relative to WT. All values are medians (± 95% CI) of four replicates and are significantly different from WT levels (i.e. dashed line; FDR-corrected independent sample t-tests (monoculture) and paired sample t-tests (coculture): P≤0.05, n=4), except those marked by 'ns'. See Table S1 for abbreviations of metabolite names.

Taken together, these results indicate that the loss of essential biosynthetic genes from the genome of *E. coli* generally resulted in strong and significant fitness advantages over the prototrophic WT when the required compounds were sufficiently present in the environment. The extent of fitness advantage, however, was context-dependent and strongly

affected by i) the concentration of the metabolite in the environment, ii) the identity of the metabolite, and iii) the absence/ presence of a competitor.

3.3 Fitness benefits depend on which gene of a biosynthetic pathway is lost

Amino acid biosynthesis involves the action of multiple enzymes that are encoded by different genes. Thus, the fitness benefit a strain gains by not having to carry out a certain biosynthetic step may differ depending on which gene has been lost. Observing different fitness benefits when different genes of the same pathway are lost may reflect differences in the biosynthetic costs incurred at each step or regulatory interactions among genes.

To verify this possibility, several genes were individually deleted from the biosynthetic pathways for Arg (6 genes), His (4), and Trp (4) (Fig. 3, and Table S2) whose deletion renders the resulting mutant auxotroph for the corresponding amino acid. All generated mutants were individually competed against WT in three environments, which differed in the concentration of the focal amino acid. The range of these concentrations covered a span (± 50 μM) around the concentration at which the terminal deletion mutant of each pathway had reached maximum fitness relative to cocultured WT (Fig. 2B, Fig. S6).

Fitness consequences resulting from the loss of a conditionally essential gene from one of the three multi-step pathways analysed strongly depended on both the identity of the lost gene as well as the concentration of amino acids available (Fig. 3). A pattern that seemed to emerge was that as the amino acid concentration in the environment increased, deletion of terminal genes tended to be more advantageous than the loss of more anterior genes (FDR-corrected paired-sample t-tests and Brown Forsythe tests followed by a LSD or Tamhane's T2 post-hoc test : $P<0.05$, $n=8$; Fig. 3). This trend was evident in two out of the three amino acid concentrations assayed for each of the three pathways analysed (Fig. 3). Furthermore, one of the three amino acid concentrations tested for each biosynthetic pathway caused a significant positive correlation between the mutants' relative fitness and the position of the deleted gene within the pathway (Pearson product-moment correlation: Arg pathway, 100 μM: $r = 0.33$, $P = 0.012$, $n = 48$; His pathway, 200 μM: $r = 0.5$, $P = 0.05$, $n = 28$; Trp pathway, 200 μM: $r = 0.55$, $P < 0.001$, $n = 32$). Interestingly, when the amount of protein invested by *E. coli* to catalyse different steps of these biosynthetic pathways was taken into account, the protein investment also increased towards the end of these pathways (Spearman's rank correlation: $\rho=0.55$, $P=0.02$, $n=17$; Fig. S7). Calculating the energetic cost for the individual coding sequences of these three pathways as well as the corresponding

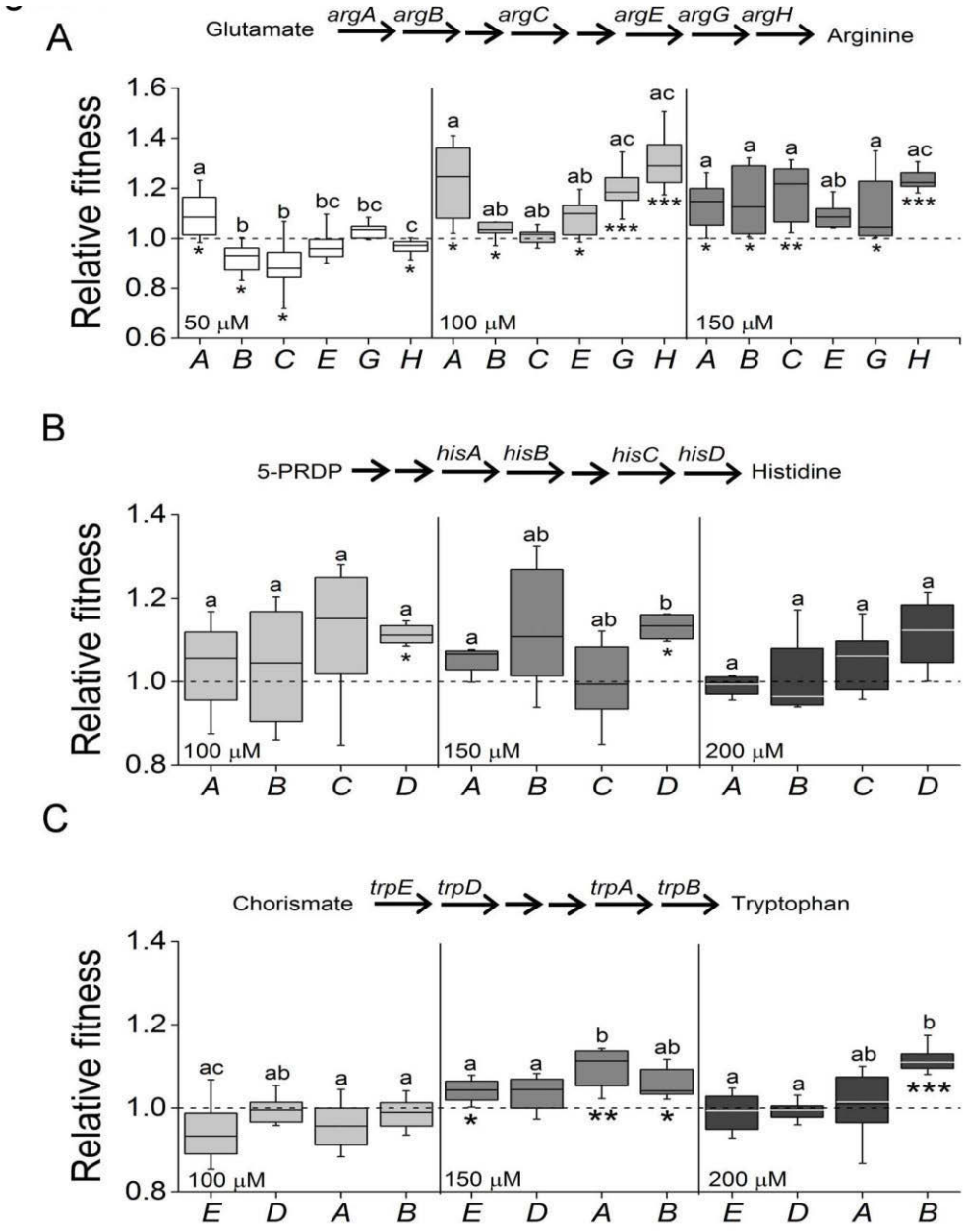


Figure 3. Competitive fitness of auxotrophic *Escherichia coli* mutants that lack different genes of the same biosynthetic pathway. Fitness of different deletion mutants that are auxotrophic for (A) Arg, (B) His, and (C) Trp was determined relative to WT. Experiments were conducted in minimal medium to which 50 μM, 100 μM, and 150 μM of Arg (A) or 100 μM, 150 μM, and 200 μM of either His (B) and Trp (C) has been supplemented. X-axes are labelled with the last letter of the focal gene's name (e.g. A for *argA, hisA or trpA*). Asterisks denote significant differences from WT levels (i.e. dashed line; FDR-corrected paired sample t-tests: *P<0.05, **P<0.01, and ***P<0.001). Different letters indicate significant differences among deletion mutants (univariate ANOVA followed by a LSD or Tamhane's T2 post-hoc test: P<0.05; n=8). Boxplot: median (horizontal lines in boxes), interquartile range (boxes, 1.5x- interquartile range (whiskers). Pathway insert: The flow of biosynthetic steps in each pathway. Unlabelled arrows represent non-essential genes. 5-PRDP: 5-phospho-α-D-ribose 1-diphosphate.

protein machinery in *E. coli* (Supporting Methods) revealed a significant greater protein cost (Wilcoxon signed rank test: P=0.002, n=10) that exceeded DNA biosynthesis costs by factor 34 (Table S3). Hence, these results suggest that a saving of protein costs may be involved in explaining the observed gain in fitness.

In case of the Arg biosynthetic pathway, the *argA* deletion mutant displayed a particularly strong fitness increase over WT in two of the three Arg concentrations tested (FDR-corrected paired sample t-test: P<0.05, n=8; Fig. 3A). Interestingly, the gene product of *argA* (i.e. N-acetylglutamate synthase) catalyses the first step in the Arg biosynthesis pathway and is the target enzyme for feedback inhibition by arginine (Vyas and Maas 1963). However, deletion of *trpE* and *trpD* (i.e. anthranilate synthase), which fulfil the same function in the Trp biosynthesis pathway (Pabst et al. 1973), did not result in a similar effect (Fig. 3C). Thus, the particularly strong fitness advantage gained by *argA* deletion mutants in the presence of sufficient amounts of Arg points to a special regulatory role this gene plays within the Arg biosynthesis pathway.

Together, these results demonstrate significant gene-specific fitness effects that arise upon deletion of different genes of the same metabolic pathway and suggest the saving of protein costs may be involved in explaining these differences.

3.4 Also *Acinetobacter baylyi* auxotrophs gain a fitness advantage upon gene loss

All *Acinetobacter baylyi* auxotrophs (i.e. *ΔhisD, ΔleuB, and ΔtrpB*) except the *ΔargH* mutant gained a significant fitness advantage upon gene loss when the corresponding amino acid was present (FDR-corrected paired sample t-tests: P≤0.05, n=8; Fig. 4). As previously observed in *E. coli*, the fitness advantage gained by *A. baylyi* auxotrophs was strongly dependent on the concentration of the focal metabolite, yet followed a completely different, downright opposite pattern (Fig. 4). Interestingly, only one of the four *A. baylyi* auxotrophs tested (i.e. *ΔhisD*) gained an advantage in relative fitness that was significantly increased over the fitness levels that the corresponding *E. coli* auxotrophs achieved under the same conditions (FDR-corrected independent sample t-tests: P≤0.05, n≥4). In sum, these results corroborate that the loss of essential biosynthetic genes can be selected for when the required metabolite is present in the environment, yet point to significant, species-specific differences.

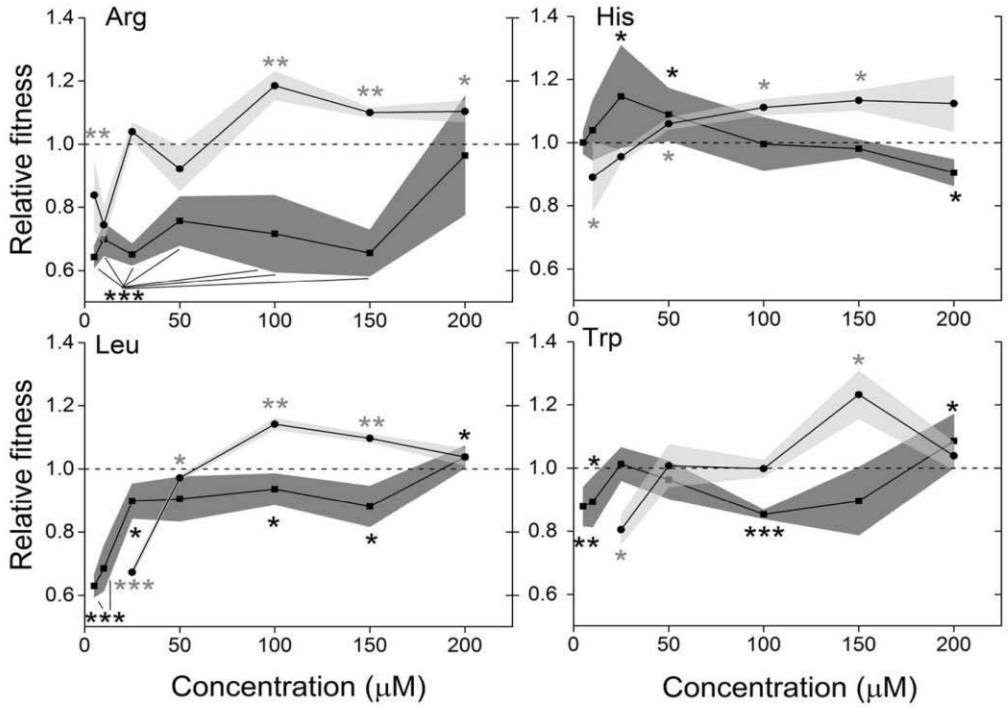


Figure 4. Competitive fitness of *Acinetobacter baylyi* and *Escherichia coli* auxotrophs relative to WT in increasing concentrations of the focal amino acids. Fitness of *E. coli (circles) and A. baylyi (*squares) mutants auxotrophic for Arg, His, Leu, and Trp relative to the corresponding WT. All values are medians of four replicates for *E. coli* and eight replicates for *A. baylyi*. The grey and dark grey regions mark the 95% confidence intervals for *E. coli* and *A. baylyi,* respectively, and the grey and dark grey asterisks mark significant differences of the *E. coli* and *A. baylyi* auxotrophs to WT levels (i.e. dashed line; FDR-corrected paired sample t-tests: *P<0.05, **P<0.01, and ***P<0.001, n≥4).

4. Discussion

Our analysis revealed that the loss of conditionally essential genes, which likely results in metabolic auxotrophies, is not limited to bacterial endosymbionts, but equally prevalent amongst free-living bacteria. However, why do microorganisms loose genes at the expense of their metabolic autonomy? For endosymbiotic bacteria, this question is commonly answered by pointing to their small population sizes and a lack of genetic recombination. These factors should result in a relaxed selection even for essential genes and - combined with a strong effect of genetic drift - could explain the rapid erosion of symbiont genomes [47,48,50,63,74,82,152]. However, free-living and gut-dwelling bacteria drastically differ from bacteria with an intracellular lifestyle in terms of their population biology as well as the selective environment they experience. Also, the high degree of metabolic complementarity and mutual inter-dependency that has been frequently observed among co-occurring endosymbionts [66,69,79] is likely favoured and maintained by natural selection.

To experimentally determine the potential role of selection in favouring mutants that lack essential genes, different biosynthetic genes were individually deleted from the genomes of two bacterial species and the resulting auxotrophic mutants systematically analysed. This analysis revealed that i) the loss of essential biosynthetic genes was generally beneficial when the required metabolite was sufficiently present in the cells' growth environment, ii) the metabolite concentration an auxotroph required to attain WT growth levels differed significantly depending on the metabolite as well as the species analysed, iii) the loss of different genes from the same metabolic pathway resulted in differential fitness consequences for the corresponding mutants, and iv) auxotrophs of two species that lacked the same biosynthetic gene responded very differently when exposed to the same concentrations of the required amino acid.

4.1 What causes the unexpectedly strong fitness advantage?

A key finding of this study is that the loss of different biosynthetic genes gave rise to different fitness benefits when the focal metabolite was sufficiently present in the mutants' growth environment. This was not only true for genes of different metabolic pathways, but also when genes of the same biosynthetic pathway were considered. A number of relevant insights emerge from this analysis. First, it made a significant difference whether a mutant's phenotype was indirectly compared to WT (monoculture) or directly competed against WT (coculture). Here, both i) the minimally required metabolite concentration as well as ii) the maximally achieved advantage over WT differed between the two perspectives. These findings cannot be exclusively explained by the costs auxotrophs save for the production of the focal metabolite relative to WT. Instead, other factors like the cells' requirement for a given metabolite and/ or the auxotrophs' transport efficiency with which they can take up different metabolites may have caused this pattern. Second, the finding that the deletion of different genes from the same biosynthetic pathway engendered different fitness consequences for the resulting auxotrophic mutants, suggests the unexpectedly strong fitness advantage of auxotrophs is at least partially caused by effects emanating from the loss of individual genes rather than a systemic response. The seeming increase of the fitness advantage auxotrophic mutants gained when terminal genes of a given pathway were deleted together with the concurrent enlarged investment of protein mass towards the end of these pathways implies the saving of protein costs may contribute to the observed gain in fitness. This interpretation is in line with empirical evidence, which suggests protein costs can significantly limit bacterial growth [90,109,167] or cause redistributions of metabolic fluxes to less expensive pathways [168]. However, Dykhuizen (1978), who addressed this question previously in *E. coli* did not find evidence for a cost-saving of Trp auxotrophs relative to prototrophic revertants [86]. Another possibility is a metabolic or regulatory rewiring that

renders auxotrophs more efficient in coping with amino-acid deficient conditions [169]. This could be achieved by an enhanced uptake of amino acids or a reallocation of the cell-internal protein pool. Future work should scrutinize these hypotheses.

4.2 Distribution of metabolic auxotrophies in nature

Our *in silico* analysis provides a first systematic assessment of the prevalence of putative metabolic auxotrophies among eubacteria. Even our conservative estimation indicated that the vast majority of genomes analysed lacked conditionally essential biosynthetic genes. A recent study corroborates these findings: reconstructing metabolic models of 55 sequenced *E. coli* and *Shigella* strains revealed multiple auxotrophies for vitamins and amino acids in twelve of these strains [67]. Taken together, these analyses suggest metabolic auxotrophies may be more widespread than previously thought.

However, can the distribution pattern of auxotrophies predicted for Enterobacteriaceae (Fig. 1D) be explained with the different fitness advantages observed in this study (Fig. 2; Fig. S6)? A series of experimental tests in which different kinetic parameters determined in this study (i.e. maximum OD/ relative fitness reached after 24 h, slope of metabolite dependency curve (i.e. metabolite concentrations versus OD/ relative fitness after 24 h)) was correlated to the predicted enterobacterial auxotrophy frequencies did not detect significant relationships between these parameters (Spearman rank correlation: $P > 0.05$ in all cases). This result is likely caused by fact that the frequency with which certain biosynthetic genes are lost is due to the availability of the corresponding metabolites in the strains' natural environments and not the potentially gained fitness advantage. Moreover, since strains are likely auxotrophic for more than one metabolite (Fig. 1A), epistatic interactions among these mutations may affect the fitness consequences of individual mutations.

4.3 Adaptive gene loss and the formation of inter-organismal networks

Our results imply that whenever local metabolite concentrations exceed certain threshold levels, strong selection pressures build up that favour the loss of the corresponding biosynthetic functions in bacteria. Thus, our analysis provides a plausible adaptive explanation for the widespread loss of conditionally essential biosynthetic genes (Fig. 1). Accordingly, amino acid concentrations in certain bacteria-inhabited environments such as soil or insect guts generally exceeded the levels required for auxotrophic mutants to outcompete prototrophic cells by orders of magnitude (Figure S8). In contrast, freshwater lakes exhibited only meagre amounts of free amino acids (i.e. 2.6 - 4124 nM) [170], which may explain why prototrophic *E. coli* strains seem to dominate in these environments [171]. Finally, metabolic auxotrophies have been observed to readily emerge in laboratory evolution

experiments with e.g. *Pseudomonas aeruginosa* [172], mutator strains of *E. coli* [173] that adapted to the mouse gut or *Legionella pneumophila* parasites adapting to mouse macrophages [174].

As these examples illustrate, metabolites can either originate from the growth environment or be produced by another organism [51,175]. The '*compensated gene loss*' resulting from the latter [175] can account for the rapid reduction of genome size of both parasitic [176]) and mutualistic bacterial symbionts and is likely also driving the formation of tightly integrated metabolic networks of co-occurring bacterial endosymbionts [69,177]. Our finding that this phenomenon is not restricted to organisms that interact over long periods of time, but also occurs among seemingly independent and free-living bacteria (Fig. 1) implies a pervasive role of adaptive gene loss for driving the evolution within microbial communities. An unavoidable leakage of vital metabolites during bacterial growth and subsistence [51] combined with the enormous and prevalent fitness advantages gained upon gene loss as observed in this and other studies [24,85,86], should result in the formation of intricately connected, inter-cellular networks. By mutually exchanging metabolites as '*public goods*', while at the same time specialising in the production of a reduced subset of metabolites, both the individual genotype and the whole bacterial community might benefit [24,178]. In particular, the difference in the concentration-dependent fitness advantage observed between two bacterial species (Fig. 4) may facilitate interspecific cross-feeding interactions. The general difficulty to isolate bacterial species from the wild [179,180] may be a reflection of this pattern.

5. Conclusions and outlook

Our study provides strong empirical support for the hypothesis that adaptive fitness advantages can account for the frequently observed loss of biosynthetic functions in bacteria. Our findings have a number of significant ramifications that should be investigated in future studies. First, the molecular causes underlying the unexpectedly strong fitness advantage upon gene loss should be identified. Second, as evidenced in our study, the natural bacterial isolates analysed were rarely auxotrophic for just one metabolite, but commonly lacked multiple biosynthetic capabilities simultaneously (Fig. 1A). Hence, future studies should address the question whether fitness effects combine additively when multiple auxotrophies are combined in one genetic background, or whether epistatic interactions limit an even further increase of the auxotrophs' fitness. Third, the '*black queen hypothesis*' [51] predicts for bacterial strains, which loose costly metabolites by leakage and that coevolve within a microbial community, to continuously loose biosynthetic genes until an equilibrium is reached, at which the benefit of gene loss is outweighed by its costs. Our study provides a first estimate of these benefits, thus allowing to further explore how they affect the race for biosynthetic disarmament within microbial communities.

Acknowledgements

We thank Ulrike Gerischer and Iris Steiner for sharing the A. baylyi ADP1 wild type strain as well as Dominik Refardt for providing us with strain E. coli REL 607. Constructive criticism by Erika Kothe and the EEE group as well as support by W. Boland is gratefully acknowledged. This work was supported by the Jena School for Microbial Communication (JSMC), the Volkswagen Foundation, and the Max Planck Society.

Supporting Information

Chapter 2

Chapter 2

Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes

Published in: Evolution (2015); Volume 69; Pages: 1244-1254; DOI:10.1111/evo.12640

Only minor typographical modifications have been made to the published article in this chapter.

Authors: Glen D'Souza*, Silvio Waschina*, Christoph Kaleta, and Christian Kost

* These authors contributed equally to this work

Abstract

Many bacterial lineages lack seemingly essential metabolic genes. Previous work suggested selective benefits could drive the loss of biosynthetic functions from bacterial genomes when the corresponding metabolites are sufficiently available in the environment. However, the factors that govern this 'genome streamlining' remain poorly understood. Here we determine the effect of plasticity and epistasis on the fitness of *Escherichia coli* genotypes from whose genome biosynthetic genes for one, two, or three different amino acids have been deleted. Competitive fitness experiments between auxotrophic mutants and prototrophic wild type cells in one of two carbon environments revealed that plasticity and epistasis strongly affected the mutants' fitness individually and interactively. Positive and negative epistatic interactions were prevalent, yet on average cancelled each other out. Moreover, epistasis correlated negatively with the expected effects of combined auxotrophy-causing mutations, thus producing a pattern of diminishing returns. Moreover, computationally analysing 1,432 eubacterial metabolic networks revealed that most pairs of auxotrophies co-occurred significantly more often than expected by chance, suggesting epistatic interactions and/ or environmental factors favoured these combinations. Our results demonstrate that both the genetic background and environmental conditions determine the adaptive value of a loss-of-biochemical-function mutation and that fitness gains decelerate, as more biochemical functions are lost.

1. Introduction

Bacterial genomes are not static entities, but are highly dynamic on evolutionary time-scales in terms of both size and composition [36]. Variation in the size of prokaryotic genomes can be caused by the duplication of existing genes, the acquisition of new genetic information from the environment (i.e. *horizontal gene transfer*), or, alternatively, by gene loss. Reductive genome evolution is a feature that characterizes many bacterial taxa and comparative genomics indicates that gene loss appears to be more important for shaping prokaryotic genomes than gene gain by horizontal gene transfer [36]. In many cases, one or more essential biosynthetic genes are lost, thus rendering the resulting auxotrophic bacteria dependent on an environmental uptake of the required metabolites [22,67,181]. Surprisingly, the loss of essential biosynthetic functions is not limited to endosymbiotic bacteria or intracellular parasites where essential nutrients can potentially be obtained from the host, but also prevails in free-living taxa such as saprophytes [53,182] or marine bacteria [83,183].

Two main scenarios can account for the frequently observed loss of conditionally essential biosynthetic genes from prokaryotic genomes: First, genetic drift may weaken selection even for essential genes and could thus explain the fixation of maladaptive mutations. This effect is likely strongest in small bacterial populations [50,74,76] such as endosymbiotic bacteria, which repeatedly undergo severe population bottlenecks during host-to-host transmission [79]. Second, the loss of biosynthesis genes may be selectively favored when the required metabolite is either sufficiently present in the growth environment or provided by co-occurring organisms [51,52]. Under these conditions, mutations that deactivate the biosynthetic machinery for a certain metabolite may result in the saving of production costs or could induce regulatory changes to economize the cell's resources, for example by rerouting metabolic fluxes, which allow the bacterial cell to better cope with starvation for the required metabolite.

Several studies using different bacterial species support the hypothesis that adaptive benefits may drive the loss of essential biosynthetic functions. In these cases, pairwise competition experiments between prototrophic bacterial cells and mutants lacking the ability to biosynthesize a certain metabolite pointed to a significant fitness advantage auxotrophs gain over prototrophic genotypes when the required metabolite is sufficiently present in the cells' growth environment [24,85,86,181]. Even though these studies suggest that metabolic loss-of-function mutants can be selectively favored, very little is known on how metabolic auxotrophies evolve.

Given that theoretical evidence predicts multiple auxotrophy-causing mutations are frequently co-occurring in the same genetic background [181], the extent to which these mutations interact with each other (i.e. *epistasis*) remains poorly understood. In other words, do the previously observed positive fitness effects combine additively as more loss-of-

function mutations accumulate in the same genome, or do epistatic effects constrain the fitness achievable by a multiply auxotrophic genotype? Moreover, natural habitats of bacteria are usually quite complex and may not only contain several primary metabolites (e.g. amino acids or vitamins), but also differ in the available carbon source. Since fluxes through metabolic networks change depending on the carbon source used [116,184], fitness and ultimately also epistatic interactions among mutations are expected to depend on the carbon source metabolized (i.e. *plasticity*).

An increasing number of studies suggest both epistasis [117,185,186] and plasticity [99,187] can significantly influence the trajectories of beneficial mutations accessible to evolving bacteria. The general pattern that seems to emerge from these analyses is that negative epistasis is prevalent and often results in diminishing returns as more beneficial mutations accrue in a single genetic background [117,186,187].

Understanding the constraints that determine evolutionary routes leading to multiply auxotrophic bacteria requires insight into how plasticity and epistasis influence the fitness consequences upon loss of metabolic genes. However, examining these effects in natural isolates is hampered by difficulties of cultivating auxotrophic genotypes under laboratory conditions or to manipulate the genome of non-model organisms. Thus, deleting a defined number of genes from the genome of a well-characterized model organism and evaluating the fitness consequences under carefully controlled growth conditions provides a tractable approach to quantify how environmental and genetic effects determine the fitness of multiply auxotrophic genotypes. Here we use a combination of computational and experimental approaches to address the following questions: (1) Do certain combinations of biosynthetic genes show an increased propensity to be jointly lost from bacterial genomes in nature? (2) How do fitness effects combine as multiple auxotrophy-causing mutations accumulate in the same genome? (3) Does the available carbon source affect fitness consequences of auxotrophy-causing mutations? (4) Do epistasis and plasticity interactively influence the effects of auxotrophy-causing mutations?

2. Materials and methods

2.1 Co-occurrence prediction of multiple auxotrophies

A previously published data set of amino acid auxotrophies that were predicted for different bacterial species [181] was updated to include the most recently available sequenced genomes and the resulting 1,432 eubacterial metabolic networks were subjected to further examination. To test, whether pairs of auxotrophies were statistically over- or underrepresented, the presence of reactions required for amino acid biosynthesis [181] was randomized, whilst controlling for the number of deletions (i.e. absence of a particular

reaction) per species and the number of species possessing a particular reaction. A total of 8,000 samples were randomly drawn from the [species x reaction existence] space using the Rasch Sampler [188] and auxotrophy frequencies were re-calculated from these random samples. Then, the frequencies of auxotrophy pairs predicted in the original data set to co-occur were compared to the expected distribution of double-auxotrophies inferred from the randomized data set. This approach allowed correcting the frequencies of predicted double-auxotrophies by the expected co-occurrence pattern of auxotrophies, which are simply due to chance (e.g. genetic drift) or the structure of the metabolic network (i.e. shared reactions in the biosynthetic pathways of two or more amino acids).

To test whether the observed co-occurrence pattern was reflecting the distribution of amino acids in natural environments, predicted auxotrophy frequencies were correlated with a published data set of 69 different aquatic, terrestrial, and host-associated environments [189]. For this, the medians of the pairwise products of relative amino acid abundances were correlated with the pairwise co-occurrence of predicted amino acid auxotrophies. In this analysis, the amino acids Glu/ Gln and Asp/ Asn were not considered, because the data set used did not allow distinguishing these pairs of amino acids.

2.2 Bacterial strains and their construction

Eleven different single gene deletions that each would render *E. coli* auxotrophic for a single amino acid were identified and constructed as described [181,190] (Table S1). All deletion alleles were transferred from existing strains [149] into *E. coli* BW 25113 [149] using P1 phage-mediated transduction [161] and recombinants were selected for their ability to grow on kanamycin-containing LB plates (50 μg ml^{-1}). In addition, 50 of the 55 possible combinations of double deletion mutants and 16 of 165 possible triple deletion mutants were successfully generated (Table S1). For this, single deletion mutants were first cured of the kanamycin resistance by excising the kanamycin cassette from the mutant's genomes using the pCP20 plasmid, which harbours the FLP recombinase [191]. Subsequently the second deletion allele was transferred into the resulting strains and successful recombinants were again selected for their resistance to kanamycin. A subset of double deletion mutants was cured of the kanamycin resistance cassette using the above-mentioned approach to yield triple deletion mutants. All generated genotypes used for subsequent experiments thus contained one copy of the kanamycin cassette in their genome, albeit at different chromosomal locations.

To examine the possibility that unintended, secondary mutations have been co-transduced, a control experiment was performed where the same deletion allele was repeatedly reintroduced into the same recipient genotype via P1 phage transduction. For this, three different deletion alleles were randomly selected (i.e. ⊿*asnB,* ⊿*mdh,* ⊿*argH*). After

transduction of *E. coli* BW 25113, the kanamycin cassette was cured and the same phage lysates were used to re-infect the recipient now carrying a deletion allele. This procedure was repeated three-times to mimic the number of transduction steps required to construct triple auxotrophic mutants. Fitness of the genotypes resulting from each mutational step was determined relative to the ancestral WT as described below. In none of the three cases was the fitness of the resulting mutants significantly affected by the number of transduction rounds (one-way ANOVA: $P > 0.05$, n=8 for each genotype). Thus, the phage transduction procedure used is very unlikely to have produced unintended, secondary mutations.

Conditional lethality of multiple auxotrophies was verified by inoculating 10^5 colony-forming units (CFUs) of these genotypes into 1 ml minimal medium for *Azospirillium brasilense* (MMAB) [158] without biotin and using fructose (5 g l^{-1}) as a carbon source. The optical density (OD) the corresponding mutant strain achieved during 24 h of growth was determined spectrophotometrically at 600 nm using a Tecan Infinite F200 Pro platereader (Tecan Group Ltd., Switzerland). The mutation was deemed conditionally essential when the auxotroph's growth did not exceed the OD_{600nm} of uninoculated minimal medium. This was the case for all double- and triple-gene deletion mutants generated. Gene deletions were confirmed by sequencing the corresponding genomic regions. To phenotypically distinguish genotypes in fitness experiments, the arabinose utilization locus (Ara^+) from strain REL 607 [162] was introduced into BW 25113 using P1 phage-mediated transduction [161].

2.3 Culture conditions

Cultures were incubated at 30 °C under shaking conditions and experiments were performed in MMAB minimal medium [158] without biotin and using either fructose (5 g l^{-1}) or disodium succinate (8.86 g l^{-1}) instead of malate as carbon source. The concentration of fructose and succinate was chosen such that, at least theoretically, the same amount of biomass could be produced under both carbon sources (see Supporting Methods for details). In addition, both media were supplemented with a mixture of all 11 amino acids, each at a concentration of 100 μM.

2.4 Competitive fitness assays

Competitive fitness experiments were performed in 96-deepwell plates (Eppendorf, Germany) with a culture volume of 1 ml. Auxotrophs were pre-cultured at 30 °C in MMAB medium supplemented with amino acids and the corresponding carbon source. For competitive fitness assays, ~10^5 CFUs ml^{-1} of WT and a focal auxotrophic mutant were co-inoculated into 1 ml MMAB medium (ratio: 1:1) supplemented with amino acids and the respective carbon source (i.e. fructose or succinate) and cell numbers were determined at 0 h

and 24 h by dilution plating. *E. coli* auxotrophs (Ara^-) were differentiated from WT (Ara^+) using the arabinose utilisation marker as described (Lenski et al. 1991). Competitive fitness of auxotrophs versus WT was determined by calculating the Malthusian parameter (M) of both genotypes: $M = (\ln (N_f/ N_i)/ 24)$, where N_i is initial number of CFUs at 0 h and N_f is the final CFU count after 24 h (Lenski et al. 1991). Relative fitness was calculated as the ratio of Malthusian parameters. Each competition assay was replicated eight times. Competition experiments between WT that did or did not contain the Ara marker provided no evidence for a fitness cost of the marker in either environment (i.e. fructose and succinate; independent sample t-test: $P > 0.05$, $n = 8$).

Given that all auxotrophic mutants generated contained a kanamycin resistance cassette in their genome, a possible fitness cost of this marker could theoretically affect the determined epistasis values: erroneously considering the cost of the marker multiple times when calculating the fitness expected for double and triple mutants from the fitness values of single gene deletion mutants, yet just once when determining the observed fitness of double- and triple mutants, could have resulted in an overestimation of the true epistatic effect (see below). To assess if the kanamycin marker incurred a cost to the auxotrophic strains, the Malthusian parameter of eleven single, four double, and four triple auxotrophic strains containing the kanamycin marker was determined in coculture with the same genotypes that have been cured from the marker (Table S2) as described above. Each competition experiment was initiated by mixing ~10^5 CFUs ml^{-1} of both competitors (ratio: 1:1) into 1 ml MMAB medium that contained fructose as the sole carbon source. The number of CFUs at 0 h and 24 h was determined by plating on LB agar plates that did or did not contain kanamycin (50 μg ml^{-1}) and the Malthusian parameter was calculated as described above. Finding that in these competition experiments the Malthusian parameter of none of the kanamycin-resistant mutants differed significantly from its kanamycin-sensitive counterpart (independent sample t-test: $P > 0.05$, $n = 10$ for all mutants, Table S2) provided no evidence for a fitness cost of this marker. The estimated minimum difference detectable by these tests [192] ranged between 0.24% - 2.8% (Table S2), which was well below the size of epistatic interactions determined (Tables S3 and S4), suggesting that a possible fitness cost of the kanamycin resistance marker used is very unlikely to have affected our results.

2.5 Calculating epistasis

Epistasis for multiple deletion mutants was calculated as the difference between the *observed* and *expected fitness*. *Expected fitness* was calculated by applying the multiplicative model [119,120,187]. Accordingly, for a genotype bearing two auxotrophy-causing mutations, the expected fitness would be the product of the observed relative fitness of the two mutations when individually present in a genotype. Epistasis was estimated as:

$$\mathcal{E}xy = Wxy - WxWy \quad (1)$$

$$\mathcal{E}xyz = Wxyz - WxWyWz \quad (2)$$

Equation 1 shows the calculation of epistasis for double deletion mutants and equation 2 for higher order (i.e. triple deletion) mutants. W is the relative fitness, W_{xy} and W_{xyz} is the relative fitness of strains with the entire set of two or three mutations respectively, and W_x, W_y, and W_z is the relative fitness of genotypes with just one deletion mutation. For higher order interactions (eq. 2), the sum of the effect of the lower order mutations was subtracted from $\mathcal{E}xyz$ in eq. 2 to obtain the net effect of higher order epistasis as show in equation 3:

$$\mathcal{E}'xyz = \mathcal{E}xyz - (\mathcal{E}xy + \mathcal{E}yz + \mathcal{E}xz) \quad (3)$$

Error for the estimated value of $\mathcal{E}$ was calculated using the method of error propagation [119,120] and epistasis was considered significant for a given combination of deletion alleles if $\mathcal{E}$ was outside the error.

2.6 Statistical analysis

The statistical relationship between the co-occurrence of predicted auxotrophies and the distribution of the corresponding amino acids in natural environments was assessed via a Kendall's rank correlation. Normal distribution of data was assessed using the Kolmogorov-Smirnov test. Homogeneity of variances was determined by applying Levene's test and variances were considered to be homogeneous when P>0.05. Fitness differences between auxotrophic mutants and their wild type competitors as well as between auxotrophic mutants in the two environments were determined with independent sample t-tests. One-way ANOVAs followed by Least Significant Difference (LSD) post hoc tests were employed to test if the mutants' fitness in either environment depended on the number of mutations. Significant deviations of epistasis from zero (no epistasis) were determined by applying one sample t-tests to the values of all mutants quantified in both environments. The statistical relationship between expected fitness and epistasis was analysed via a Pearson's product-moment correlation. The False Discovery Rate (FDR) procedure of Benjamini et al. (2006) was applied to correct P values after multiple testing [165]. The relationship between expected and observed fitness was analysed using a type II regression model. Slopes of regression lines were considered to be significantly smaller than 1 when their 95% confidence intervals did not include the 45° line (i.e. the perfect correlation between expected and observed fitness, which is the null hypothesis for non-epistatic interactions). A general linear model with '*fitness*' as dependent variable and '*environment*' as well as the presence of one of 11 mutations as '*mutation 1*', '*mutation 2',* and '*mutation 3*' as fixed factors was calculated to identify interactive effects among mutations and/ or the environment. Statistical analyses were performed using the SPSS package (version 17.0, IBM, USA) and the R software [166].

3 Results

3.1 Prevalent positive co-occurrence of auxotrophies in eubacterial genomes

A recent analysis of 949 eubacterial genomes and their inferred metabolic networks suggested that biosynthetic functions for amino acids, nucleotides, and vitamins are frequently lacking in the corresponding metabolic networks, indicating that auxotrophies are prevalent in natural populations of bacteria [181]. Interestingly, by re-analysing a more recent collection of 1,432 eubacterial genomes, 37% of all bacteria analysed were auxotrophic for two or more metabolites. If mutations that deactivate biosynthetic functions interact epistatically (i.e. non-additively), pairwise co-occurrence patterns of auxotrophies are expected to significantly deviate from a random distribution. Testing this prediction for a subset of 458 eubacteria that were predicted to be auxotrophic for multiple amino acids revealed for most pairwise comparisons (152 of 190) a significant positive association (FDR-corrected one-sample Wilcoxon test: $P < 0.05$, $n = 8{,}000$, Fig. 1), indicating that auxotrophies co-occur more frequently than expected by chance. A smaller fraction of 37 auxotrophy pairs co-occurred significantly less often than expected by chance (FDR-corrected one-sample Wilcoxon test: $P < 0.05$, $n = 8{,}000$) and only two pairs (Gly-Thr, Thr-Gly) showed a distribution that was statistically undistinguishable from a random distribution (FDR-corrected one-sample Wilcoxon test: $P > 0.05$, $n = 8{,}000$). Finding that virtually all amino acid double-auxotrophies deviate significantly in their frequency from the frequency expected by chance suggests epistatic interactions and/ or environmental factors favored these combinations.

To test whether auxotrophy co-occurrences were caused by an increased propensity of certain amino acids to co-occur in natural environments, the predicted auxotrophy frequencies were correlated with quantitative measurements of relative amino acid concentrations in 69 different environments [189]. At first, the median of pairwise products of relative amino acid concentrations did not correlate significantly with the auxotrophy co-occurrence data (Kendall's rank correlation: $R_\tau = 0.04$, $P = 0.49$, $n = 120$). However, a closer look at this correlation revealed that in most environments the amino acids alanine (Ala) and glycine (Gly) were relatively abundant, while the corresponding auxotrophies were relatively rare [181]. As the two smallest proteinogenic amino acids, Ala and Gly are the metabolically cheapest to produce [87,88]. Thus, Ala and Gly auxotrophies might not be very frequent in eubacteria, because the energetic savings to lose these biosynthetic functions (i.e. the selective advantages) are relatively low. Moreover, several possible alternative biosynthetic reactions for Ala and Gly exist in prokaryotes [181], which might limit the frequency of auxotrophies. Excluding Ala and Gly from the analysis for these reasons resulted in a highly significant positive correlation between the frequency of double-auxotrophies and the pairwise abundance of amino acids in the environment (Kendall's rank correlation, $R_\tau = 0.22$,

P = 0.003, n = 91, Fig. S1), which is consistent with an environmentally favoured loss of metabolic genes. Taken together, the analysis of amino acid auxotrophy distributions in eubacteria suggests epistatic interactions and/ or an environmentally compensated gene loss may have caused the observed co-occurrence pattern of amino acid auxotrophies.

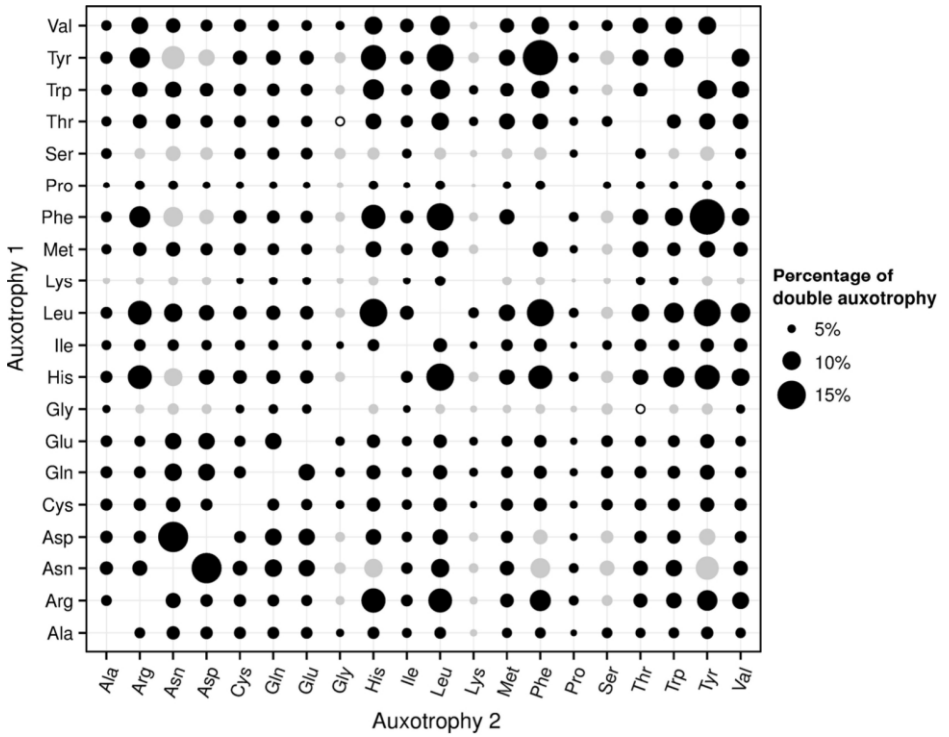


Figure 1. Predicted pairwise co-occurrence of amino acid auxotrophies in eubacterial genomes. Sizes of circles represent the proportion of genotypes (%) predicted to be simultaneously auxotroph for the two corresponding amino acids. Filled circles indicate pairs of auxotrophies, which co-occurred significantly more (black) or less often (grey) than expected by chance (FDR-corrected one-sample Wilcoxon test: P < 0.05, n = 8,000), while unfilled circles depict pairs with a random co-occurrence pattern (P > 0.05, n = 8,000). The data set included 1,432 eubacterial genomes that were predicted to be auxotrophic (584) for one or more of 20 different amino acids or prototrophic (848) for all amino acids.

3.2 Negative epistasis causes diminishing returns with fitness of multiply auxotrophic genotypes

How does cellular fitness scale with an increase in the number of auxotrophy-causing mutations? To address this question, one, two, or three of 11 genes that render the resulting mutant auxotrophic for amino acids were deleted from the same genetic background of *E. coli*. Altogether, 11 mutant strains bearing one (hereafter: '*single mutants*'), 50 mutants bearing two (hereafter: '*double mutants*'), and 16 strains bearing three different amino acid auxotrophy-causing mutations (hereafter: '*triple mutants*') were generated (Table S1). Subsequently, the competitive fitness against prototrophic WT was determined for all 77 strains in two distinct growth environments that contained an equimolar concentration of 11 amino acids (100 μM each), yet differed in the carbon source available (i.e. either fructose or succinate). These

carbon sources were chosen as they are important intermediates in the primary metabolism of most bacterial species, but derive from different points of the cells' metabolic network: fructose-6-phosphate being a core metabolite in glycolysis, while succinate is part of the tricarboxylic acid pathway.

This analysis indicated for the fructose-containing environment that both double and triple mutants were significantly less fit than the corresponding single gene deletion mutants (one-way ANOVA followed by a LSD post-hoc test: $P < 0.05$, d.f. = 76; Fig 2 A). In contrast, the relative fitness of single-, double-, and triple-gene deletion mutants did not differ in the succinate environment (one-way ANOVA followed by a LSD post-hoc test: $P > 0.05$, d.f. = 76; Fig 2 B).

Quantitatively assessing the degree with which the effects of focal mutations deviated from expected finesses revealed on average no predominant influence of either positive or negative epistatic effects (Fig. S2). This pattern held true for both the fructose (mean epistasis: -0.05 ± 0.05, one sample t-test: $P = 0.26$, d.f. = 65, Fig. S2) and the succinate environment (mean epistasis: 0.003 ± 0.03, one sample t-test: $P = 0.93$, d.f. = 65, Fig. S2). However, analysing epistatic effects for all genotypes individually uncovered for the fructose environment 20 cases of

significantly negative and 13 cases of significantly positive epistatic interactions (Table S3), while in the succinate environment 26 instances showed significant negative and 25 cases significant positive epistatic interactions (Table S4) (one sample t-test: $P < 0.05$, n = 8; Table 1). Thus, positive and negative epistatic effects cancelled each other out, thereby causing the abovementioned non-significant average deviation.

When the relation between expected relative fitness effects of multiple gene deletions as predicted from individual mutations and observed epistasis was scrutinized, a negative correlation (Pearson product-moment correlation: $r = -0.27$, $P = 0.03$ for fructose and $r = -0.27$, $P = 0.03$ for succinate) was observed for both carbon environments tested (Fig. 2 C and D). In other words, epistatic interactions among mutations became more negative as the predicted fitness increased. Theoretically, this relationship could also be caused through a phenomenon called regression-to-the-mean [193], in which measurement error alone can cause a negative correlation between expected fitness and epistasis due to the statistical non-independence between expected fitness and epistasis. To test whether this phenomenon could explain the observed diminishing fitness returns, a Type II regression was applied. Finding no correlation between observed and expected fitness in either carbon environment (Type II regression: $P > 0.05$, n = 66, Fig. S3), while both slopes were significantly smaller than 1 corroborated that the fitness of multiply auxotrophic genotypes showed a pattern of true diminishing returns.

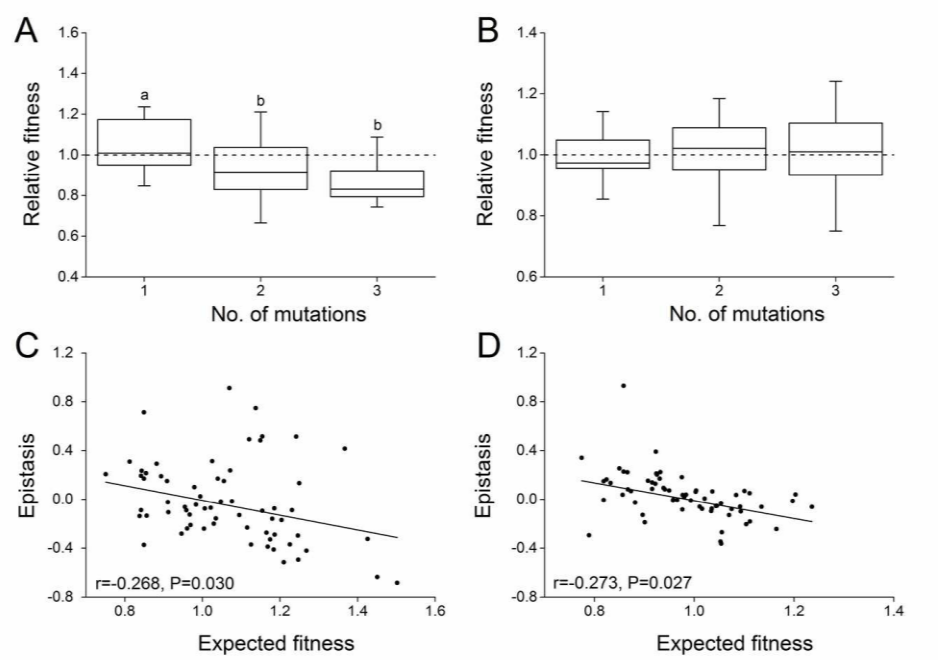


Figure 2. Change of relative fitness with increasing numbers of auxotrophy-causing mutations and relation between epistasis and expected relative fitness. (A, B) Competitive fitness of mutants bearing one, two or three auxotrophy-causing mutations relative to prototrophic WT cells in minimal media containing either (A) fructose or (B) succinate. The dashed line represents fitness levels of the WT. Different letters indicate significant differences among deletion mutants (univariate ANOVA followed by a LSD post-hoc test: $P < 0.05$; n = 11 (single deletions), 50 (double deletions), and 16 (triple deletions)). Boxplots: median (horizontal lines in boxes), interquartile range (boxes, 1.5x- interquartile range (whiskers). (C, D) Relation between absolute epistasis and expected fitness determined in minimal medium containing (C) fructose and (D) succinate. Values of all double- and triple gene deletion mutants are shown and both panels include the results of a Pearson's product-moment correlation.

Table 1. Number of epistatic interaction identified in 50 double- and 16 triple gene deletion mutants in both carbon (C) environments analysed.

	Epistasis[1] in double mutants			**Epistasis[1] in triple mutants**		
C-environment	negative	zero	positive	negative	zero	positive
Fructose	14	27	9	6	6	4
Succinate	22	14	14	4	1	11

[1]For details on how epistatic interactions were determined please see Materials and methods.

Taken together, these experiments showed that the fitness consequences of losing conditionally essential biosynthetic genes did not increase linearly with the number of biosynthetic functions lost and that negative epistasis caused diminishing returns with mutant fitness.

3.3 Fitness consequences of auxotrophy-causing mutations depend on the available carbon source

Directly comparing the fitness levels the single gene deletion mutants achieved in both environments revealed that only the fitness of one of eleven auxotrophs tested (i.e. $\Delta ilvA$) was plastic with respect to the carbon source present and significantly fitter in the fructose- than in the succinate-containing environment (FDR-corrected independent samples t-test, $P < 0.05$, d.f. ≥ 8; Fig. 3 A). When multiply auxotrophic genotypes were also considered, a tenth (5/50) of the double mutants and two of the 16 triple mutants tested attained a significantly higher relative fitness when grown in fructose than when grown in succinate (FDR-corrected independent samples t-test: $P < 0.05$, d.f. ≥ 10 ; Fig 3 B and C). Conversely, about half of the other double and triple mutants (22/50 and 10/16, respectively) were fitter in the succinate than in the fructose-containing environment (FDR-corrected independent samples t-test: $P < 0.05$, d.f. ≥ 10; Fig. 3 B and C). However, the fitness of 23 double mutants (46%) and four triple mutants (25%) was unaffected by the available carbon source used (FDR corrected independent samples t-test: $P < 0.05$; Fig 3 B and C). Together, these results suggest that the fitness of auxotrophic mutants is highly dependent on the ambient environmental conditions.

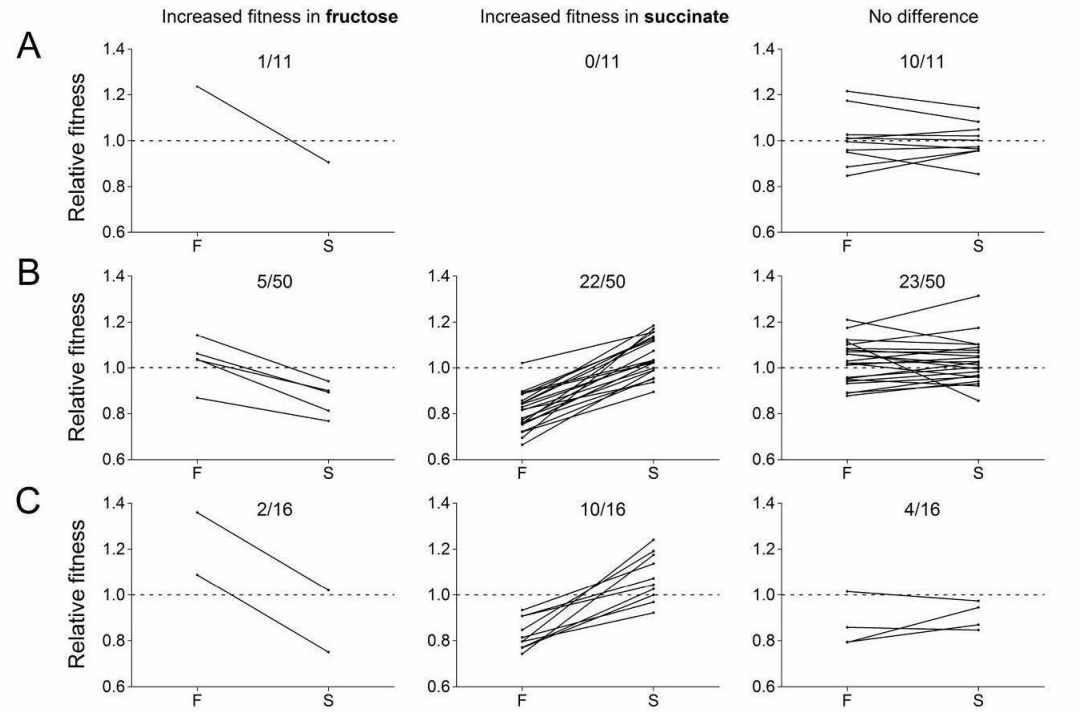


Figure 3. Reaction norms of competitive fitness of different auxotrophic genotypes against prototrophic wild type in two different carbon environments. Each line depicts the competitive fitness of genotypes having (A) one, (B) two, or (C) three auxotrophy-causing mutations. Competition experiments against prototrophic WT (dashed line) were conducted in minimal media containing either fructose (F) or succinate (S). Differences in the mutants' relative finesses in both carbon environments were assessed using FDR-corrected independent sample t-tests ($P < 0.05$, d.f. ≥ 8). Numbers above panels indicate the number of cases (left) and the total number of mutants tested (right).

3.4 Plasticity and epistasis jointly influence the fitness of multiply auxotrophic mutants

The above findings suggested that interactions among mutations (i.e. *epistasis*, G x G), interactions between mutations and the environment (G x E), and possibly also interactions of epistasis and the environment (G x G x E) determined the fitness of defined auxotrophic mutant genotypes. Statistically evaluating the effect of these three parameters on the mutants' fitness indicated indeed a highly significant effect of epistasis (univariate ANOVA: $P < 0.0001$), G x E (univariate ANOVA: $P < 0.001$), and G x G x E (univariate ANOVA: $P < 0.0001$). Together, these findings show that the fitness associated with loosing conditionally essential biosynthetic genes is strongly affected by other metabolic mutations in the genome as well as the given nutritional environment.

4. Discussion

Knowledge on how fitness effects of a given mutation depend on other mutations present in the genome, the selective environment or both is key to understanding adaptive processes, because the topology of the genotype-phenotype map determines the evolutionary trajectories that are accessible to organisms evolving within these fitness landscapes. Here we focus on the fitness consequences upon loss of one or more conditionally essential amino acid biosynthesis genes from bacterial genomes. Our computational analysis of 1,432 eubacterial genomes uncovered that in the vast majority of cases pairs of different auxotrophy-causing mutations co-occurred significantly more often than is expected by chance. Experimentally evaluating the fitness consequences resulting from introducing one, two, or three auxotrophy-causing mutations into the genome of *E. coli* in the presence of the required amino acids and in one of two carbon environments unravelled that (i) both positive and negative epistasis were prevalent among auxotrophy-causing genes, (ii) epistasis produced diminishing returns with increasing expected genotype fitness, and (iii) both the fitness of auxotrophic mutants and epistatic effects strongly depended on the carbon source available in the environment.

In our computational analysis, auxotrophy-causing genes showed a strong tendency to co-occur, which in most cases significantly exceeded what would be expected if mutations were randomly distributed (Fig. 1). Three main mechanisms may, independently or in combination, have contributed to this pattern: First, an increased co-occurrence of two amino acid auxotrophies could reflect the likelihood of the two corresponding amino acids to co-occur in the respective genotype's natural environment, thus favouring mutants that loose the corresponding biosynthesis genes. Indeed, the partial correlation observed between the co-occurrence of amino acids in nature and the co-occurrence of amino acid

auxotrophies in bacterial genomes (Fig. S 1) supports this scenario. Second, the probability of two amino acid biosynthesis genes to be simultaneously lost might be indicative of the fitness consequences arising upon loss of both genes in the corresponding bacterial strains. Third, amino acid biosynthesis genes that are localized in close spatial proximity on a bacterial chromosome might be simultaneously lost in large chromosomal deletion events, thus causing an increased co-occurrence of two amino acid auxotrophies.

In contrast, drift is unlikely to produce the observed co-occurrence pattern, because randomly fixing deletion alleles should rather display a frequency distribution that is not different from a random distribution. Since this was the case in only two of the 190 pairwise comparisons considered, drift is unlikely to be a major determinant of the observed co-occurrence pattern.

However, statistically evaluating the relationship between the experimentally determined relative fitness or epistasis of different *E. coli* mutants with the frequencies, with which auxotrophy-causing mutations have been predicted to co-occur in eubacterial genomes (Fig. 1), did not yield significant correlations in either case (Spearman's rank correlation: $P > 0.05$). The lack of a statistical relationship between these parameters could be due to one or a combination of several of the following factors. First, bacteria frequently lose large portions of their genome. A simultaneous loss of multiple biosynthetic genes could thus explain the mismatch between the distribution of auxotrophy-causing mutations and expectations based on their epistatic interactions. Similarly, other mutations in the genome that were not considered in the present study could interact with auxotrophy-causing mutations, thus affecting the fitness of multiply-auxotrophic genotypes. Second, the effective size of bacterial populations is likely to affect the probability with which bacteria loose biosynthetic genes. Genetic drift is more effective when population sizes are small as is the case for most endosymbiotic bacteria. Under these conditions even non-adaptive alleles can fix in the population. Third, environmental conditions that the analysed eubacterial strains experience in their natural environments were not evaluated in the current study, yet can affect the fitness of multiply-auxotrophic genotypes. Fourth, epistatic interactions identified for *E. coli* might not be representative for the taxonomic diversity of eubacterial genomes analysed (Fig. 1). Fifth, in our study, exclusively structural biosynthetic genes were deleted. However, in an amino acid-containing environment, natural selection might also favour mutations in regulatory elements, which could lead to the simultaneous deactivation of multiple biosynthetic pathways. Their subsequent loss from the mutants' genome would reflect regulatory relationships among groups of genes rather than epistatic interactions among multiple genes that were individually lost. Thus, future work is necessary to elucidate how auxotrophies evolve and to which extent epistatic interactions determine the mutational paths taken.

Previous work showed positive fitness effects generally accompany the loss of a conditionally essential biosynthetic gene when the focal metabolite is sufficiently available in the environment [24,85,86,181,194]. However, our study revealed that the sign and magnitude of fitness consequences can drastically change depending on the environment and the presence of additional auxotrophy-causing mutations. As such, our results strikingly matched theoretical predictions of a recent study, in which a flux-balance analysis of the metabolic network of *E. coli* identified strong effects of the carbon source used on epistatic interactions among deleted metabolic genes [116]. However, which mechanisms caused this functional relationship? One factor that could account for these observations, is that the loss of conditionally essential biosynthetic genes is likely to trigger a strong regulatory response that allows bacterial cells to survive despite amino acid starvation [195]. Most probably, these changes involve an up-regulation of amino acid transporters as well as a rerouting of metabolic fluxes through multiple pathways [195] leading to a globally restructured metabolism [196]. Since such systemic changes may be specifically tailored to compensate specific shortages associated with losing certain sets of genes, these regulatory differences could explain the observed plasticity and epistatic interactions among mutations. Second, our competition experiments were performed in relatively complex nutritional environments that, besides one of two carbon sources, also contained eleven different amino acids. Thus, epistatic effects could be caused by a competitive inhibition of amino acid uptake systems [197], competition of transporters for membrane space [198], or effects resulting from alterations of cell-internal amino acid pools [199]. Future work should examine these possibilities.

Interestingly, increasing the number of metabolic auxotrophies did not result in an additive increase of fitness effects caused by individual mutations, but on average mostly showed an overall decline or neutral effect in the succinate and fructose environment respectively (Fig. 2 A and B). This observation is consistent with previous experimental works [96,117,186,187] showing negative epistasis acts to diminish mutational effects. Finding this pattern also for auxotrophy-causing mutations suggests a common mechanism caused the beneficial fitness effects of different single gene deletion mutants. Intriguingly, a saving of protein expression costs has been previously suggested as a mechanistic cause for the fitness effects upon loss of conditionally essential biosynthetic genes from the genome of *E. coli* [181] as well as for the diminishing returns epistasis observed when four beneficial alleles were analysed in *Methylobacterium extorquens* AM1 [96]. In any case, if epistatic interactions determine the order in which auxotrophy-causing mutations are fixed in bacterial genomes, the current work provides several testable hypotheses that could be verified in a laboratory-based evolution experiment.

Finally, a particularly strong beneficial effect upon loss of the first metabolic gene may act as a spring-loaded mechanism that facilitates the establishment of metabolic cross-feeding interactions within microbial communities [23,24,51,200]; or aids the establishment of symbiotic associations between microbial symbionts and their host [201,202].

5. Conclusions and outlook

Our study provides first empirical insights into the selective consequences bacterial genotypes face when losing multiple auxotrophy-causing mutations. Especially the observed impact the ambient environment and the number of genes lost had on the fitness of auxotrophic genotypes implies a strong context-dependency of metabolic loss-of-function mutations that needs to be taken into account when such mutations are interpreted. Observing that epistasis produced diminishing returns with increasing expected genotype fitness points to a yet unknown molecular mechanism that constrains the fitness achievable by multiply auxotrophic genotypes. Identifying this mechanism will not only shed light on what causes the strong fitness benefits conferred by auxotrophy-causing mutations, but will also help to understand the molecular links that connect different biosynthetic genes. In particular laboratory-based evolution experiments, in which bacterial populations evolve under carefully controlled environmental conditions, provide a unique opportunity to identify which genes (e.g. regulatory versus structural genes) are prime targets of natural selection during the adaptive evolution of metabolic auxotrophies. Together with the approaches used in this study, such experiments would allow to further dissect how phenotypic plasticity and epistasis interactively guide the adaptive loss of biosynthetic functions.

Acknowledgements

We thank the Experimental Ecology and Evolution group, Arjan de Visser as well as three anonymous referees for constructive criticism, Wilhelm Boland for support, and Dominik Refardt for providing us with strain *E. coli* REL 607. We are especially grateful to David Berger, Vaughn Cooper, and Kenneth Flynn for advice on statistical analyses. This work was supported by the Jena School for Microbial Communication (JSMC), the Volkswagen Foundation, and the Max Planck Society.

Supporting Information

The following supporting information is available for this chapter on page 196

Figure S1. Correlation between the frequency of double auxotrophies and the pairwise abundance of amino acids in natural environments.

Figure S2. Frequency distribution of epistatic effects for 55 double- and 16 triple gene deletion mutants as determined in the fructose and the succinate environment.

Figure S3. Type II regression of observed and expected fitness as determined in the fructose and succinate -containing environment

Table S1. Strains used in this study.

Table S2. Fitness cost of the kanamycin resistance marker in the fructose-containing environment.

Table S3. Relative fitness and epistatic interactions among auxotrophy-causing mutations in the fructose-containing environment.

Table S4. Relative fitness and epistatic interactions among auxotrophy-causing mutations in the succinate-containing environment.

Chapter 3

Chapter 3

Metabolic network architecture and carbon source determine metabolite production costs

In major revision at The FEBS Journal (as of 14th December 2015)
Only minor typographical modifications have been made to the submitted manuscript in this chapter.

Authors: Silvio Waschina, Glen D'Souza, Christian Kost, and Christoph Kaleta

Abstract:

Metabolism is essential to organismal life, because it provides energy and building block metabolites. Even though it is known that the biosynthesis of metabolites consumes a significant proportion of the resources available to a cell, the factors that determine their production costs remain less well understood. In this context, it is especially unclear how the nutritional environment affects the costs of metabolite production. Here we use the amino acid metabolism of *Escherichia coli* as a model to show that the point at which a carbon source enters central metabolic pathways is a major determinant of individual metabolite production costs. Growth rates of auxotrophic genotypes, which in the presence of the required amino acid save biosynthetic costs, were compared to the growth rates that prototrophic cells achieved under the same conditions. The experimental results showed a strong concordance with computationally-estimated biosynthetic costs, which allowed us, for the first time, to systematically quantify carbon source-dependent metabolite production costs. Thus, we demonstrate that the nutritional environment in combination with network architecture is an important but hitherto underestimated factor influencing biosynthetic costs and thus microbial growth. Our observations are highly relevant for the optimization of biotechnological processes as well as for understanding the ecology of microorganisms in their natural environments.

1. Introduction

Most bacterial species are heterotrophic and thus derive their carbon from breaking down organic compounds [203]. The structural diversity of organic compounds bacteria encounter in their natural environments is remarkable and for several species it is known that they can utilize an extremely wide range of chemically different carbon sources [61]. *Escherichia coli,* for instance, is able to utilize more than 80 compounds as sole source of energy and carbon [204]. However, since carbon sources differ drastically in terms of their energy content as well as the molecular routes how a given bacterial cell can import and degrade the corresponding chemical, bacterial growth depends decisively on the nature of the carbon source used [205]. In this context, it has been proposed that biochemical constraints in the allocation of resources may limit the growth rate of bacterial cells [206,207]. In particular, such a pattern could be caused by the distribution of fluxes through the metabolic network to provide an optimal supply of building block metabolites (i.e. amino acids, nucleotides, and lipids) and growth factors (i.e. vitamins and co-factors) for cell growth.

Assuming that the architecture of a cell's metabolic network determines fluxes through the network, flux distributions should depend on the point at which a given carbon source enters the metabolic network. Indeed, it has been shown that the entry points of a given carbon source can cause considerably higher relative fluxes through reactions closer to the entry point than fluxes of reactions more distant to the entry point of the carbon source [208]. As a consequence, locally increased fluxes could also affect the biosynthetic costs of metabolites by locally increasing the amount of substrate available to a given biosynthetic reaction relative to all other growth-related functions. Thus, differences in flux distributions caused by different carbon sources should also translate into different biosynthetic costs of metabolites. Here we test this hypothesis by combining theoretical predictions with targeted experiments using amino acid biosynthesis of *E. coli* as a tractable model. Amino acid metabolism was chosen as a test case, because the biosynthesis of amino acids diverts an immense fraction of the total carbon source budget of a bacterial cell during growth [209]. It is therefore not surprising that bacterial species are under strong selective pressure to economize their amino acid usage [88,210]. We used a genome-scale model of the metabolic network of *E. coli* to estimate the biosynthetic cost for each of 20 proteinogenic amino acids depending on the utilized carbon source. Next, we validated these predictions by comparing the growth rates of genotypes auxotrophic for individual focal amino acids and the prototrophic wild type of *E. coli* grown on different carbon sources, while supplementing increasing concentrations of the focal amino acid to the growth environment. Under these conditions, auxotrophic genotypes increasingly saved the costs to biosynthesize the focal amino acid relative to the prototrophic wild type, and could thus invest the economized carbon source in other cell growth-related functions. By gradually relaxing the amino acid limitation for the growth of

auxotrophs in this way, and comparing their maximum growth rates relative to the growth rates achieved by prototrophic wild type cells, allowed quantifying the carbon source-dependent costs to produce individual amino acids.

Both theoretical predictions and experimental results revealed strong differences in the production costs of central metabolites in bacteria depending on the point at which the utilized carbon source enters the cell's metabolic network. The observed shifts of biosynthetic costs depending on the utilized carbon source are physiologically relevant and are caused by the structure of the underlying metabolic network.

2. Materials and methods

2.1 Prediction of biosynthetic costs

The biosynthetic costs of all 20 proteinogenic amino acids for 61 different carbon sources (see Table S3), which theoretically support the growth of *E. coli*, were estimated using flux balance analysis (FBA). Biosynthetic costs $p_{k,x}$ of an amino acid k were defined as the proportion of carbon source x, which is at least required to produce 1 mmol gDW^{-1} h^{-1} of the amino acid relative to the amount of carbon source x required to form 1 mmol gDW^{-1} h^{-1} biomass. The estimation incorporates the '*dual costs of amino acids*' [88]: (1) the resources required to generate energy in form of high-energy phosphor bonds (ATP and GTP) as well as the reducing power in form of NADH, NADPH, and $FADH_2$, which is consumed by enzymes of the biosynthetic pathway, and (2) the resources required to produce precursors for amino acid synthesis.

Two optimizations were performed within the FBA-framework using a genome-scale metabolic network reconstructions of *E. coli* K12 [211]: (1) $n_{k,x}$ was defined as the minimum amount of a carbon source x (in mmol gDW^{-1} h^{-1}; DW = dry weight) to produce one unit (i.e. 1 mmol gDW^{-1} h^{-1}) of an amino acid k. $n_{k,x}$ was determined by minimizing the influx of the carbon source x and fixing an outflow reaction of the amino acid k to a flux value of 1 mmol gDW^{-1} h^{-1}. (2) m_x was defined as the minimum amount of a carbon source x required to form 1 mmol gDW^{-1} h^{-1} biomass. m_x was calculated by constraining the flux through the biomass reaction (with 53.95 GAM estimate) of the metabolic model to a value equal 1 mmol gDW^{-1} h^{-1} and by minimizing the influx of the carbon source x. The optimizations were performed within Matlab 7.14 (Mathworks, USA) using the COBRA Toolbox version 2.0.5 [212] and the TOMLAB /CPLEX version 7.9 (TOMLAB Optimization, USA) as linear programming solver. The media elements used for the genome-scale model were Ca^{2+}, Cl^{-}, CO_2, Co^{2+}, Cu^{2+}, Fe^{2+}, Fe^{3+}, H^{+}, H_2O, K^{+}, Mg^{2+}, Mn^{2+}, molybdate, Na^{2+}, NH_4^{+}, Ni^{2+}, O_2, phosphate, SO_4, tungstate, and Zn^{2+}.

Finally, the biosynthetic cost estimations $p_{k,x}$ for all amino acid – carbon source combinations were calculated as

$$p_{k,x} = n_{k,x}/m_x.$$

2.2 Bacterial strains

Amino acid auxotrophic genotypes used in this study have been generated as described previously [181] (Table S2). The auxotrophic strains were derived from the *E. coli* BW25113 strain, which is the prototrophic wild type. Mutant strains were cured of the kanamycin resistance marker by excising the kanamycin cassette from the mutant's genome using pCP20 plasmid, which harbors the FLP recombinase [191]. For unknown reasons, it was not possible to cure the tyrosine auxotroph of the kanamycin resistance. Thus, the original kanamycin resistant mutant was used for growth kinetic assays instead. However, it has been previously demonstrated that this resistance marker does not incur detectable fitness effects under non-selective (i.e. antibiotic-free) conditions [213].

2.3 Culture conditions

All cultures were incubated under shaking conditions at 30 °C and grown in Minimal Media for *Azospirillum brasilense* (MMAB) [158] containing K_2HPO_4 (3 g L^{-1}), NaH_2PO_4 (1 g L^{-1}), NH_4Cl (1 g L^{-1}), $MgSO_4 \cdot 7H_2O$ (0.3 g L^{-1}), KCl (0.15 g L^{-1}), $CaCl_2 \cdot 2H_2O$ (0.01 g L^{-1}), $FeSO_4 \cdot 7H_2O$ (0.0025 g L^{-1}), $Na_2MoO_4 \cdot 2H_2O$ (0.05 g L^{-1}), and using different carbon sources. The concentrations of the carbon sources were 5 g L^{-1} D-fructose, 8.86 g L^{-1} disodium succinate, 8.61 g L^{-1} potassium L-lactate, 4.42 g L^{-1} glycerol, 8.17 g L^{-1} sodium pyruvate, 10.64 g L^{-1} disodium L-malate, 5 g L^{-1} D-maltose monohydrate or 5.06 g L^{-1} D-xylose. The concentrations of carbon sources were chosen such that – at least theoretically – the same amount of biomass could have been produced under all nutritional conditions. For this, we used the above-introduced value m_x, i.e. the minimum amount of carbon source x required to form one unit of biomass. The final concentration of carbon source x was calculated as $c_x = c_{Fru} \cdot m_x/m_{Fru}$ using 5 g L^{-1} fructose (c_{Fru} = 27.75 mM) as reference. This procedure is similar to the approach described by [214]), where concentrations were adjusted to match the number of reducible carbon atoms. However, using the genome-scale metabolic network of *E. coli* allows to take the physiological capabilities of the cell to transform a certain carbon source into biomass more precisely into account.

The eight carbon sources fructose, maltose, xylose, glycerol, pyruvate, lactate, succinate, and malate were chosen, because these substrates enter the central metabolic network of *E. coli* at different points (Fig. 1A) and the predicted biosynthetic costs of amino acids differed considerably between these carbon sources (Fig. 1B). Fructose, maltose, and glycerol are

catabolized via the Embden-Meyerhof-Parnas (EMP) Pathway. Xylose is converted to the pentose phosphate pathway intermediate D-xylulose 5-phosphate. L-lactate can be oxidized to pyruvate, a central metabolite, which links the glycolysis and the tricarboxylic acid (TCA) cycle. The carbon sources succinate and L-malate are intermediates of the TCA cycle.

2.4 Growth kinetic assays

The response of seven amino acid auxotrophic *E. coli* mutants and the prototrophic wild type strain in terms of the maximum growth rate to the supplementation of the focal amino acids was quantified in growth kinetic assays. For this, seven genotypes that were auxotrophic for one of the following amino acids were selected (deleted gene in brackets): histidine (*hisD*), tyrosine (*tyrA*), phenylalanine (*pheA*), tryptophan (*trpB*), leucine (*leuB*), lysine (*lysA*), and isoleucine (*ilvA*). These amino acids were chosen based on three criteria: (1) these amino acids cannot be catabolized and utilized as carbon source by *E. coli*. Tryptophan was the exception, which can be partially degraded to pyruvate and indole (indole cannot be further degraded) [215]. (2) No other cellular functions besides protein synthesis is known for these seven amino acids. For example, *E. coli* cannot degrade methionine, but can utilize it also as a precursor for S-adenosyl-L-methionine (SAM), the major methyl group donor in the cell. (3) No other reaction is known, with which *E. coli* can transform another metabolite into the focal amino acid [190]. The above criteria were applied to exclude unwanted confounding effects to influence the growth kinetics.

For each amino acid, six *E. coli* BW25113 wild type colonies and six colonies of the corresponding auxotrophic genotype were used to inoculate 1 ml overnight cultures (16 h) with fructose as carbon source. The media used to cultivate auxotrophic strains was supplemented with the amino acid the focal auxotroph required to grow (see Table S1 for exact amino acid concentrations). Each of these cultures was used to inoculate eight 1 ml pre-cultures (96-deep-well plates, Eppendorf, Germany), each containing one of the eight different carbon sources (i.e. fructose, maltose, xylose, glycerol, pyruvate, lactate, succinate, and malate) and the focal amino acid (see Table S1 for exact amino acid concentrations). Pre-cultures were incubated for 26 h at 30 °C under shaking conditions (220 rpm).

To test whether the maximum growth rates of prototrophic wild type cells was sensitive to increasing amino acid concentration in the growth medium, wild type precultures were used to inoculate 50 µl cultures in 384-well plates (flat bottom and transparent, Greiner Bio-One, Kremsmünster, Austria) with an initial cell density of 10^5 colony-forming units (CFUs) mL^{-1}. The MMAB medium used for these experiments contained the same carbon source as the preculture, yet in addition one of eight different concentrations of the focal amino acid, with the lowest level corresponding to no amino acid supplementation (see Table S1 for exact amino acid concentrations). In this way, each of the 64 combinations of eight carbon sources

and eight amino acid concentrations was independently replicated six times. A second 384-well plate with the exact same media layout was inoculated accordingly from the precultures of auxotrophic genotypes. Wells without amino acid supplementation were inoculated with the wild type strain as control. Growth kinetics were determined in a Tecan Infinite 200 Pro plate reader (Tecan Group, Männedorf, Switzerland) for automated kinetic measurements for 48 h at 30 °C and a 10 min kinetic cycle consisting of 7.5 minutes of orbital shaking (2 mm amplitude), 1 min waiting (no shaking), and 1.5 minutes for measuring the optical density at 600 nm (OD_{600nm}, 10 nm bandwidth) with 5 flashes.

2.5 Statistical data analysis

For each culture of the growth kinetic experiment, the maximum growth rate μ_{max} was determined. Since *E. coli* reaches substantially different maximum growth rates in the eight different carbon source regimes and to compare the increase of μ_{max} with increasing amino acid concentration, the μ_{max} values were normalized by the median of the maximum growth rates the wild type strain achieved under the same carbon condition without amino acid supplementation. Hereafter, we will refer to the normalized maximum growth rates as $\bar{\mu}_{max}$. The increase of the growth rate per μM of the focal amino acid (7 data points for the auxotrophic genotypes, 8 for the wild type strain) was calculated for each cognate population (i.e. populations which originated from the same clonal colony) and for each carbon source by linear regression [166,216].

For correlation analysis between increases of growth rates with either amino acid concentration or with predicted metabolic costs, a linear mixed-effects model was fitted considering the '*cognate population identity*' as random effect and the '*amino acid concentration*' or '*predicted metabolic costs*', respectively, as fixed effects. Models were fitted by maximizing the restricted log-likelihood until convergence. Conditional R^2 values of the fitted models were calculated according to [217].

Principle component analysis (PCA) and k-means clustering were performed to analyze the variance of biosynthetic costs of amino acids under various carbon sources. Only the main PCA axes, which together explained more than 90% of the observed variation, were used for k-means clustering. The algorithm by [218], for k-means clustering was applied starting with 25 random initial sets and optimization (minimizing within-group sum of squares) until convergence. P-values were corrected after multiple testing using the false discovery rate (FDR) procedure of [165]). All statistical analyses were using the *R* software (version 3.1.1) [166].

3. Results

3.1 Metabolic costs of amino acid depend on carbon source

To determine whether or not the metabolic costs to biosynthesize each of 20 proteinogenic amino acids depend on the carbon source used, amino acid production costs were computationally estimated for all 61 carbon sources, which are known to support the growth of *E. coli* K12 as sole source of carbon and energy [219] (Fig. 1). The mean costs of amino acids predicted in this way quantitatively matched previous predictions of energetic costs of amino acid biosynthesis (Pearson's product-moment correlation: $R = 0.96$, $N = 20$, $P < 0.0001$, supporting information Fig. S1, [88]). This correlation shows that the cost prediction method presented here is in line with previous estimations, but further enables to systematically assess metabolite production costs differences between various carbon substrates. To identify whether and to which extent the metabolic costs of a single amino acid were affected by the available carbon source, the metabolic cost estimated for all amino acid and 61 carbon sources were analyzed by principle component analysis (PCA) (Fig. 1C). Clustering for the correlation values of the observed amino acid costs with these three main principal components revealed three distinct groups: group 1 (blue) consisted of amino acids with precursors in glycolysis and/or the pentose phosphate pathway (i.e. Cys, Gly, His, Met, Phe, Ser, Trp, Tyr, and Ser), group 2 (green) contained pyruvate-derived amino acids (i.e. Ala, Leu, and Val), and group 3 (red) comprised amino acids with precursors from the tricarboxylic acid (TCA) cycle (i.e. Arg, Asn, Asp, Glu, Gln, Ile, Lys, Pro, and Thr) (Fig. 1A). The differences between groups reflect diverging biosynthetic costs associated to different classes of carbon sources: Amino acids of group 1 are metabolically cheaper to produce when glycolytic substrates (sugars/ sugar alcohols) are utilized as carbon source, yet more cost-intensive when only gluconeogenic substrates (e.g pyruvate, lactate, TCA-cycle intermediates) are available, and *vice versa* for the amino acids of group 2 and 3 (Tukey multiple comparisons of means: $P < 0.05$, for samples sizes see Fig. 1D). Consequently, there is a cost trade-off between the different groups of amino acids: reduced costs to produce amino acids in one group come at the expense of higher costs to synthesize amino acids of another group.

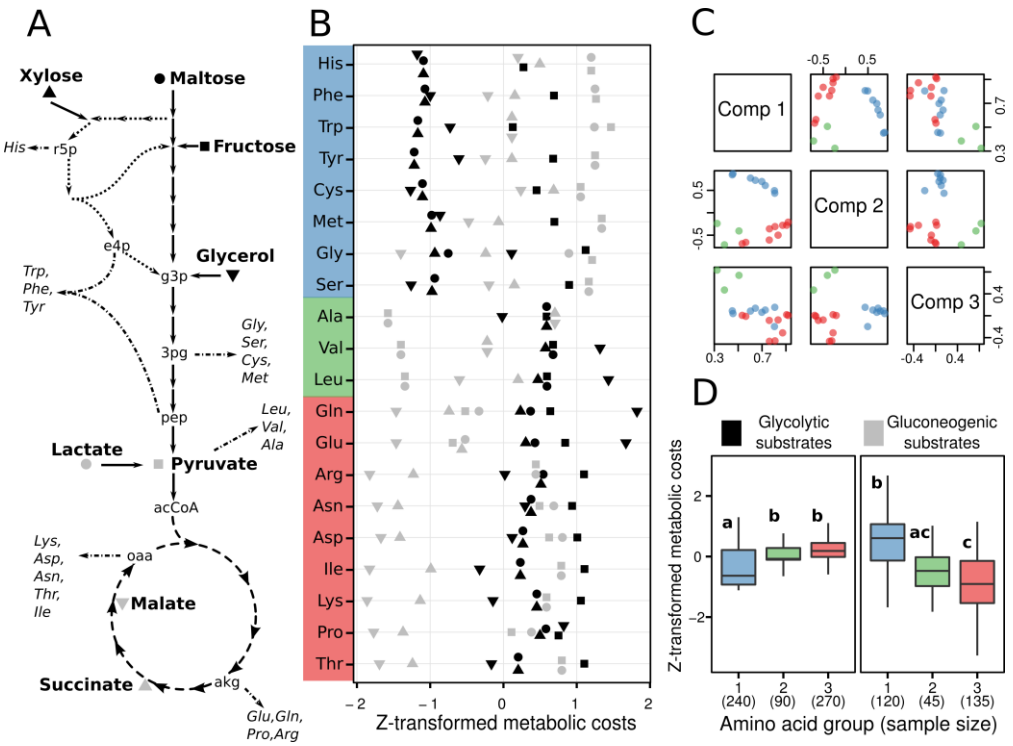


Figure 1. ***In silico*** **estimations of carbon source- and network structure-dependent metabolic costs of proteinogenic amino acids.** (A) Schematic representation of the central metabolism of *Escherichia coli* (glycolysis – solid arrows, TCA cycle – dashed arrows, pentose phosphate pathway – dotted arrows). Carbon sources used in this study are shown in boldface, amino acids in italics. (B) Estimated metabolic costs of amino acids for 8 carbon sources including 4 organic acids (grey) and 4 sugars/ sugar-alcohol (black). For a better visualization, metabolic costs of each amino acid were z-transformed (same range of values). (C) Principle component analysis (PCA) of estimated metabolic costs of amino acids based on 61 carbon sources. Shown are the correlations of the metabolic costs of each amino acid with the three main PCA components (Comp 1-3), which together explain >91% of the observed variation. Data points are colored according to k-means clustering with three centers: (group 1, blue): Cys, Gly, His, Met, Phe, Ser, Trp, and Tyr; (group 2, green): Ala, Leu, and Val; (group 3, red): Asn, Asp, Arg, Gln, Glu, Ile, Lys, Pro, and Thr. (D) Estimated (z-transformed) metabolic costs of amino acids for glycolytic- and gluconeogenic carbon sources. Amino acids are grouped according to the k-means clustering in (C). Different letters denote significant differences (Tukey multiple comparisons of means: $P < 0.05$, numbers below amino acid groups denote sample sizes).

Furthermore, the quantitative impact of different types of carbon sources on the absolute metabolic costs of amino acids varied among amino acids (Fig. 2). The highest variability of metabolic costs were observed for the leucine and glutamate, whose biosynthetic costs varied up to 13% from the mean metabolic costs based on the estimations assuming 61 different types of carbon sources. The lowest variability of 6% was observed for alanine.

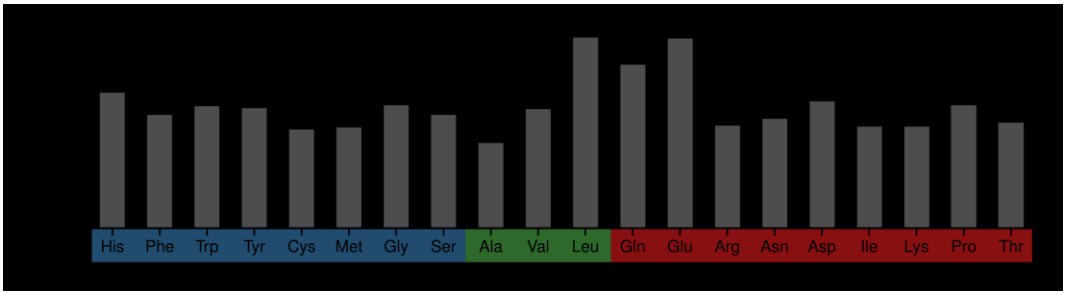


Figure 2. Variability of amino acid metabolic costs. The variability of metabolic costs was calculated as the 95% confidence interval size divided by the mean metabolic cost of the respective amino acid based on the costs estimations assuming 61 different types of carbon sources.

3.2 The response of auxotrophs to amino acid supplementation depends on the carbon source

The maximum growth rates of all seven auxotrophs were tested under amino acid supplementation and eight different carbon source conditions. Auxotrophic strains were chosen to study the effect of amino acid supplementation and to ensure that the cells actually save the biosynthetic costs to produce the focal amino acids and use amino acids from the media. The seven auxotrophies were chosen for the experiments, because no other cellular function than the incorporation into proteins have been described for the respective amino acids. Other effects, than the saving of metabolic costs, of the amino acid supplementation on the growth of the *E. coli* auxotrophs could therefore be prevented.

The maximum growth rate of all seven auxotrophs increased significantly with increasing amino acid supplementation (FDR-corrected linear mixed-model fit by maximizing the restricted log-likelihood: $P < 0.05$, n=42, Fig. 3 and Fig. S2). The only exception to the otherwise consistent pattern was the case of the isoleucine auxotroph using succinate as carbon source (FDR-corrected linear mixed-model fit by maximizing the restricted log-likelihood, $P = 0.42$, $n = 42$). In contrast, the maximum growth rates of populations of prototrophic *E. coli* wild type cells did not respond significantly to increasing amino acid concentrations in 32 out of 56 amino acid-carbon source combinations analyzed (Fig. S3). In 22 cases, the maximum growth rate increased significantly with amino acid supplementation, in two cases (i.e. histidine and xylose/ succinate) it even decreased significantly with increasing amino acid concentrations (FDR-corrected linear mixed-model fit by maximizing the restricted log-likelihood: $P < 0.05$, n=42, Fig. S3).

In virtually none of the cases examined did the auxotrophic genotypes reach maximal growth levels of WT populations (Figs. 3 and S2), indicating that under the focal conditions growth of auxotrophic genotypes was mainly limited by the availability of the required amino acid. After normalizing the auxotroph's growth rate by the growth rate the prototrophic WT strain had achieved under the same carbon source condition without amino acid supplementation (in the following, $\bar{\mu}_{max}$ refers to the normalized growth rate), it became clear that the increase

of the relative maximum growth rate $\bar{\mu}_{max}$ strongly depended on the carbon source provided for growth (FDR-corrected repeated measures ANOVA: all $P < 0.001$, $df_{carbon\ sources} = 7$, $df_{error} = 35$, Figs. 3 and S2). For example, the tryptophan auxotroph responded to tryptophan supplementation with an increase of 4.4 $\bar{\mu}_{max}$ (mM Trp)$^{-1}$ (mean) with fructose as carbon source, whereas with 10 $\bar{\mu}_{max}$ (mM Trp)$^{-1}$ the increase was significantly higher when utilizing pyruvate (paired t-test: $P < 0.001$, $n = 6$, Fig. 3).

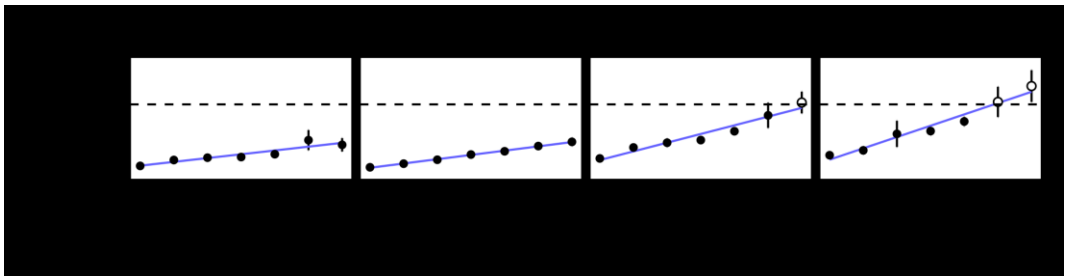


Figure 3. Carbon source-dependent growth rate response of the tryptophan auxotrophic genotype to increasing tryptophan supplementation. Shown are the mean maximum growth rates (± 95% confidence interval) of the tryptophan (Trp) auxotrophic genotype relative ($\bar{\mu}_{max}$) to the prototrophic wild type (=1, dashed line) in four carbon source regimes and seven different Trp concentrations. Filled circles indicate growth rates of the auxotrophs, which are significantly lower than the maximum growth rate of the prototrophic wild type under the same carbon source conditions without Trp supplementation (FDR-corrected Welch two sample t-tests: $P < 0.05$, $n = 6$). Unfilled circles denote no statistical difference. This figure is representative for the complete data set shown in Fig. S2.

Taken together, the growth-kinetic assays revealed a strong effect of the carbon source on the growth physiology of the seven amino acid auxotrophic strains tested when the availability of the required amino acid was limiting growth.

3.3 Biosynthetic costs can explain the growth rate increase upon amino acid supplementation

A significant positive correlation between estimated biosynthetic costs and growth rate increases was observed for five of the seven amino acids tested: histidine ($P < 0.05$), tryptophan ($P < 0.01$), leucine ($P < 0.001$), lysine ($P < 0.05$), and isoleucine ($P < 0.001$, FDR-corrected linear mixed-model fit by maximizing the restricted log-likelihood: n=48, Fig. 4). These five amino acids represent all three main groups identified in the above-mentioned *in silico* analysis of the costs to biosynthesize the 20 proteinogenic amino acids (Fig. 1C). In other words, the same metabolic trade-offs in the efficiencies to synthesize amino acids that were theoretically predicted (Fig. 1D) were also found experimentally (Fig. 4).

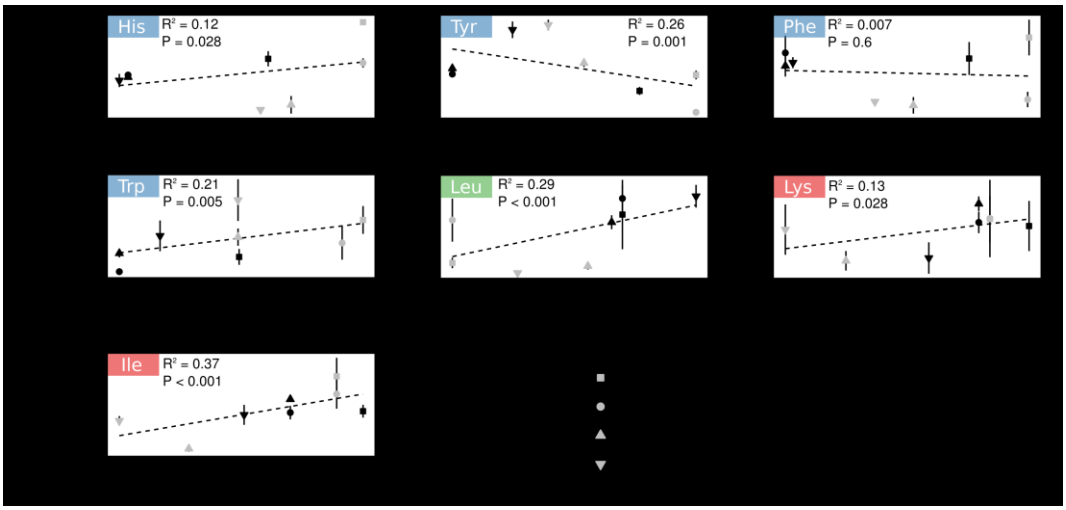


Figure 4. Correlation of predicted- and measured biosynthetic costs. The response in growth rate of auxotrophic genotypes to amino acid supplementation can be explained by the metabolic network structure. Shown are the correlations of predicted metabolic costs $p_{k,x}$ (x-axes) for amino acid k and carbon source x and the experimentally-determined increase of the relative growth rates $\bar{\mu}_{max}$ of auxotrophs with increasing amino acid concentration (Y). Mean values ± 95% confidence intervals are shown. See Table S1 for amino acid abbreviations.

A significantly negative correlation was observed between the predicted biosynthetic costs and growth rate increases for tyrosine (FDR-corrected linear mixed-model fit by maximizing the restricted log-likelihood: P = 0.001, n=48, Fig. 4), while no statistical relationship between these two parameters could be detected for phenylalanine (FDR-corrected linear mixed-model fit by maximizing the restricted log-likelihood: P = 0.6, n=48, Fig. 4).

4. Discussion

Microorganisms invest a significant proportion of their available carbon resources in the biosynthesis of metabolites. The amount of carbon source a cell needs to produce individual metabolites (i.e. biosynthetic costs) can be estimated based on the organisms' genome sequence and information on the nutritional composition of the natural habitat [220]. However, natural environments can fluctuate widely in the availability of different resources [221,222] and many microorganisms are able to utilize a broad range of different carbon sources [61]. Two main questions arise from these facts: i) How are metabolite production costs affected by the nutritional environment?, and ii) How variable are these costs within an organism? Here, we tested for the first time whether the variability of biosynthetic costs within a given organism can be explained by the carbon source used. The main findings of this study are that the structure of the metabolic network determines biosynthetic costs and

that these costs are variable depending on (1) the position of the precursor metabolites within the metabolic network, and (2) the point at which the carbon source enters central metabolism (Fig. 1D).

A genome-scale metabolic network of *E. coli* was employed to predict differences in amino acid production costs depending on the nutritional environment. To test the *in silico* cost estimations, growth kinetic assays of amino acid auxotrophic *E. coli* strains and the prototrophic wild type were performed for eight different carbon sources and seven amino acids in increasing concentrations. By comparing the maximum growth rates achieved by auxotrophic and prototrophic genotypes under specific conditions, it was possible to experimentally determine the biosynthetic costs, which auxotrophic genotypes saved by not having to synthetize the respective amino acid autonomously. The experimental measures derived in this way matched theoretical predictions of the carbon source-dependent biosynthetic costs for five of the seven amino acids tested: histidine, isoleucine, leucine, lysine, and tryptophan (Fig. 4). A discrepancy between cost prediction and experimental approximation was observed only for phenylalanine and tyrosine. The biosynthetic pathways for these two amino acids are closely connected: both amino acids originate from the common precursor chorismate and the biosynthetic pathways consist both of three reactions where only the second step is catalyzed by distinct enzymes [223]. Furthermore, both pathways are tightly co-regulated [224]. This interconnection of both pathways might cause additional effects besides the focal biosynthetic costs, when only one of the two amino acids is supplemented to the media.

To avoid confounding factors affecting the growth kinetics that are independent of biosynthetic costs, we focused our analysis on amino acids, which cannot be degraded and, hence, cannot be utilized as an alternative carbon source. Also, by using auxotrophic genotypes that cannot convert any other metabolite into the focal amino acid [190], it was possible to directly and precisely control the amount of the focal amino acid that was available to the cells. Taken together, our study provides, for the first time, a comparison of the growth response of *E. coli* to amino acid supplementation with a metabolic model using the flux balance analysis framework. The detected cost differences between carbon sources strongly influenced bacterial growth and thus significantly affected bacterial fitness.

To understand the evolution of a microbial metabolic network requires knowledge on the factors that determine biosynthetic costs within a given organism. The results presented in this work provide first evidence that metabolite production costs are affected by environmental factors such as the available carbon source. Most notable differences were found for the amino acids leucine, glutamate, and glutamine, whose costs varied by up to 13% between carbon sources. The observation of carbon source-dependent metabolic costs of amino acids is in line with recent findings that gene deletion mutations, which lead to the

loss of biosynthetic functions, can have different fitness effects depending on which carbon source is provided for growth [213]. In addition, it has been shown that synthetically generated amino acid-, nucleotide- , and vitamin auxotrophic mutants of *E. coli* had a significant fitness advantage over their prototrophic ancestor in environments where the respective metabolite was sufficiently present – even when both strains directly competed against each other [181]. These fitness benefits are likely to be due to the biosynthetic costs, which the auxotrophs save by not having to synthesize the respective metabolite [225].

Another interesting outcome of our study was, that the comparison of the biosynthetic costs of all 20 proteinogenic amino acids for 61 different carbon sources pointed to a metabolic cost trade-off between the efficiencies to produce different classes of amino acids (Fig. 1D). Thus, amino acids that are less costly to produce utilizing one specific carbon source (e.g. amino acids derived from TCA cycle intermediates) relative to another carbon source come at the expense of higher costs for other amino acids (e.g. derived from glycolysis intermediates). Biochemical trade-offs are thought to play a key role for metabolic specialization [226]. Hence, our results provide a plausible explanation for the evolution and maintenance of metabolic cross-feeding interactions where subpopulations, which specialized on preferentially performing certain metabolic functions, share the products of these functions [24]. Based on our results, metabolite cross-feeding could be especially promoted in environments, where multiple carbon sources are simultaneously present and subpopulation have specialized on utilizing distinct carbon sources. Sympatric specialization to utilize different carbon sources has been observed in laboratory evolution experiments of *E. coli* [227,228]. In a prominent example of a long term evolution experiment, in which *E. coli* was serially propagated in glucose minimal media, an adaptive mutation emerged in one population after 31,500 generations, through which the newly evolved variants acquired the ability to utilize citrate as carbon source, which has been included as part of the media formulation [228]. In another long-term continuous culture of *E. coli,* where glucose was provided as sole carbon source, two subpopulations evolved: one, which utilized glucose and produced acetate as a metabolic by-product and a second subpopulation, which specialized to utilize the exogenously available acetate [227]. Consequently, the utilization of different carbon sources can cause significant differences in the distribution of metabolic fluxes [208,229,230] and, as shown in this study, different biosynthetic costs. Interestingly, the amino acid biosynthetic cost differences between the two specialized subpopulations in the two above mentioned examples are highly reciprocal, because glucose is a glycolytic carbon source, whereas citrate, or acetate, respectively are gluconeogenic substrates (Fig. 1D). These differences in turn could favor the evolution of amino acid cross-feeding, where each specialized subpopulation can receive mutual benefits by saving biosynthetic costs.

Our results are not only relevant to understand adaptive processes of bacteria that are exposed to different nutritional environments, but have also implications for more applied contexts, for example the optimization of biotechnological processes where microorganism are used to produce value-added compounds such as biofuels, amino acids, or recombinant proteins. Metabolic engineering uses recombinant DNA techniques to modify the structure of metabolic networks by introducing new biosynthetic capabilities to the cell or improving the production rate of a specific molecule [231]. Another way to optimize production rates of desired metabolites is to rationally design the nutritional environment that is used as culture media [232]. Based on the presented results, it will be possible to increase the yield of a desired compound by rationally choosing a carbon source that minimizes production costs of the focal metabolite. Thus, a better understanding of the (environmental) factors that determine the production costs of desired compounds can significantly improve biotechnological production processes.

All growing cells allocate resources to different biosynthetic pathways in response to the nutritional environment. The resource costs associated with the biosynthesis of metabolites strongly affect the fitness of a species. In this study, the interplay between the chemical nature of a carbon source and its conversion into cell constituents was systematically assessed. The presented results unravel the link between a cell's nutritional environment and the architecture of its metabolic network as a key determinant of biosynthetic costs and microbial growth. As the structure of a metabolic network has evolved in response to natural selection, the here observed variability of biosynthetic costs depending on the available carbon source is indicative of the crucial role of the environmental context for the evolution of biochemical networks and the ecology of microorganisms. Future work is necessary to extent the economical concept of metabolic costs in more natural settings where multiple microbial species with diverse metabolic capabilities coexist and where several different substrates are available for cell growth.

Acknowledgements

We thank Holger Merker, Gabriele Diekert, and Wilhelm Boland for constructive criticism and support. This work was supported by the Jena School for Microbial Communication (JSMC), the Excellence Cluster *Inflammation at Interfaces*, the Volkswagen Foundation, and the Max Planck Society.

Supporting Information

The following supporting information is available for this chapter on page 209

Figure S1. Biosynthetic cost estimations from this study are in line with previously reported estimations.

Figure S2. Maximum growth rates of auxotrophs under various carbon sources and amino acid concentrations relative to the maximum growth rate level of the wild type growing under the same carbon source and without amino acid supplementation

Figure S3. Maximum growth rates of the *E. coli* wild type strain under various carbon sources and amino acid concentrations relative to the maximum growth rate level of the wild type growing under the same carbon source and without amino acid supplementation

Table S1. Final amino acid concentrations (μM) in the media used for the precultures of auxotrophs and for the growth kinetic assays.

Table S2. Strains used in this study.

Table S3. Carbon sources considered for biosynthetic cost estimation.

Chapter 4

Chapter 4

Experimental evolution of metabolic dependency in bacteria

Manuscript to be submitted to PLoS Biology

Authors: Glen D'Souza and Christian Kost

Abstract

Bacteria that adapt to nutrient-containing environments frequently lose biosynthetic genes, thus making them dependent on an environmental uptake of the corresponding metabolite. Despite the ubiquity of this '*genome streamlining*', it is generally unclear whether the concomitant loss of biosynthetic functions is favored by natural selection or rather caused by random genetic drift. Here we demonstrate experimentally that a loss of metabolic functions is strongly selected for when the corresponding metabolites can be derived from the environment. Serially propagating replicate populations of the bacterium *Escherichia coli* in amino acid-containing environments revealed that auxotrophic genotypes rapidly evolved in less than 2,000 generations in almost all populations. Moreover, auxotrophs also evolved in environments lacking amino acids – yet to a much lesser extent. Loss of these biosynthetic functions was adaptive and due to mutations in both *structural* and *regulatory genes*. Interestingly, auxotrophic mutants derived amino acids not only via an environmental uptake, but also by cross-feeding from coexisting prototrophs. Our results provide strong quantitative evidence that adaptive fitness benefits favor loss-of-function mutants when the corresponding metabolite can be obtained from the biotic and abiotic environment, which can drive the establishment of metabolic interactions within microbial communities.

1. Introduction

Bacterial genomes are highly dynamic in terms of both size and composition [36]. The extensive variation in gene repertoires that characterizes prokaryotic genomes can be caused by genome expansion via horizontal gene transfer and gene duplication or, alternatively, contraction due to gene loss. Interestingly, comparative analyses have provided evidence that gene loss may in fact be quantitatively more important for determining the size of prokaryotic genomes than the gain of new genetic information [36,42,46]. Indeed, as sequencing technologies improve, more and more microorganisms are being discovered that feature tremendously small genomes [73]; some of which are even smaller than the suggested minimal genome size for cellular life of ~300 kb [49]. Analyzing the genetic content of these reduced genomes revealed - besides a lack of dispensable elements such as extrachromosomal replicons or redundant genes [48] - also the elimination of seemingly essential biosynthetic functions [181]. For example, reconstructing metabolic networks from sequence data to predict the phenotype of the focal organism unraveled that the majority of bacterial genomes analyzed lacked the biosynthetic capability to produce several essential building block metabolites such as amino acids, vitamins, or even nucleobases [22,67,181,213]. Surprisingly, the list of genotypes that cannot produce certain metabolites autonomously (hereafter: *auxotrophic genotypes*) does not only include host-associated bacteria such as pathogens [174] or endosymbionts [32,47,69], which potentially obtain the required metabolites from the host cytoplasm, but also free-living bacteria such as *Prochlorococcus* and *Pelagibacter* [183,233] that are known to mainly inhabit nutrient-poor environments. The ubiquity of biosynthetic loss-of-function mutations in bacteria that inhabit ecologically disparate environments begs an explanation: Which evolutionary mechanisms have favored a loss of biosynthetic genes over metabolic autonomy in these bacteria?

Two main hypotheses have been put forward to explain these striking observations. First, genetic drift has been suggested to drive gene loss in bacteria which are in obligate associations with eukaryotic hosts and usually experience nutrient-rich or constant environments [234]. These bacteria undergo frequent reductions in population sizes due to periodic bottlenecks, and - as a consequence thereof -render the effects of drift to be stronger than those of selection [46]. Furthermore, a lack or reduced frequency of recombination may accelerate the fixation of non-beneficial or deleterious mutations to fix in the population [47]. This hypothesis is supported by comparative genomic datasets and experimental evidence for this idea comes from an experimental evolution experiment where bacterial populations were subjected to periodic single-cell bottlenecks resulting in the bacteria with strikingly reduced genomes [76].

The second main hypothesis that has been proposed to explain the apparent '*genome streamlining*', especially in bacteria with free-living lifestyles and those that inhabit low nutrient

environments, is natural selection. The large population effective sizes in such environments would render the benefits of selection to be strong and thus cause fitness increasing deletions to fix in the populations [46]. Thus, gene loss in environments where the gene is no longer required may benefit bacteria with genomes of a reduced size over conspecifics with a larger genome - primarily due to the adaptive benefits stemming from a more optimized cellular economization [82]. An increased cellular economy can result from the rerouting of resources from redundant functions, which are supplemented from the environment, to other cellular processes. This hypothesis has received experimental support from various studies where deletions were found to be selectively beneficial in evolving bacterial populations [92,235]. Also, previous studies that compared the growth of synthetically engineered auxotrophic bacterial strains to prototrophic ones, demonstrated that metabolic auxotrophies can be beneficial when the required biosynthetic product is sufficiently available in the environment [85,181,213]. Although these studies highlight that selection can account for losses of biosynthetic functions, it is unclear whether adaptive mutants which have lost biosynthetic functions can rapidly evolve and spread to detectable frequencies in populations in rich environments. Therefore, direct experimental evidence implicating natural selection in the loss of biosynthetic functions and the subsequent metabolic specialization when metabolites are present in the environment is lacking.

However, based on previous experimental studies [85,181] it can be hypothesized that adaptive benefits can result in bacterial populations losing genes and concomitantly the encoded biosynthetic functions, in rich environments, where the requisite metabolite is abundant. Altogether, insights from genomic and experimental studies strongly favor the argument that selection can be an especially strong force in driving the evolution of metabolic auxotrophies. To unravel whether fitness advantages can indeed drive the loss of biosynthetic functions in nutrient-containing environments, we experimentally evolved replicate populations of the prototrophic bacterium *Escherichia coli* in amino acid-replete (hereafter: AA-regime) or -deficient environments (hereafter: non-AA regime) We observed that auxotrophic mutants, which required multiple amino acids rapidly evolved in multiple replicate populations propagated in presence of amino acids. Surprisingly, auxotrophic mutants also evolved in the non-amino acid environment, albeit at lower frequencies than the amino acid-environment. Loss of biosynthetic functions in the auxotrophic genotypes was adaptive only in the presence of amino acids in the environment. Interestingly, auxotrophic genotypes also derived amino acids from coexisting prototrophic strains and this interaction was maintained by negative frequency-dependent selection. A genomic analysis of the evolved auxotrophic genotypes revealed that distinct genetic changes in both *structural* and *regulatory* genes potentially resulted in the adaptive loss of biosynthetic functions.

2. Material and methods

2.1 Strains, media, and growth conditions

The isogenic ancestor of the evolution experiment was *Escherichia coli* BW25113 Ara$^-$ or Ara$^+$. The Ara+ phenotypic marker was used to discriminate the strains for future experiments as described previously [181]. Unless stated otherwise, liquid minimal medium for *Azospirillum brasillense* (MMAB) [158] with 0.5 % fructose as the carbon source and without biotin was used for cultivating bacteria in all experiments. For some experiments, all 20 amino acids (+AA regime) were supplemented to the MMAB medium - each at a final concentration of 100 μM. When solid MMAB medium was used, 1.5 % agar was supplemented to the liquid minimal medium. For all experimental assays, strains and populations were first pre-cultured in the conditions they evolved in before inoculating the experimental assay (a 1:100 dilution). All pre-cultures and growth assays were performed at 30 °C with shaking at 220 rpm for 18 or 24 hours. All experiments involving monocultures were initiated using ~10^5 cells ml^{-1} of the focal strain or population and co-cultures were inoculated with ~10^5 cells ml^{-1} of each strain or population.

2.2 Evolution experiment

Eight independent lineages were founded from four isogenic clones of either *E. coli* BW25113 (Ara$^-$) [149] or *E. coli* BW25113 (Ara$^+$) [181]. The two pairs of lineages differed only with respect to the presence of the *araA* gene [181,213]. Its presence or absence allows to phenotypically discriminate the two strains from their evolutionary ancestor (red and white differentiation) [162]. The eight lineages were serially propagated by daily transfers into fresh 1 ml MMAB medium for 2,000 generations containing a mixture of all of the 20 essential amino acids. This treatment represented the AA regime of the experiment (i.e. with amino acids). Similarly, eight additional lineages, which descended from the same ancestral colonies as the lineages mentioned above, were propagated for 2,000 generations by daily transfers into MMAB medium that did not contain any amino acid. Each day, 1 μl of culture medium was transferred to 999 μl of fresh medium (a 1:1,000 dilution). The transfer volume contained ~10^3 cells and the population expanded to ~10^6 cells after 24 hours of growth, resulting in ~10 generations ever day (generations = [(log number of cells after 24 h) - (log number of cells initially present)]/ 0.301). Evolving populations were frozen at -80 °C at 250, 500, 1,000, 1,500, and 2,000 generations along with 50% glycerol for subsequent experimental assays.

2.3 Measurement of growth parameters

Productivity of the evolving populations was measured by determining the number of CFUs ml^{-1} on Lysogeny Broth (LB) agar plates at 15; 180; 450; 750; 870; 1,170; 1,365; 1,590; 1,695; 1,920; and 1,995 generations. Growth kinetic parameters like the maximal growth rate (μ_{max} h^{-1}) and the maximum optical density reached (OD_{max}) were determined for populations and clones isolated from different time points over the course of the evolution experiment. For this, frozen samples were revived by inoculating 10 µl into 1 ml of the corresponding medium as described above. Growth assays were performed in 50 µl of medium in a 384 micro-well plate (Greiner, Germany) and growth kinetics were monitored in a Tecan Infinite Pro Microplate reader (Tecan, Austria) by recording the OD every eight minutes for 24 h at 30 °C with shaking at 2.5 Hz in the interim. Differences in kinetics between different growth conditions for a specific population were determined by calculating the $\Delta\mu_{max}$ h^{-1} (μ_{max} $h^{-1}{}_{-AA}$ - µMax $h^{-1}{}_{+AA}$).

2.4 Determination of auxotrophies

The appearance of auxotrophic mutants in the evolving populations was determined by resuscitating and pre-culturing frozen samples from generations 250; 500; 1,000; 1,500; and 2,000 of both selection regimes. These cultures were then serially diluted such that each population contained ~10^3 CFUs ml^{-1} and plated on MMAB agar plates that contained all amino acids. 1,000 colonies from each population of the two regimes were then inoculated onto a new MMAB agar plate that contained all AAs. The colonies were then replicated using a 96-pin replicator on MMAB agar plates without any AAs to identify colonies that will be unable to grow. Any colonies that failed to grow on MMAB without AAs were deemed as auxotrophs. These colonies were then selected and replica-plated on MMAB plates without AAs and 20 different MMAB 'dropout' media, each containing a different combination of 19 AAs, leaving out one specific AA (Table S1). This approach allowed determining for which specific amino acids the focal strains were auxotrophic for. Any strain that did not grow on un-supplemented MMAB media, but grew on MMAB medium containing all AAs or one or more of the media that were lacking one AA was scored as auxotroph. Strains that were unable to grow on MMAB without AA supplementation yet could grow on all 20 dropout media were deemed 'unassigned'.

2.5 Competitive fitness assays

The fitness of each auxotrophic or prototrophic type that has been isolated after 1,000; 1,500; and 2,000 generations of growth in the AA-regime and after 500 and 1,500 generations in the non-AA regime was determined in competition experiments against the evolutionary

ancestor. For this, ~10^5 cells of the derived clone as well as of the evolutionary ancestor that was carrying the respective other Arabinose utilization marker (i.e. Ara^+ versus Ara^- and *vice versa*) were pre-cultured and subsequently co-inoculated into 1 ml of MMAB with or without AA. The relative fitness of strains from the AA-regime was determined in the absence or presence of AAs, while the strains from the non-AA regime were only analyzed in the absence of AAs. The number of CFUs ml^{-1} was determined at the start (i.e. 0 h) and the end of the co-culture period (i.e. after 24 h) by plating on TA agar with arabinose and TTC [162]. After that the Malthusian fitness parameter was calculated as described previously [162]. Each strain type in the experiment was replicated 8-times.

2.6 Metabolic dependency

To determine to which extent the growth of derived auxotrophs depended on the availability of AAs in the environment and/or the presence of a co-cultured prototroph, auxotrophic strains that have been isolated from the AA-regime after 2,000 generations or after 500 and 1,500 generations from the non-AA regime were pre-cultured in AA-supplemented MMAB medium. The corresponding prototrophs (which have been isolated from the same population and the same time points as the auxotroph) or the evolutionary ancestor were similarly pre-cultured in MMAB medium with or without AA-supplementation, depending on the medium of the main experiment. ~10^5 cells of the focal auxotrophic mutant were inoculated into 1 ml MMAB with or without AAs. The same experimental set-up was repeated three times: (i) with the auxotrophs grown in monoculture, or by co-inoculating ~10^5 cells per ml of (ii) a co-evolved prototrophic strain that has been isolated from the same replicate and time-point, or (iii) the evolutionary ancestor. Each experimental treatment was replicated 4-times. All of these cultures were incubated for 24 h. The number of CFUs ml^{-1} was determined for each strain at the start (0 h) and the end of the experiment (24 h). Since the derived auxotrophic and prototrophic types in each co-culture carried the same Ara marker, because they descend from the same ancestor, both types were distinguished by plating on an un-supplemented MMAB plate and LB plates. Thus, the number of CFUs ml^{-1} of prototrophic or ancestral strains (i.e. CFUs on the MMAB plate) and CFUs ml^{-1} of the auxotroph (i.e. CFUs on the LB plate minus CFUs on the MMAB plate) were determined and the Malthusian parameter of the auxotrophic strain calculated as previously [162].

2.7 Invasion-from-rare experiment

Cultures of auxotrophs and the corresponding prototrophs that have been isolated after 2,000 generations of evolution in the AA-regime or after 500 and 1,500 generations of growth in the non-AA regime were prepared. Subsequently, co-cultures between both

partners were inoculated together at different initial frequencies (i.e. 1:100 or 100:1), such that the initial cell density of the co-culture was 10^5 CFUs ml^{-1}. Each experimental treatment was replicated 4-times. The prototrophic or ancestral strains that derived from the same population were plated and distinguished as described above at the onset (0 h) and after the experiment (24 h). The number of CFUs ml^{-1} was determined for both types and the selection coefficients of the invading (rare) type was calculated as described previously [236].

2.8 Whole-genome sequencing

One representative auxotrophic was selected from six populations of the AA-supplemented environment where these types were detected and one prototroph of 4 of the same populations was selected for genome sequencing (10 in total). In addition, 2 auxotrophs from the 2 populations that evolved under un-supplemented conditions and 2 prototrophic isolates from this regime (i.e those which coevolved with auxotrophs) were also selected for sequencing (4 in total). Genomic DNA was extracted after these strains had been grown in LB medium for 24 h using the Epicentre MasterPure DNA extraction kit (Biozym Scientific, Germany). Quality control and library preparation (TruSeq, Illumina) was performed by the Max Planck Genome Centre Cologne, Germany (http://mpgc.mpipz.mpg.de/home/) and sequencing was performed on the Ilumina HiSeq2500 platform. The resulting raw Illumina sequences were aligned to the published reference genome of *Escherichia coli* BW25113 (CP009273_1) [237] using the breseq pipeline [153,238] and mutations were thus identified.

2.9 Statistical analysis

Normal distribution of data was assessed using the Kolmogorov-Smirnov test. Homogeneity of variances was determined by applying Levene's test and variances were considered to be homogeneous when $P>0.05$. Independent sample t-tests were used to compare fitness and growth rates of populations evolved under the two regimes as well as fitness of auxotrophic or prototrophic strains relative to the evolutionary ancestor. Paired samples t-tests were used to compare growth rates of AA-evolved populations in the presence or absence of amino acids. The productivity of populations over evolutionary time was fitted using exponential fits and the slopes of the fitted lines were compared using independent sample t-tests [239]. One-way ANOVAs followed by LSD post-hoc tests were used to distinguish the Malthusian fitness in the metabolic dependence experiments and the growth kinetic parameters of different genotypes in the growth experiments. One sample t-tests were used to determine if selection coefficients of the auxotrophic genotypes were significantly different from 0.

3. Results

3.1 The metabolic environment influences adaptive capabilities of the evolving populations

To determine how the metabolic environment can affect the evolution of adapting bacterial populations, 8 replicate populations of *Escherichia coli* were serially propagated by daily transfer of 1,000 cells in minimal medium that did or did not contain all of 20 different amino acids (AA). Quantifying the productivity that each population achieved at different time points of the evolution experiment as the number of colony forming units (CFUs) ml^{-1} indicated that the mere presence of AAs already benefitted the ancestral genotype as indicated by a significantly increased productivity relative to AA-deficient conditions (independent sample t-test: $P<0.05$, $n=8$, Fig. 1A). This pattern consistently remained as these populations were further propagated, resulting in a significantly increased slope of the populations that evolved in the presence of amino acids as compared to populations that were selected in un-supplemented minimal medium (linear regression of fitted lines: $P<0.05$, $n=88$, Fig 1A). In other words, the presence of additional nutrients increased the rate of adaptation. Interestingly, a comparison of growth rates of derived populations revealed that the increase of this kinetic parameter over evolutionary time was statistically indistinguishable between populations that evolved in the two selection regimes (independent samples t-test, $P<0.05$, Figure S1). This finding indicates that although the rate at which cells divided was similar, the presence of AAs likely offered more resources for the AA-evolved populations to exploit.

To determine whether the increased productivity of populations that evolved in the presence of AAs was indeed due to the increased availability of nutrients, the growth rates ($\mu max\ h^{-1}$) the focal populations achieved in the presence (i.e. the selection regime) of AAs was subtracted from the growth rates of the same populations determined under AA-deficient conditions. The resulting value ($\Delta\ \mu_{max}$) was significantly greater than zero for 4 of 8 of the ancestral populations (paired samples t-test: $P<0.05$, $n=4$, Fig. 1B), indicating that the growth of the evolutionary ancestor was reduced when AAs were present. However, over the course of evolution, omission of amino acids from the test medium resulted in a decline of the $\Delta\ \mu_{max}$ values for most of the replicate populations to the point that at 2,000 generations seven out of eight populations showed reduced growth rates when AAs were absent (paired samples t-test: $P<0.05$, $n=4$, Fig. 1B).

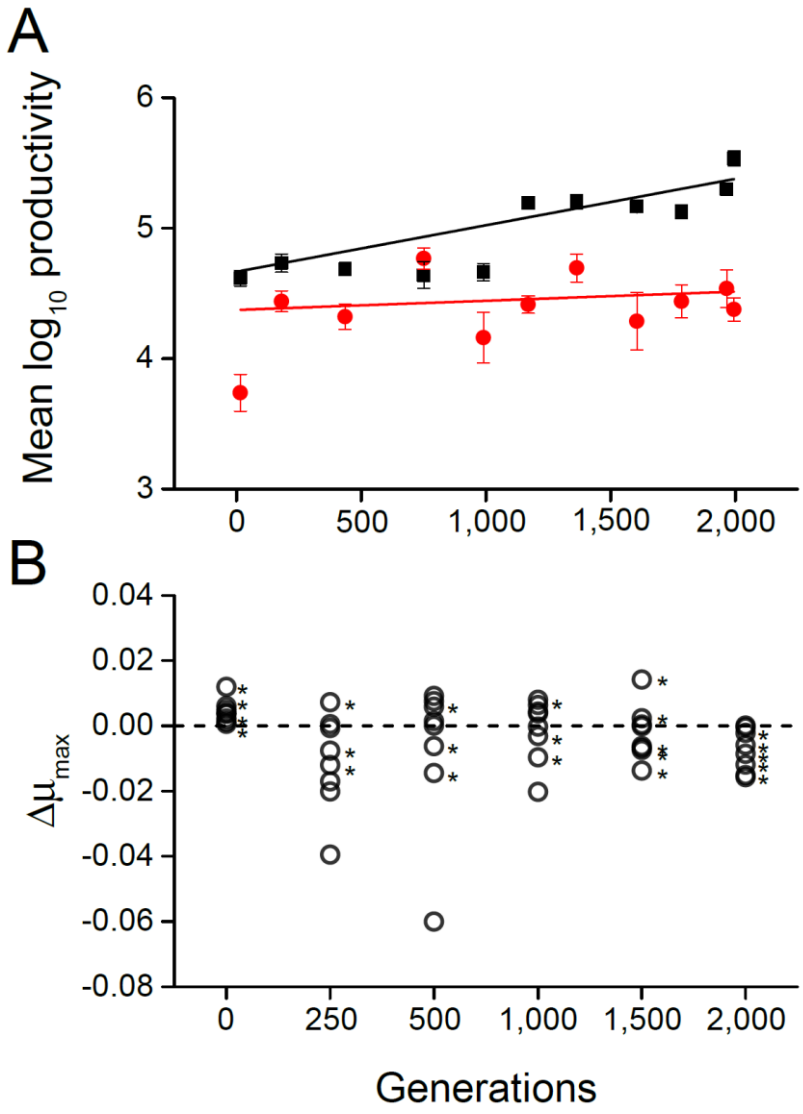


Figure 1: Growth dynamics of evolved populations and amino acid dependency of AA-evolved lines. (A) Mean productivity of the populations that evolved in the presence (black boxes) or absence of AA (red circles) over the course of the evolution experiment. Each data point represents the mean productivity (± 95% confidence interval) of eight replicate populations that are displayed as the $\log_{10}$ of the number of colony forming units per ml (CFU ml^{-1}) at different time points of the experiment. The black and red line represent the exponential fit of the individual points (linear regression of fitted lines: $P<0.05$, $n=88$). (B) Differences in the maximal growth rates (μ_{max} h^{-1}) of the AA-evolved populations when growing in media with or without AA supplementation. Differences were calculated by subtracting the growth rates populations achieved under the AA-replete conditions from the growth rates achieved in un-supplemented minimal medium. Circles represent the different replicate populations and asterisks represent significant differences in growth rates in the two environments (paired samples t-test: $P<0.05$, $n=4$).

Taken together, these analyses showed that the presence of amino acids in the growth environment increased the rate with which populations of *E. coli* adapted to these conditions,

yet also rendered the increase of the population-level productivity contingent on an environmental availability of AAs.

3.2 AA auxotrophies rapidly emerge in AA-containing environments

One possibility to explain the AA-dependent increase in the productivity of populations that evolved in AA-containing environments could be the emergence of AA auxotrophic genotypes that benefitted from utilizing AAs that are available in the environment. These strains would be unable to grow in the absence of AAs, yet show an increased productivity in the presence of AA [181]. To determine whether and to which extent auxotrophic genotypes evolved in both selection regimes, 1,000 colonies of each replicate population from different evolutionary time points were screened for the presence of auxotrophic genotypes. After 0, 250, and 500 generations, no auxotrophic CFU was detected in any of the populations that evolved in the presence of AAs (Fig 2A.). However, when populations from later time points were sampled, 50% (1,000 generations), 25% (1,500 gens.), and 75% (2,000 gens.) of the AA-evolved populations contained auxotrophic genotypes (Fig 2A). After 1,000 generations, the proportion of auxotrophic genotypes detected in these populations ranged between 0.8% and 2.5%, after 1,500 generations between 5.7% and 20%, and between 0.6% to 7.5% after 2,000 generations of evolution in an AA-containing environment (Fig 2A).

Intriguingly, 3 of the 8 populations that evolved without an external supply of AAs also featured AA auxotrophic strains (Fig 2B): Replicate population 3, which had evolved for 500 generations, contained nearly 2% of auxotrophic genotypes, while replicate populations 7 and 8 comprised approximately 1% of auxotrophic strains after 1,500 generations of evolution (Fig 2B). Detecting auxotrophic genotypes in populations that evolved in the absence of AAs is surprising and suggests that these loss-of-function mutants likely obtained the AAs they required for growth from the coexisting prototrophic genotypes, which were present at high frequencies.

A striking pattern that arose in both selection regimes was the dynamics that characterized the emergence of auxotrophic genotypes. Even though the number of auxotrophic genotypes generally increased over evolutionary time, their distribution and abundance within replicate populations showed a high degree of fluctuation around the detection limit of 1,000 cells (Fig. 2). Given that a frequency change of 1% corresponds to at least 10^4 auxotrophic cells, the observed fluctuations are significant on a population-level. Moreover, the fact that their frequency rose to detectable levels ($\geq 10^4$ cells) implies these mutants were likely selectively favored.

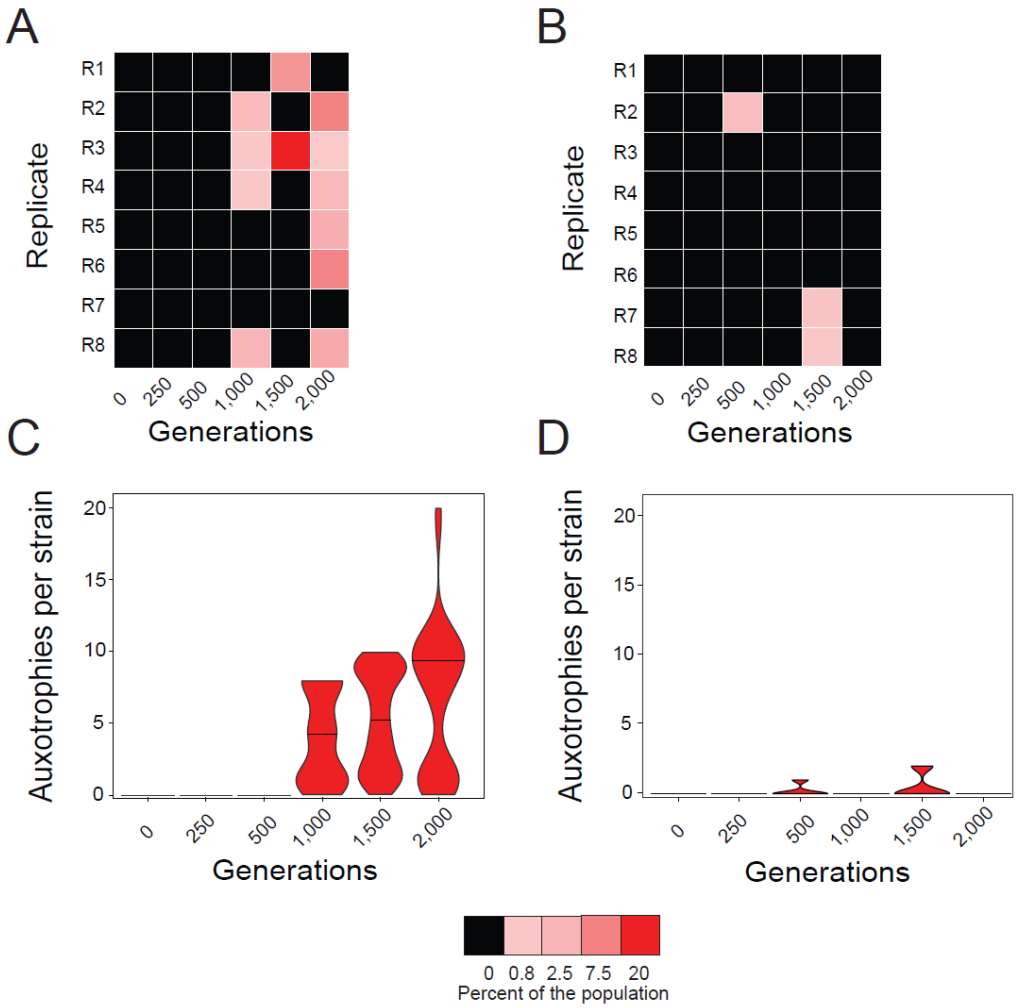


Figure 2: **Rapid evolution of amino acid auxotrophies.** (A,B) Heatmaps depicts the population-level frequency that the AA-auxotrophic genotypes reached per 1,000 CFUs per population analyzed when evolving in (A) the presence of AAs, and (B) the absence of AAs. Rows represent replicate populations (R1-R8) over different time points (columns). Black fields indicate prototrophic populations, whereas different shadings of red represent different frequencies of auxotrophic strains (legend in the figure). (C,D) Distribution of the number of amino acid auxotrophies in the auxotrophic fraction of populations over the course of evolutionary time in the (C) AA and (D) non-AA regimes. Violin plots are scaled to the same maximum width and lines within plot indicates the median of the distribution.

Analyzing the number of different metabolic auxotrophies that were found on a genotype-level revealed that strains isolated from the AA-containing environment were generally impaired in the biosynthesis of more than four different AAs simultaneously, while auxotrophs that evolved in the AA-deficient environment depended on average on only one or two different AAs (Fig. 2C,D). Some of the strains, which have been isolated after 2,000 generations of evolution in the AA-containing environment, even required all 20 AAs (Fig 2C) for growth. Furthermore, the auxotrophic types that emerged in the individual

populations after 2,000 generations of adapting to an AA-containing environment showed a striking congruence in the identity of amino acid auxotrophies that evolved in these populations (Fig. S2), suggesting adaptive advantages likely drove this pattern.

Taken together, these results demonstrate that AA auxotrophies evolved in both selection regimes, yet to a significantly larger extent when AAs have been supplemented to the environment. The finding that their frequency increased under both conditions to detectable levels implies that these loss-of-function mutants were selectively favored.

3.3 Evolved auxotrophies are adaptive when amino acids are environmentally available

To identify whether the rapid emergence and spread of auxotrophic genotypes was driven by selective advantages that resulted from the loss of biosynthetic functions, auxotrophic and prototrophic genotypes that evolved in either the absence or presence of all AAs were competed against their evolutionary ancestor. Determining competitive fitness in this way revealed that auxotrophic strains, which have evolved in presence of AAs, were significantly fitter than the ancestral genotype when all 20 AAs were present in the environment (i.e. the selection regime) (independent sample t-tests: $P<0.05$, $n=130$, Fig. 3A). However, when AAs were omitted, these auxotrophic genotypes were significantly less fit than their ancestor (independent sample t-tests: $P<0.05$, $n=130$, Fig 3A). This observation suggests that auxotrophic genotypes increased in frequency, because they gained an adaptive advantage when AAs were present in the environment. In contrast, the evolutionary success of derived prototrophs was independent of an environmental availability of AAs, as indicated by the observation that their fitness was significantly increased over ancestral levels independent of whether or not AAs were present in the environment (independent sample t-tests: $P<0.05$, $n=130$, Fig. 3A). A qualitatively similar picture emerged when the derived strains of the non-AA regime were analyzed: Supplementing AAs to the growth medium resulted in an increased fitness of auxotrophic genotypes relative to their prototrophic ancestor (independent sample t-tests: $P<0.05$, $n=12$, Fig. 3B). However, under un-supplemented conditions (i.e. the selection regime), auxotrophs were less fit than the ancestor (independent sample t-tests: $P<0.05$, $n=12$, Fig. 3B), while prototrophic isolates gained a significant fitness advantage over the ancestor (independent sample t-tests: $P>0.05$, $n=12$, Fig. 3B) that was quantitatively comparable to the advantage AA-evolved prototrophs gained in the presence of AAs.

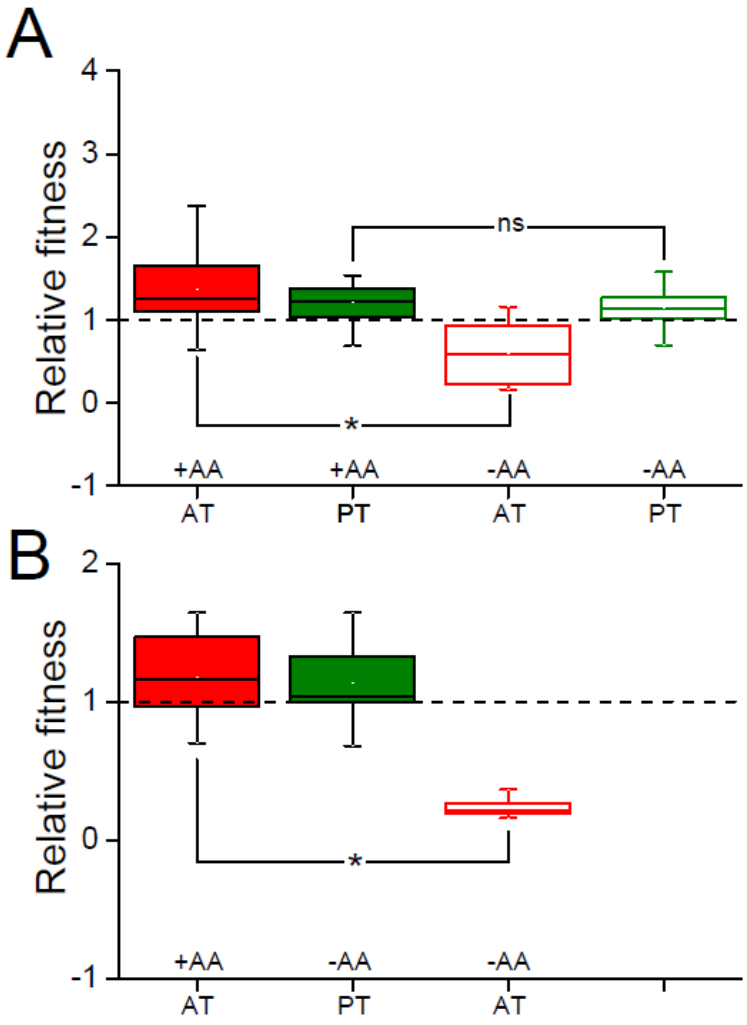


Figure 3: **Fitness advantage of auxotrophs, but not prototrophs is AA-dependent.** Relative fitness of auxotrophs that evolved in (A) the presence of AAs or (B) absence of AAs. Competitive fitness of auxotrophs (red boxes) or prototrophs (green boxes) relative to the evolutionary ancestor was measured in the presence (filled boxes) or absence (filled boxes) of amino acids. Asterisks indicate significant fitness differences relative to the prototrophic ancestor while 'ns' denotes non-significant differences (independent samples t-tests: P>0.05, n=130 for AA regime and n= 12 for non-AA regime). The dashed line at y=1 represents the fitness of the ancestor. Box plots consist of medians (horizontal lines within boxes), interquartile range (boxes), and 1.5x-interquartile range (whiskers).

Taken together, these findings imply that an environmental availability of amino acids favored mutants that have lost the ability to biosynthesize certain AAs autonomously. In contrast, the fitness advantage gained by prototrophic genotypes was independent of the presence of AAs in the environment.

3.4 Auxotrophs evolved a metabolic dependency on coexisting prototrophs

Two findings of the abovementioned experiments beg an explanation. First, AA auxotrophies evolved even when no AAs were present in the environment (Fig. 2B). Second, even though AA-evolved auxotrophs were less fit than the ancestor when no AAs were present in the environment (independent sample t-test: $P<0.05$, $n=130$, Fig. 3A), these genotypes still grew to detectable levels and were not lost from these populations. From where did these auxotrophic genotypes obtain the amino acids they needed to grow? A likely source could be the prototrophs that coexisted with the auxotrophic genotypes in the two abovementioned experiments. To test this possibility, auxotrophic genotypes were grown in monoculture, together with the prototroph they co-evolved with, or the evolutionary ancestor (both co-cultures: 1:1 ratio). This test was performed in either the absence or the presence of AAs and the productivity of the auxotrophs (i.e. the Malthusian parameter) was determined by plating.

As expected, auxotrophs were unable to grow when no AAs were present in the environment, yet grew when all AAs were supplemented to the growth medium. This held true for all auxotrophs analyzed from both selection regimes (Fig. 4A,B). However, co-culturing auxotrophic genotypes in the absence of AAs together with the prototrophic strain they had coevolved with, resulted in productivity levels of the auxotrophs that were statistically indistinguishable to the levels they have reached in monoculture in the presence of AAs (one-way ANOVA followed by a LSD post-hoc test: $P<0.05$, $n=130$ (AA-regime), $n=12$ (non-AA regime), Fig. 4A,B) indicating that auxotrophs derived amino acids from the co-cultured prototrophs. Supplementing these co-cultures with additional AAs further increase the productivity of AA-evolved auxotrophs over the productivity they reached under un-supplemented conditions (one-way ANOVA followed by a LSD post-hoc test: $P<0.05$, $n=130$, Fig. 4A), while the growth of non-AA-evolved auxotrophs did not change upon AA supplementation (one-way ANOVA followed by a LSD post-hoc test: $P>0.05$, $n=12$, Fig. 4B). Interestingly, when the focal auxotrophs were co-cultured with the evolutionary ancestor and not the coevolved prototrophs, the auxotrophs isolated from the AA regime showed generally reduced productivity levels compared to the situation when the coevolved prototroph was present (one-way ANOVA followed by a LSD post-hoc test: $P<0.05$, $n=130$, Fig. 4A). Still, AA supplementation significantly enhanced the productivity of AA-evolved auxotrophs in co-culture with the ancestor over the productivity reached in unsupplemented medium (one-way ANOVA followed by a LSD post-hoc test: $P<0.05$, $n=130$, Fig. 4A). This pattern is consistent with a coevolutionary change between derived auxotrophs and prototrophs that is absent when the ancestral prototroph is the interaction partner. In contrast, for the auxotrophs that evolved in the non-AA regime, it did not make a difference whether the coevolved prototrophs or the evolutionary ancestor was present

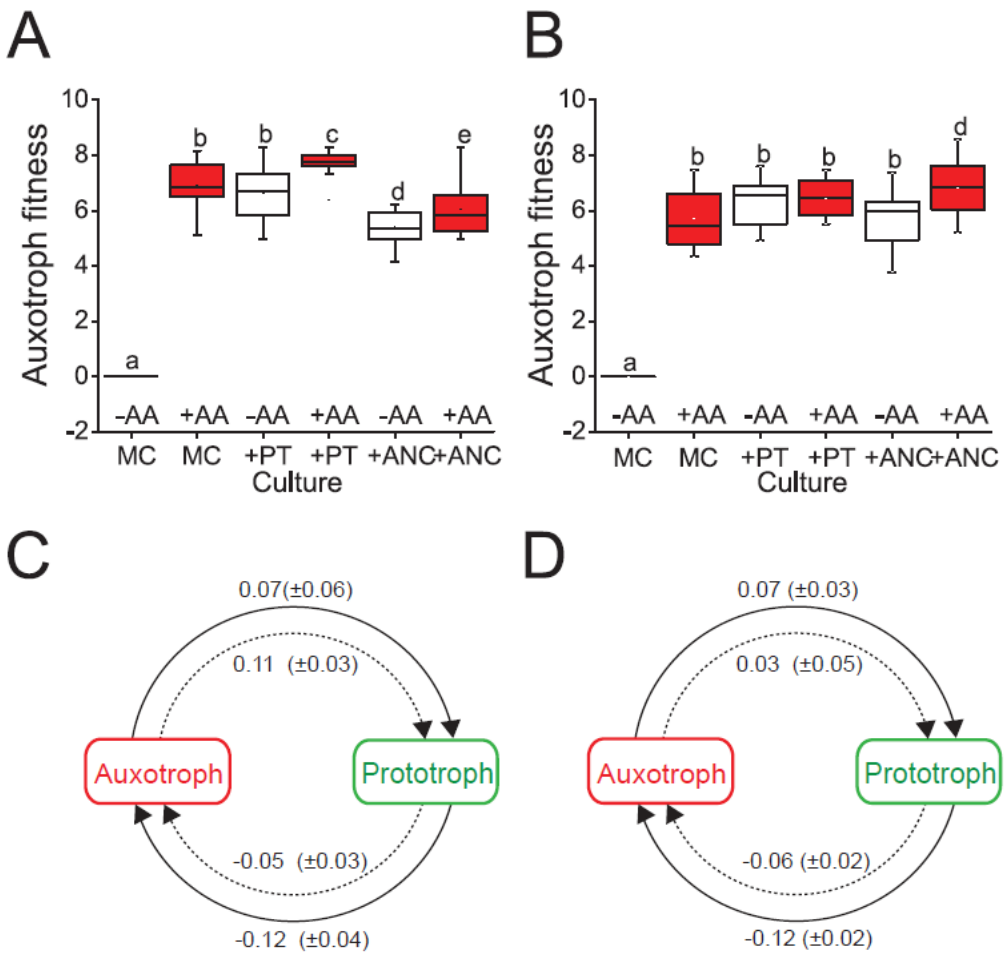


Figure 4: **Auxotrophic strains evolve frequency-dependent interactions with coexisting prototrophs.** (A,B) Malthusian growth parameters of auxotrophic strains that have been isolated after 2,000 generations of growth under (A) AA-replete and (B) AA-deficient conditions. Growth was determined in monoculture (MC) or in co-culture with either a prototrophic strain isolated from the same time-point and population (PT) or the evolutionary ancestor (ANC) in the presence (red boxes) or absence of amino acids (empty boxes). Different letters in panels A and B indicate significant differences in the auxotrophs' fitness (one-way ANOVA followed by a LSD post-hoc test: $P<0.05$, n=130 for AA regime and n= 12 for non-AA regime). (C,D) Reciprocal invasion-from-rare experiment shows the interaction between co-occurring auxotrophs and prototrophs is stabilized by negative frequency-dependent selection. Selection coefficients of the auxotrophs from (C) the AA regime and (D) non- AA regimes. Selection coefficients of the auxotrophic strains were measured when a resident population of auxotrophs or prototrophs was invaded at rare frequencies by the corresponding type, indicated by the direction of the arrows, in the presence (solid arrows) or absence (dashed arrows) of AAs. All comparisons are statistically significant (one sample t-test: $P<0.05$, n=130 for AA regime and n= 12 for non-AA regime).

when the co-culture was exposed to un-supplemented minimal medium (one-way ANOVA followed by a LSD post-hoc test: $P>0.05$, n=12, Fig. 4B). However, when amino acids were provided to these cultures, the auxotrophs reached the highest productivity of all

experimental conditions analyzed (one-way ANOVA followed by a LSD post-hoc test: $P>0.05$, $n=12$, Fig. 4B).

Altogether, these observations suggest that either the auxotrophs' ability to derive amino acids from the coexisting prototrophs increased over time or, alternatively, the prototrophic cells increased their amino acid production levels. In either way, these results demonstrate that evolved auxotrophs utilized the AAs that are available in the growth environment as well as those they could obtain from other, coexisting strains.

3.5 Negative frequency-dependent selection maintains genotypic diversity

Which ecological mechanism maintained the evolved genotypic diversity (i.e. both auxotrophic and prototrophic genotypes) in the evolution experiment? A likely mechanism that has been previously identified to be key for maintaining synthetically engineered cross-feeding genotypes that reciprocally exchanged essential amino acids is negative frequency-dependent selection [24]. To determine whether the same mechanism also stabilized our naturally evolved genotypes, the ability of the derived auxotrophs and prototrophs to invade a population of the respective other strain when rare was determined in the absence and presence of AAs in the environment. Under all conditions tested was the initially rare type (i.e. auxotroph or prototroph, initial ratio: 1:100) able to invade a resident population of the respective other strain as evidenced by the significantly increased selection coefficients of the invading type (one-sample t-test: $P<0.05$, $n=24$, Fig. 4C,D). This finding corroborated the hypothesis that AA cross-feeding between prototrophic donor cells and auxotrophic recipients is maintained by negative frequency-dependent selection.

Finally, growing monocultures of AA-evolved genotypes and the evolutionary ancestor in the presence of AAs and comparing their growth kinetic parameters revealed that the derived auxotrophic and prototrophic strains grew significantly faster and achieved a higher productivity than the ancestral strain (one-way ANOVA followed by a LSD post hoc test: $P<0.05$, $n=20$, Fig S3A). Moreover, the growth of auxotrophic strains was characterized by a significantly shorter lag phase (one-way ANOVA followed by a LSD post hoc test: $P<0.05$, $n=20$, Fig S3A) and an earlier onset of the stationary phase than was the case for both derived and ancestral prototrophic strains (one-way ANOVA followed by a LSD post hoc test: $P<0.05$, $n=20$, Fig S3B), indicating that auxotrophic strains likely utilized environmentally available AAs until this pool was depleted. In contrast, prototrophic genotypes remained much longer in the exponential growth phase than auxotrophs (one-way ANOVA followed by a LSD post hoc test: $P<0.05$, $n=20$, Fig S3C), suggesting that their growth was limited by the carbon source rather than the amount of AAs present in the environment.

Taken together, these observations suggest that because of their inability to produce certain amino acids autonomously, auxotrophs utilize first the pool of environmentally available AAs, before exploiting the AAs produced by other, coexisting strains. This bi-phasic growth pattern along with the obligate dependency of auxotrophs on coexisting prototrophs likely maintained auxotrophs by negative frequency-dependent selection.

3.6 Mutations in both structural and regulatory genes caused auxotrophic phenotypes

To unravel the genetic basis of the metabolic auxotrophies that emerged in the course of the evolution experiment, genomes of 8 representative auxotrophs and 6 representative prototrophs from different replicate populations, which had evolved for 2,000 generations in the absence or presence of AAs, were re-sequenced and compared to the genome of the evolutionary ancestor. This analysis revealed that prototrophs and auxotrophs had very few mutations in common (Fig. 5, Table S2 and S3), thereby confirming that the two coevolved strains represented indeed genetically distinct subpopulations.

Next we focused on those mutations that occurred exclusively in genomes of auxotrophic types to determine if these could be causal for these phenotypes. Both, auxotrophs isolated from the non-AA regime carried identical SNPs in the *rpoB* and *yqiB* gene. The former gene encodes the β subunit of RNA polymerase (RNAP) [240] and the latter a predicted dehydrogenase (b3033, ECK3024; Ecocyc [219]). Interestingly, both of these genotypes were auxotrophic for lysine and tryptophan. The mutation identified in the *rpoB* gene has been predicted to negatively affect protein activity [241] (Table S3). It has been shown that similar mutations, such as deletions in the β-subunit of RNAP, confer increased growth rates in minimal medium to the corresponding *E. coli* mutants and affect the ability of RNAP to bind to certain operons or promoters, especially those involved in amino acid biosynthesis, resulting in a general down-regulation of genes of this class [242].

In contrast, the auxotrophies that arose in the AA-containing environment were caused by completely different mechanisms including the loss of *regulatory* and *structural genes*. One auxotroph (isolated from population R2) lost approximately 13 kb from its genome (Fig. 5, Table S3). The deleted fragment comprised 14 genes, which included the ones encoding the BaeSR two-component regulatory system. Deletion of these two genes has previously been shown to result in down-regulation of multiple regulatory and amino acid biosynthesis-associated proteins [243]. Another auxotroph (population R5) carried a non-sense mutation, which inactivated the *sspA* gene that encodes the stringent starvation protein A (Fig. 5, Table S2 and S3). This gene is known to be involved in the regulation of amino acid biosynthesis and mutants lacking this gene have been previously shown to lose viability under arginine-limiting conditions [244]. Moreover, the sspA protein is known to inhibit the accumulation

of H-NS, a global gene regulator [245]. H-NS in turn is known to inhibit the expression of several genes and regulons, especially those involved in amino acid biosynthesis such as the leucine regulatory protein (lrp) [246]. Lrp in turn controls multiple amino acid biosynthetic operons and regulons [247]. Thus, a loss of function mutation causing a loss of protein activity in the auxotrophic strains would result in increased H-NS levels, which in turn would repress lrp and multiple biosynthetic regions, thereby causing the auxotrophic phenotypes that have been seen in the mutants.

The only case in which a biosynthetic gene has actually been lost from the genome was in case of an auxotroph that has been isolated from population 6. This mutant had lost a 13 kb region from its genome that contained 20 genes, including *proA* and *proB* that are essential for the biosynthesis of the AA proline [248]. In addition, this strain also carried a mutation in the *stpA* gene, which encodes for the starvation protein A (Fig. 5, Table S3. Interestingly, starvation protein A is known to interact with H-NS and lrp in regulating the expression of amino acid biosynthesis genes in *E. coli* - especially under conditions of amino acid starvation [246]. Although it is not known whether the observed SNP affects protein function (Table S3) [241], this mutation could have resulted in a reduced binding ability of starvation protein A to H-NS, leaving the latter free to repress biosynthetic genes in a lrp-mediated manner, similar to the case described above. Surprisingly, the genomes of the other auxotrophs (R3, R4, and R8; Fig. 5, Table S3) showed no mutations with obvious links to amino acid biosynthesis.

In sum, analyzing the genomes of the evolved strains identified numerous mutations in genes that have been previously implicated in amino acid metabolism, suggesting that they are likely causal for the observed auxotrophic phenotypes.

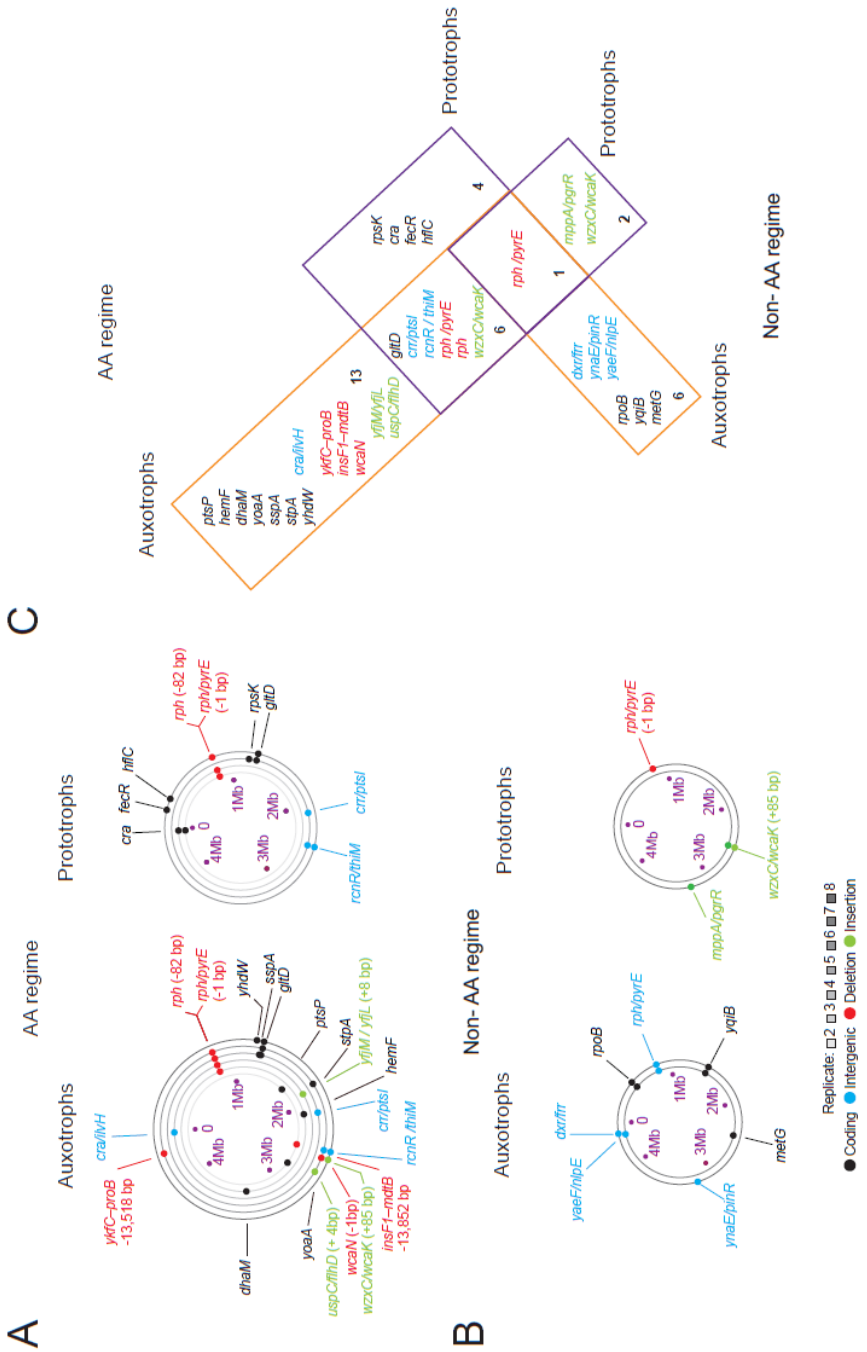


Figure 5: **Genomic consequences of adaptation to the two selection regimes.**

(A, B) Each ring represents the genome of a single auxotrophic or prototrophic clone that has been been isolated from different replicate populations, which have evolved for 2,000 generations under (A) AA-replete and (B) AA-deficient conditions. The order of rings represents the replicate populations from which the cognate auxotroph and prototroph have been isolated. Labels and lines connecting dots on the rings represent mutations and their positions on the genome with the color of the label and dot specifying the type of mutation (see legend in the figure). (C) Venn diagrams depict mutations which are unique to auxotrophs (orange rectangles) and prototrophs (purple rectangles) and those that are common between both types (region encompassing both rectangles) for the strains from the AA (top rectangles) and non-AA regimes (bottom rectangles). Numbers indicate the number of common mutations between the strains.

4. Discussion

Why are metabolic auxotrophies so common in natural microbial communities? Hypothesizing that adaptive benefits may account for the frequently observed loss of metabolic functions, our evolution experiment revealed that prototrophic *Escherichia coli* cells rapidly evolved metabolic auxotrophies when adapting to environments that contained all of 20 different amino acids. Interestingly, also serial propagation in AA-free environments resulted in the emergence of genotypes that had a lost the ability to autonomously produce some amino acids, yet the number of auxotrophies per strain, the number of auxotrophic strains per population, and the number of populations containing auxotrophs was significantly lower relative to populations that evolved under AA-replete conditions. In line with the initial hypothesis, auxotrophic genotypes that evolved in AA-containing environments gained an adaptive advantage over their evolutionary ancestor, yet the observed fitness benefit was contingent on the presence of AAs in the environment. Surprisingly, all evolved auxotrophs also derived amino acids from coexisting prototrophic cells and this interaction was stabilized by negative frequency-dependent selection. Finally, multiple genetic routes lead to the inactivation of amino acid biosynthetic abilities, including mutations in regulatory and structural genes.

A main outcome of the evolution experiment was that adaptive benefits drove the rapid loss of biosynthetic functions when the focal metabolites were sufficiently present in the cell-external environment. These findings are in line with previous analyses, which revealed a significant fitness advantage synthetically engineered, auxotrophic mutants gained over competing prototrophic types when AAs were sufficiently present in the environment [24,85,181,213]. A prediction that follows from these observations is that metabolic auxotrophies should rapidly evolve whenever bacteria are cultivated in AA-rich media or inhabit environments with increased AA-availabilities [181,249,250]. Indeed, metabolic auxotrophies have been repeatedly reported to arise in laboratory-based evolution experiments [172-174] or have been detected in natural microbial communities [22,24,37,58,67,82,181,213,251]. In our experiment, derived auxotrophs always coexisted together with metabolically autonomous prototrophs. A strikingly similar pattern has been previously observed in populations of *Pseudomonas aeruginosa* that adapted to the lungs of cystic fibrosis (CF) patients: both prototrophic and auxotrophic strains have been isolated from the amino acid-rich mucus that fills the lungs of these CF patients [251,252].

Independent of whether or not AAs were present in the selective environment, auxotrophs that evolved in our evolution experiment always obtained AAs also from other community members such as the coexisting prototrophs. Two mechanisms are conceivable how auxotrophs obtained the AAs they required for growth: metabolites might be exchanged among genotypes via diffusion through the cell-external environment [26,51,113] or,

alternatively, in a contact-dependent manner [253,254]. Recently it has been described that auxotrophic cells of *E. coli* can produce so-called *nanotubes* to directly obtain cytoplasmic AAs from other bacterial cells [254]. By reducing the loss of AAs to the cell-external environment, these structures likely minimize the costs to the AA-producing cell and might thus be interpreted as a strategy of bacteria to survive under AA-limiting conditions.

Analyzing the genomes of derived mutants unraveled the mutational causes that gave rise to the observed phenotypes. In contrast to expectations, deactivation of amino acid biosynthetic pathways via a loss of the corresponding structural genes was much less common than a loss or inactivation of key regulatory elements (Fig. 5, Table S3). Interestingly, auxotrophies that evolved in the non-AA regime were mostly due to a mutational down-regulation of biosynthetic gene expression, while most auxotrophies that evolved in the AA-containing environment were caused by a complete loss of enzyme-encoding regions or an inactivation of the corresponding regulatory elements. This pattern likely mirrors differences in the two selective regimes. While the environment that did not contain AAs likely penalized any newly evolved auxotroph, whose metabolic deficiency could not be compensated by any of the prototrophic types present, the AA-replete condition likely favored many more different auxotrophs. Indeed, the only auxotrophs that could be detected in the lines that evolved under AA-free conditions had lost the ability to produce leucine, lysine, and tryptophan, which incur relatively low metabolic costs to produce [87] and are thus cheaper for the corresponding prototrophs. In contrast, in the AA-replete environment many more auxotrophic mutants evolved; with all replicate populations displaying a core set of common auxotrophies (Fig. S2), indicating that such parallel loss of these biosynthetic abilities was a convenient and tractable adaptive path for these bacteria, adequately compensated for by the environment.

Theoretical predictions of metabolic auxotrophies have been largely based on whether or not a biosynthetic route exists in a given bacterial genome [22,181,213]. Due to a lack of understanding of the underlying regulatory networks, these approaches usually neglect the multifarious genetic routes that can possibly cause metabolic auxotrophies. Consequently, previously published estimates that only consider the absence or presence of biosynthetic genes [22,181,213] likely underestimate the true number of auxotrophic prokaryotes in nature dramatically. Given that the fitness advantage multiply auxotrophic bacteria gain are strongly affected by epistatic interactions among the auxotrophy-causing mutations [213], mutationally-induced regulatory changes could represent an effective bypass of this evolutionary constraint.

The adaptive loss of metabolic capabilities and the emergent dependence on other co-occurring strains as observed in this study have significant ramifications for the evolution of bacterial genomes. A striking pattern that emerges when genomes of multiple different

bacterial clones are sequenced that have coexisted together for extended time periods, is not only the frequent loss of many biosynthetic functions from their genomes, but often also a high degree of metabolic complementarity on the genomic level. Examples involve both free-living bacterial communities [255] and consortia of endosymbiotic bacteria, whose metabolite production is intricately interwoven between their eukaryotic host [47,79,152] and other coevolving bacteria [32,66,177]. Given the ease, with which metabolic auxotrophies can evolve, thereby rendering the resulting mutants dependent on other coexisting organisms, it is conceivable how this event can set the stage for a co-evolutionary race, in which the interacting partners may benefit from losing additional metabolic functions. This race will most likely favor those loss-of-function mutants, who are fitter than other, competing genotypes given the presence of a donor that can sufficiently compensate for their deficiencies. In the long-run, this 'black-queen'-like process [113] may lead to co-adaptations on both sides. Indeed, our observation that the AA-evolved auxotrophs grew significantly better when co-cultured with the derived prototrophs than their evolutionary ancestor supports this interpretation (Fig. 4A). In the long-run, this process should lead to intricately connected metabolic networks between multiple different bacterial genotypes. Ultimately, the findings of our study may provide a plausible explanation of why most bacterial species known are difficult to cultivate under laboratory conditions [179,180]: most likely they have adapted to the nutritional biotic and abiotic environment, which complicates a reproduction of these conditions in the laboratory.

Acknowledgments

The authors would like to thank the EEE group for helpful discussions and constructive criticism, Prof. Wilhelm Boland for support, as well as Silvio Waschina and Daniel Preussger for performing some of the daily transfers. Ellie Harrison is gratefully acknowledged for sharing the R script that was used to generate the genome plots. This project was funded by the Volkswagen Foundation, Jena School for Microbial Communication (JSMC), and the Max Planck Society.

Supporting Information

The following supporting information is available for this chapter on page 216

Figure S1: Growth rates relative to the ancestor of populations evolving under the two different regimes

Figure S2: Amino acid auxotrophy profiles in the replicate populations over time

Figure S3: Growth kinetic parameters of the ancestral, auxotrophic and prototrophic strains from the AA-regime

Table S1: Composition of different dropout media

Table S2: Mutations present in the evolved genotypes

Table S3: Mutations unique to auxotrophic strains and their predicted functional implications.

General Discussion

General Discussion

The genomic and metabolic composition of an organism are important determinants of their adaptive capabilities in the environments they dwell in [256]. A result of reductive evolution of bacterial genomes is a concomitant decline in biosynthetic capabilities and loss of metabolic autonomy, which makes such organisms dependent on the environment for the uptake of key metabolites. The loss of biosynthetic functions involved in production of metabolites like amino acids, nucleobases, and cofactors is intriguing since such metabolites serve as constituents of important cellular macromolecules, e.g. amino acids in protein synthesis, or are required for their activity, e.g. cofactors in enzyme catalysis. Thus, the functioning of key cellular processes would be reliant on environmental availabilities of these metabolites. However, the evolutionary mechanisms that cause the loss of such biosynthetic capabilities in bacteria are generally unclear with natural selection and random genetic drift suggested as possible mechanisms.

In order to unravel the causes driving a reduction in biosynthetic functions in bacteria, a first step that was attempted was to quantify the extent of such losses in bacteria. An *in silico* analysis of 949 (later updated to 1432) completely sequenced bacterial genomes and their metabolic reconstructions revealed that loss of biosynthetic genes and functions, i.e. auxotrophy for amino acids, cofactors and nucleobases is surprisingly prevalent in the bacterial world. Interestingly, free-living bacteria also were frequently auxotrophic. What could be the fitness consequence of losing such biosynthetic functions in bacteria? Experiments using auxotrophic strains of *Escherichia coli* and *Acinetobacter baylyi* showed that in most cases a loss of biosynthetic genes resulted in the increased fitness of auxotrophs relative to the prototrophic counterparts when the requisite metabolite required by the auxotrophic strains was present in the environment. Thus, the selective advantages that the auxotrophic strains gain over prototrophic types can explain why bacteria in nature lose their metabolic autonomy and become dependent on the environment.

Another common observation from the *in silico* analysis was that bacteria spanning different lifestyles frequently also lost multiple biosynthetic functions. How can the genetic interactions between different biosynthetic gene deletions, i.e. *epistasis*, influence the fitness consequences of losing biosynthetic genes? Competitive fitness measurements of auxotrophic *E. coli* strains, which had lost one or multiple biosynthetic genes, versus the prototrophic strain revealed prevalent *epistasis* between biosynthetic functions. Interestingly, *epistasis* significantly constrains the rate of adaptation by causing a pattern of diminishing returns of adaptive benefits with an increasing number of biosynthetic mutations. Furthermore, the observed *epistatic* patterns between biosynthetic genes and selective advantages of gene loss were strongly dependent on the type of carbon source present in the

nutritional environment. These findings suggest that the adaptive loss of biosynthetic functions can be strongly influenced by *epistasis* and *environmental plasticity*.

How can the selective advantages observed in auxotrophic strains and the variation in fitness in different carbon source environments be explained? A two pronged approach of computational and experimental analyses revealed that the type of carbon source and the architecture of the metabolic network, determines the metabolic costs of production of a metabolite. These costs in turn determine the growth advantages that bacteria gain upon loss of biosynthetic genes. This result provides a rich insight into how biochemical trade-offs, which are shaped by a cell's metabolic network can influence metabolic growth strategies in bacteria. Thus, the results of chapter 1, 2 and 3 strongly make the case for adaptive benefits in driving the prevalent loss of biosynthetic functions in bacteria in environments rich in metabolites.

Does natural selection favor the loss of genes in nutrient-rich environments? An experimental evolution approach revealed that the loss of biosynthetic genes and functions is indeed favored by selection. Replicate isogenic populations of an initially prototrophic strain of *E.coli* repeatedly evolved to become auxotrophic when serially propagated in both amino acid-rich and -poor environments. However, the frequency of evolved auxotrophic strains was much higher in the amino acid-rich environment. Nonetheless, all the auxotrophic genotypes were adaptive when amino acids were present in the environment. Interestingly, auxotrophic strains also evolved to become dependent on the coexisting prototrophs, thus giving rise to a previously absent metabolic interaction. The stable coexistence and the unidirectional dependency between auxotrophic and prototrophic genotypes was stabilized by negative frequency-dependent selection.

Taken together, the findings described in this thesis provide strong experimental evidence for natural selection in shaping the evolution of microbial communities by: i) driving the loss of biosynthetic functions and thereby causing a dependency on the environment, ii) promoting metabolic specialization of bacterial genotypes, and ii) consolidating specialization and dependencies through ecological interactions. In the subsequent sections I discuss how these processes might occur in bacterial collectives and their implications for the evolution of microbial communities.

1.1 Metabolic dependencies are widespread in the bacterial world

A major revelation from this thesis is that bacteria with seemingly free-living lifestyles also lose biosynthetic genes like bacteria with specialized lifestyles i.e. endosymbionts or gut dwelling bacteria. The prevalence of auxotrophies in bacteria (chapter 1) is supported by findings of other studies. In one study, Mee and coworkers (2014) employed *in silico* approaches to quantify the presence or absence of complete and intact biosynthetic pathways

and observed that a large proportion of bacterial genomes lack complete biosynthetic pathways for multiple amino acids [22]. In another study, Monk and coworkers (2013) used genome scale reconstructions of metabolic networks to predict the anabolic capabilities of 55 *Shigella* and *E.coli* strains. The authors found that almost 22 % of the analyzed strains were auxotrophic for amino acids and multiple vitamins [67]. A caveat that needs to be accounted for here is that the approaches employed to estimate the loss of biosynthetic capabilities are explicitly based on inferences from a comparative analysis of bacterial genomes and not on the basis of phenotypic surveys of auxotrophic bacteria in nature. Can these *in silico* predictions be corroborated by occurrences of auxotrophs in nature? Auxotrophic isolates are frequently isolated from diverse habitats: vitamin requiring strains dominate microbial communities in soil and aquatic environments [257], amino acid auxotrophic strains of *Pseudomonas aeruginosa* are present in the infected lungs of human individuals who suffer from cystic fibrosis [251], endosymbiotic strains of *Rhizobium* in root nodules of plants are auxotrophic for amino acids [258,259] in addition to the numerous genome reduced endosymbiotic bacteria that are well-described in literature [32,45,69,177]. In addition, auxotrophic mutants have also been observed to rapidly evolve in experimentally evolving bacterial populations (as in chapter 4). For instance, when pathogenic strains of *Legionella pneumonophila* were adapted to mouse macrophages [174], *E.coli* strains were adapted to the mouse gut [173], *Pseudomonas aeruginosa* strains were propagated as biofilm cultures [172] or when *Bacillus subtilis* strains were adapted to nutrient-rich environments [133].

1.1.1 A diversity of mechanisms can cause metabolic dependencies

The estimates of metabolic auxotrophies that are based on genome reconstructions and lifestyles of 949 and 1437 genomes in this thesis (chapter 1 and 2) are among the first systematically reported predictions for bacteria. However, these estimates are likely very conservative for multiple reasons. First, the criteria to consider if biosynthetic capabilities were intact in a genome was very stringent and required at least 50% of a biosynthetic pathway to be lost for a bacterial species to be deemed as auxotrophic. However, as in the case for *E. coli* and *A. baylyi* (chapter 1 and 2), a single deletion was sufficient to cause an auxotrophic phenotype in the bacterium. Second, the *in silico* analyses are impeded by the limited number of completely sequenced and annotated genomes that are available. Genome databases are also inherently biased, because a large proportion of bacterial genomes in these datasets are those that are of pathogenic or economic importance (e.g. biotechnology), and thus these databases are considerably underrepresented with bacteria having alternate lifestyles [260,261]. Last, only *structural genes* are taken into account in the *in-silico* analyses and the auxotrophies that arise due to *regulatory elements*, which have been suggested to play an important role in adaptation [169,262], not considered. Understanding how mutations in

structural or *regulatory* elements influence loss of biosynthetic function in bacteria is essential to accurately predict the extent of such losses in nature.

In bacteria the biosynthesis of essential cellular metabolites is tightly governed by numerous other genetic elements and molecules in addition to enzymes. These include *structural elements* such as sequences of DNA like promoters, coding regions, and terminators, to which *regulatory elements* bind and coordinate expression of genes and pathways. Loss of any of these elements can result in an inactivation or down-regulation of a gene product involved in metabolite biosynthesis, thus causing a cell to display an auxotrophic phenotype. Interestingly, the evolution of auxotrophs in this study could be attributed to mutations in both *structural* and *regulatory elements* (chapter 4). How can diverse genetic alterations lead to convergent phenotypes (in this case the loss of functions)? For *structural elements*, the answer is simple. Loss of the DNA sequence due to mutations can cause a loss of function or abolish the binding ability of a *regulatory element*. However, in the case of *regulatory elements* the case is quite different.

The explanation is centered around the architecture of cellular networks in bacteria, which are inherently *modular* [169]. As a result, a single mutation in a *regulatory element* (For instance hns in Fig. 1A) can have consequences on multiple (90 downstream target ORFs for hns, Fig. 1A) components coherently elsewhere in the network [263]. For instance, in *E. coli* there are 300 transcription factors which are organized hierarchically with few master regulators (Fig. 1A) [263]. Of these, the regulatory proteins CRP, FNR, IHF, Fis, ArcA, NarL, H-NS, Fur and Lrp control over half of all genes, through direct and indirect interactions [263]. Thus, a loss of function mutation in any of these genes will alter or shut down expression of associated genes, and additionally trigger far-reaching changes in cellular metabolism [169].

In the specific case of amino acid metabolism in *E. coli,* the transcription factors ArgR, Lrp, TrpR and TyrR directly or indirectly regulate the biosynthetic pathways of almost all of 20 amino acids (Fig. 1B)[247,264,265]. These transcriptional factors are in turn regulated hierarchically by many more core cellular regulators (Fig. 1A) [244-246]. Therefore, a mutation that inactivates the function of any of these *regulatory elements* in bacteria can lead to a loss of the ability to autonomously biosynthesize these metabolites leading these genotypes to depend on the environment. Similar mechanisms likely also operate in other bacterial genotypes. For instance, uracil auxotrophy in isolates of *Lactobacillus plantarum* has been shown to account as a result of losses of genes involved in biosynthesis as well as mutations in *regulatory elements* [266].

The complicated hierarchy and interconnectedness of the components in cellular networks underlying amino acid metabolism in bacteria have two important implications. First, they expand the suite of genetic targets (in addition to *structural genes*) likely accessible to natural selection to mediate the inactivation of biosynthetic functions in bacteria. This, is especially

interesting in light of a recent study where it was observed that the deletions of regulatory functions were in most cases accompanied by increases in fitness [169]. This finding is indicative that the adaptive losses of regulatory elements facilitate the emergence of new phenotypes [169]. Furthermore, it is well known that mutations in *regulatory elements* play an important role in the early steps of adaptation to new environments [169,262]. Second, the inactivation of *regulatory elements* can result the in the loss of multiple biosynthetic functions simultaneously. In this study, the loss of multiple *structural genes* resulted in prominent *epistatic effects* which constrain (or diminish) the increase of adaptive benefits (chapter 2) in multiply auxotrophic strains. Inactivation of regulatory elements could reflect an alleviation of these constraints that are imposed by *epistasis* on adaptive evolution. Indeed the finding that *regulatory elements* were subject to mutations in the case of the derived auxotrophic strains (chapter 4), strengthens this argument.

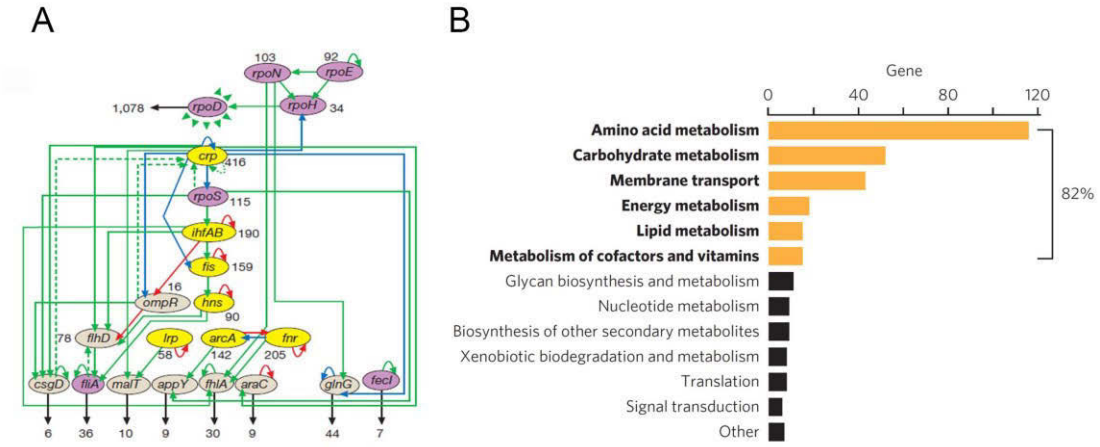


Figure 1A: Network of cellular transcription and sigma factors in *E. coli*. Green, red and blue arrows denote direct activating, repressing and dual interactions. Sigma factors, master regulators and lower-tier regulators are in purple, yellow and beige, respectively. The numbers denote the total number of direct downstream ORF–gene interactions that controlled by each node. Figure from Isalan *et al*., (2008) [263].

Figure 1B: Functional gene categories that are directly regulated by ArgR, Lrp and TrpR. Yellow bars indicate categories that are related to amino acid metabolism whereas black bars represent genes involved in energy metabolism. The percentage represents the total fraction of amino acid metabolism related functions. Figure from Cho *et al*., (2011) [247].

In summary, the loss of biosynthetic functions could be much more prevalent than estimated by *in silico* analyses because of the multiple ways that can lead to metabolic auxotrophies in bacteria.

1.2 Evolutionary adaptation rather than drift can explain loss of biosynthetic functions in bacteria

The outcomes in this thesis strongly imply that loss of biosynthetic gene loss in bacteria is primarily a consequence of adaptive processes and that selection favors the loss of costly biosynthetic functions when the metabolite is sufficiently compensated by the environment (chapter 1,2 and 4). Therefore, one implication of this finding is that selection favors the removal of constraints that bacteria face during growth and survival in their environment. It is a well-known paradigm that bacterial growth is severely constrained by the expression of unneeded proteins since their production diverts resources from making other important and beneficial proteins that are immediately relevant to the cells environment [90,109]. Most natural environments of bacteria are rich in many important metabolites required by a bacterial cell, for instance amino acids (chapter 1 and Table 1) [257]. In the case of biosynthetic functions, the environmental availabilities of metabolites make the autonomous production capabilities of these metabolites redundant and thus mean that prototrophic cells have to incur the costs of production. Although, it can be argued that regulatory mechanisms in bacterial cells act to shut down biosynthetic pathways when the metabolite is present in the environment, evidence suggests that regulatory processes might be inefficient transiently during adaptation to a new environment [267]. The costs associated with biosynthesis of amino acids constitute an important part of the metabolic costs that are incurred by a bacterial cell [88] and the results in this thesis show that both, metabolic and protein costs can determine growth rates (chapter 3) and fitness (chapter 1) of bacterial cells. Thus, by a loss of these functions, bacterial strains can be freed in part from growth related constraints and distribute the resources that were earlier associated with these functions to other important cellular functions. This argument is in line with the 'streamlining theory' which has been proposed to explain genome reduction in free-living organisms [268]. The basic idea is that selection favors a reduction in genome size in order to give rise to a cell architecture that minimizes the resources required for replication in such organisms [41,82,268]. However, what causes drive losses in endosymbiotic or host restricted bacteria?

In this study selection drove the loss of functions rather than a reduction of genome size (chapter 4). A similar result was observed in serially propagated populations of *Methylobacterium extorquens* AM1 wherein most of the evolved populations had lost *accessory genes* [91]. This result has an important implication in context of drift or selection in driving the loss functions and genes in bacteria. Based on population genetic models the propensity of drift to influence the fixation of an allele is high when $N_e s < 1$ (Box 1 in Introduction) [41,46,80]. This scenario, where $N_e s$ is less than unity, has been postulated to be applicable for endosymbionts because these bacteria have a low N_e. However, such models fail to

account for the possibility that if s is high enough, it can potentially result in a high $N_e s$ i.e. $N_e s < 1$, thus potentially over-riding the effects of low N_e.

Table 1: Relative abundances of amino acids in different environments. Amino acid abundances were weighted based on their concentrations in the total fraction amino acids in present in the environment. Table from Moura *et al*., (2013) [189].

Amino acid	Aquatic (mean ±SD)	Terrestrial (mean ±SD)	Host-associated (mean ±SD)	All environments (mean ±SD)
Ala	0.1167 ±0.05315	0.1144 ±0.01689	0.0602 ±0.00516	0.1113 ±0.04662
Arg	0.0378 ±0.03249	0.0459 ±0.03462	0.0504 ±0.01340	0.0409 ±0.03191
Asn+Asp	0.0896 ±0.04352	0.1435 ±0.04755	0.0618 ±0.00557	0.1009 ±0.04980
Cys	0.0005 ±0.00265	0.0015 ±0.00323	0.0351 ±0.00332	0.0037 ±0.01004
Glu+Gln	0.0883 ±0.04270	0.1696 ±0.13768	0.2039 ±0.01316	0.1187 ±0.08787
Gly	0.2002 ±0.11806	0.0951 ±0.04960	0.0781 ±0.00580	0.1633 ±0.11176
His	0.0134 ±0.02044	0.0328 ±0.03000	0.0199 ±0.00191	0.0189 ±0.02374
Ile	0.0375 ±0.02426	0.0266 ±0.00944	0.0349 ±0.00138	0.0345 ±0.02075
Leu	0.0604 ±0.02298	0.0439 ±0.02440	0.0723 ±0.00263	0.0572 ±0.02374
Lys	0.0139 ±0.02318	0.0356 ±0.02159	0.0333 ±0.00376	0.0210 ±0.02386
Met	0.0032 ±0.00587	0.0055 ±0.00211	0.0174 ±0.00149	0.0050 ±0.00627
Phe	0.0334 ±0.02062	0.0367 ±0.02188	0.0344 ±0.00219	0.0343 ±0.01995
Pro	0.0157 ±0.03960	0.0039 ±0.01650	0.1174 ±0.00959	0.0213 ±0.04465
Ser	0.1455 ±0.15386	0.0660 ±0.02238	0.0558 ±0.00245	0.1178 ±0.13120
Thr	0.0595 ±0.01956	0.0538 ±0.01834	0.0440 ±0.00337	0.0567 ±0.01884
Trp	0.0005 ±0.00129	0.0023 ±0.00549	0.0080 ±0.00037	0.0016 ±0.00359
Tyr	0.0179 ±0.01621	0.0563 ±0.05206	0.0126 ±0.00177	0.0272 ±0.03355
Val	0.0583 ±0.02954	0.0487 ±0.01737	0.0605 ±0.00371	0.0560 ±0.02583

In this study, a loss of biosynthetic genes resulted in enormous fitness increases gains, on average 13%, in the auxotrophic strains in the presence of metabolites, and in different carbon source environments (chapter 1,2 and 4). Since, selection coefficients and fitness are interlinked [236], a plausible hypothesis is that if the fitness effect of the mutation is high enough, it can increase s and cause selection to influence the fixation of the mutation in the population irrespective of the population size or bottlenecks. Environments containing symbionts or host associated bacteria are normally rich in amino acids such as eukaryotic cells [269], and insect [249] and mammalian guts [270] (Table 1). Thus, selective advantages of auxotrophy could easily manifest in such environments. Future work should scrutinize and experimentally test this hypothesis. Furthermore, an experimental estimation of N_e of most endosymbiotic bacteria is often not possible. However, some studies have attempted to estimate population sizes in symbionts of insects. For instance the number of *Candidatus* Streptomyces philanthi individuals in the European bee-wolf increase from ~970 cells after vertical transmission as a result of logistic growth in the offspring 3-5 days after transmission, to range in-between 10^7-10^8 [103]. Similarly, in stinkbugs, population sizes of the symbiotic bacterium *Candidatus* Ishikawaella capsulata are estimated to be around 2×10^7 cells after increasing from 1.9×10^6 cells during vertical transmission [271]. Interestingly, the

population and bottleneck sizes imposed on the evolving *E. coli* populations in chapter 4 (population sizes of 10^7- 10^8 and bottleneck size of 1,000 cells) were strikingly similar to the above mentioned estimates for symbionts in their hosts, and yet adaptive processes were operational in driving the loss of biosynthetic functions. Thus, in the case of endosymbionts, a relatively small *Ne* combined with the enormous selective benefits (*s*) can also account for loss of biosynthetic functions in bacteria. Additionally, it is difficult to reconcile the frequently observed reciprocal losses of metabolic functions between hosts and bacteria [72,73] or between multiple endosymbionts [32,66,69] in the same host by genetic drift rather than a consistent period of adaptive coevolution. However, the possibility that other less costly or neutral functions are lost from bacterial genomes as a result of drift cannot be ruled out.

If the loss of biosynthetic functions is an evolutionary adaptation to the environment, how does a loss of metabolic autonomy help auxotrophic genotypes to colonize or exploit their environments? For instance, in bacterial pathogens a reduction in biosynthetic functions could result in selective fitness benefits that could help in infecting and proliferating within their animal hosts. The pathogen *L. pneumophilia*, the causative agent of Legionnaires disease, is auxotrophic for multiple amino acids and derives these metabolites by triggering degradation of proteins in its eukaryotic hosts by using a virulence factor [272]. Thus an acquisition of amino acids from the host cytoplasm helps this organism to successfully establish an intracellular infection [272]. Auxotrophies also exist in other microbial pathogens like *Rickettsia* and *Mycoplasma* [47,48] and could reflect general nutrient adaptations in such organisms and thus potentially represent an important force for pathogenic evolution [272].

The loss of metabolic autonomy in one bacterial genotype can also be an evolutionary adaptation in response to the presence of other bacterial genotypes in the environment (chapter 4) [175]. Such a process of 'compensated trait loss', where functions are lost due to provisioning of resources by ecological interaction [175,273], is also envisioned in the BQH and could be relevant in microbial communities where multiple genotypes are present and interact with each other. An interesting case of such a pattern of 'compensated trait loss' in bacteria is that of an evolved syntrophic mutualism between *Desulfovibrio vulgaris* and *Methanococcus maripaludis* [274,275]. Both the ancestral genotypes are metabolically autonomous when grown on certain substrates but in the conditions of the experiment both genotypes were required to be dependent on each other [274,275]. Over the course of syntrophic growth both species evolved to be increasingly metabolically dependent on each other and increase their stability and productivity. Interestingly, in independently evolved populations, *D. vulgaris* genotypes arose that had accumulated adaptive loss of function mutations in genes essential for metabolic autonomy in alternate environments resulting in an obligate dependency of these genotypes on *M. maripaludis* [274].

Thus, a significant outcome of such 'compensated trait losses' is the increased potential for diverse genotypes to develop interdependencies on each other [7,51,274]. A question thus is, how can initially metabolically autonomous genotypes diversify into being metabolically specialized and dependent in response to activities of other genotypes? I address possible ecological mechanisms in the following section.

1.3 The origins of diversity and mechanisms of coexistence and dependency in bacterial communities

Adaptive radiations have been often attributed to account for the widespread biological diversity existent in nature [276-278]. Simply put, *adaptive radiation* is the differentiation of a single ancestral type into different groups that can have distinct ecological roles and role-specific adaptations [278,279]. Often such *radiation* events can be a consequence of interactions, which can be frequency dependent, between individuals of a population and result in *adaptive diversification* of the ancestral lineage [280,281]. Such diversification events which lead to the coexistence of multiple bacterial types that originated from a single ancestral strain have been a frequent outcome of many selection experiments with microbes in *sympatric*, i.e. well mixed conditions [276,277,281-283]. Interestingly, a similar pattern of *sympatric adaptive diversification* was also observed in the evolution experiment in the non AA-regime in chapter 4. The prototrophic ancestral strain diversified into two distinct variants, one of which was still prototrophic while the other was auxotrophic for multiple amino acids (chapter 4). A likely explanation for the emergence and continued existence of auxotrophic genotypes is that changes (genetic or metabolic) in the evolving prototrophic sub-population provided a new niche i.e. *niche construction*, e.g. by secretion of amino acids, for the auxotrophic sub-population to fill. The two genotypes had distinct growth profiles and their coexistence of the auxotrophic genotypes was maintained by negative frequency-dependent selection (chapter 4). Thus, *niche construction* by one genotype facilitated the evolution of a dependent genotype. Two experimental studies have documented strikingly similar patterns of *adaptive diversification* events in terms of the metabolic capabilities of bacteria.

In one study, *E.coli* populations founded from a single genotype repeatedly evolved two distinct metabolic phenotypes when propagated in medium containing glucose [281,284]. The *E. coli* cells that were evolving to the glucose medium excreted acetate into the environment, thus creating a new *niche* which another sub-population evolved to fill [281,284]. Thus, two genetically distinct populations, which evolved from the same ancestor, specialized on two distinct components of the environment, glucose and acetate [281,284]. The glucose specialized genotypes grow faster when glucose is abundant but inefficiently utilize acetate [281,284]. In contrast, the acetate specialized genotypes catabolize acetate efficiently [281,284]. Both, these genotypes are maintained by negative frequency-dependent

selection, which is generated because of a limitation of resources due to the metabolic specialization of each genotype and daily transfers [281,283,284]. In another study involving a long term evolution experiment with *E.coli* populations that were adapting to glucose, an interesting case of what can be thought of as diversification was observed [285]. In one of the twelve adapting populations of *E. coli*, the ability to utilize citrate (cit^+ cells) as a carbon source emerged [285]. These cit^+ genotypes did not completely eliminate the cit^- genotypes (which display the ancestral phenotype) and instead coexist with cit^- cells in a frequency-dependent manner [285]. The basis for their coexistence was that cit^+ genotypes secreted carbon sources such as succinate, fumarate, and malate into the environment which benefitted cit^- genotypes since these types could catabolize these compounds [285].

These experimental outcomes signify the importance of *niche construction* by a certain bacterial genotype, for instance through metabolic changes, has on the evolution of metabolic specialized genotypes. It has been suggested that *adaptive diversification* can be a key driver of speciation in microbial communities [286]. However, it can be also argued that natural environments are exceptionally complex in terms of the temporal and spatial variations in nutrient compositions and thus, *adaptive diversification* observed in consistent environments in the laboratory experiments, such as in chapter 4 and those mentioned above [281,284,285], would not necessarily proceed in a similar fashion in nature as in the environment [286]. However, the ease with which rapid specialization occurs in experimental populations repeatedly suggests that even small periods of environmental stability can result in *adaptive diversification* of bacteria in the nature [286].

One important question thus begs an explanation. How can *niche construction* occur in microbial communities? A feature of microbial growth called 'overflow metabolism' can provide the likely explanation. Many bacterial genotypes under certain conditions of growth can excrete large quantities of metabolic waste or by-products into their external environment [287]. This process is termed as overflow metabolism [287] and results in *public goods* (or alternately BQ functions in this context) like metabolites, carbon sources, enzymes, siderophores being available for exploitation by other genotypes in the environment [26,28]. The primary reasons of such metabolic export by a bacterial cell is to maintain the balance of the metabolic system as well as to maintain the chemical and physical integrity of the metabolic network in a cell [28]. However, the presence of such *public goods* can favor *adaptive diversification* like in case of the experimental observations [281,284,285] mentioned earlier or alternately favor distinct genotypes, for e.g. auxotrophic types, to utilize these metabolites. Perhaps this could also explain why prototrophic and auxotrophic strains of *Pseudomonas aeruginosa* that descended from the same ancestral genotype often inhabit the same environments, i.e. the exudates of cystic fibrosis infected patients [251,252,288] or why co-

occurrences of auxotrophic and prototrophic genotypes have also been noted in freshwater microbial communities [171,255].

Nonetheless, an important consequence of diversification would be the development of metabolic dependencies between different genotypes on each other. In the examples of adaptive diversification, the metabolic dependence was such that only one of the two genotypes is completely reliant on the other, and not *vice versa*. For instance, acetate specialists are dependent on the glucose specialists [284], auxotrophic genotypes on prototrophic genotypes (chapter 4), and cit^- genotypes are dependent on cit^+ genotypes [285]. However, most cases of metabolic dependencies are in the form of intricate bi-directional metabolic complementarities such as those observed in the case of coevolving bacterial consortia which exchange nutrients, e.g. syntrophy, [2,8,274] or coevolving endosymbionts within eukaryotic hosts which display striking complementarity in their metabolic pathways [32,47,49,66,69,274]. How can such metabolic interdependencies originate in nature? I discuss possible routes to intricate metabolic associations in the next section.

1.4 From adaptive loss of functions to inter-bacterial metabolic complementarily in microbial communities

The prevalent loss of functions coupled with by-product utilization can result in intricate and obligate interdependencies between partners in coevolving consortia [51,100]. However, classical models of mutualisms predict that the evolution of such interdependencies between two partners through by-product reciprocity would be shaped by natural selection such that receivers of by-products maximize their benefits by being cooperative to producers [7,100]. Sachs and Hollowell (2012), and Kost (2015) have postulated a possible path to the evolution of interdependencies in bacteria which are in principal similar to the predictions of the BQH (Figure 2) [51,100,289]. The first step is the selfish use by one species of by-products secreted into the environment by another species [100]. The second step will be the loss of these costly functions that are now gained from the environment [100]. A last step will then be the evolution of costly cooperative traits in order to maximize the functions that are produced by others [100]. The first two steps have been explained so far in the context of the *adaptive diversification* (section 6.3) and the adaptive loss of functions (section 6.2). How can the third step i.e. the production of costly functions come about?

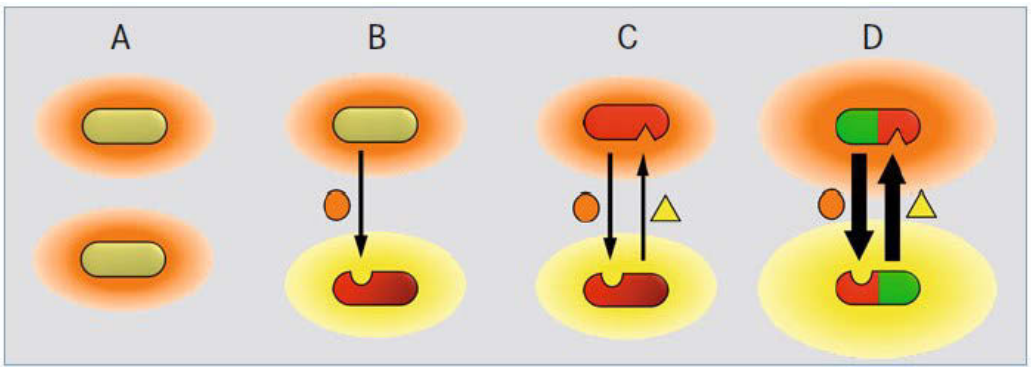


Figure 2: Steps towards the evolution of reciprocal metabolic dependencies in bacteria from initially metabolically autonomous genotypes. (A) Bacteria release metabolites into the environment (orange circle). (B) The presence of these metabolites in the environment results in the adaptive loss of metabolic autonomy in another species or subpopulation (red cells). (C) Bacteria which have lost functions can in turn produce other metabolites (yellow) which can result in the loss of metabolic autonomy in another population or species. (D) Reciprocal feedbacks can increase the strength of the interaction and give rise to intricate metabolic dependencies between the two cell types. Figure from Kost (2015) [289].

A recently proposed theory called the biological general equilibrium theory (BGET) attempts to provide a framework to explain the evolution of interdependencies (Figure 3) [29]. This theory is based on the general equilibrium theory (GET) that is used in economics to explain the behavior of trade and markets [29]. The basic premise of the BGET framework is that microbial communities will be more productive in terms of growth when their constituent members lose biosynthetic functions and enter into inter-cellular exchange compared to a scenario where the constituent members of a community are completely prototrophic [29]. This is because natural selection optimizes metabolite production and growth in specific environments by favoring specialization [29]. As a result, genotypes, which are already endowed with productivity differences and are functionally distinct can come together to function as a single *metabolic unit* [29]. Thus a key requirement for metabolic reciprocity is specialization in metabolite production. A question then is how such specialization can occur in bacteria.

Findings in chapter 3, strongly suggest that the structure of metabolic network in *E. coli* results in a trade-off in the biosynthesis of amino acids. Briefly, amino acids, which are synthesized from glycolytic precursors, entail lower metabolic costs if glycolytic carbon sources, e.g. fructose, are present as compared to costs when a gluconeogenic carbon source is used (e.g. succinate). In contrast, amino acids which have their precursors in the tricarboxylic acid cycle, induce lower metabolic costs when gluconeogenic carbon sources are present compared to growth on glycolytic carbon sources. This trade-off probably also accounts for why distinct epistatic patterns emerged in the fructose and succinate environments (chapter 2). Metabolic trade-offs in bacteria can result from biochemical

conflicts resulting from distinct metabolic processes competing for resources like energy (ATP), bio-elements (carbon and nitrogen) or core cellular machinery like metabolic enzymes and protein synthesis [226]. These resources are limiting in a bacterial cell and if a cell invests more in one metabolic process, it can invest less resources in another metabolic process [226]. Importantly, such biochemical trade-offs can promote metabolic specialization wherein specialized cell types that can only perform a certain set of metabolic processes optimally, can emerge [226]. The evolution of acetate specialists described in section 6.3 is an example of the evolution of metabolic specialists as a result of trade-offs in central carbon metabolism in bacteria [281,282,284].

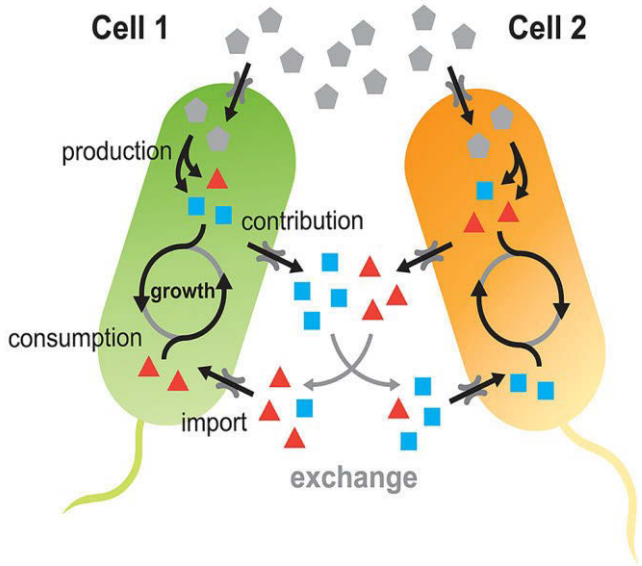


Figure 3: Evolution of metabolic exchanges between metabolically specialized microbial cell-types under the BGET framework. Microbes can convert catabolites (grey pentagons) to fuel the biosynthesis of metabolites required for synthesis of essential cellular macromolecules. Instead of performing all of the biosynthetic reactions in one cell-type, cells can produce metabolites by splitting these functions between different cells. Thus, cell 1 can specialize on the production of a single metabolite (blue squares) while cell 2 produces more of another metabolite (red triangles). Export of these anabolites into the environment can lead to a reciprocal exchange of these compounds, helping such specialized cell types to cooperate and maximize individual growth rates. Figure from Tasoff *et al.* (2015) [29].

Thus, as result of metabolic trade-offs, dependent genotypes will specialize on the production of only certain sets of metabolites, which they can exchange and thus maximize their productivity (Figure 3) [29]. This idea is supported by empirical work, which has shown that two or more auxotrophic genotypes can reciprocally exchange amino acids to complement each other's metabolic needs [20,23,24,28,126,254,290] and that such metabolic exchange can be favored by natural selection [24,290]. For instance, Pande and coworkers

(2014) constructed synthetically engineered consortia composed of two *E.coli* strains, each of which was auxotrophic for a certain amino acid while also overproducing a different set of amino acids [24]. Thus, these genotypes could metabolically complement each other's amino acid deficiencies by reciprocally exchanging amino acids [24]. The authors observed that in general cross-feeding consortia had a pronounced fitness advantage over prototrophic populations [24]. In another study, Zhang and Reed studied the dynamics of a cross-feeding consortia composed of two auxotrophic strains over the course of time found that these consortia adaptively evolved to increase their growth rates up to 3 fold compared to the ancestral consortia [290].

An important factor or problem for the evolution of interdependencies in bacteria is often the requirement for dependent bacteria to always have available genotypes which can provision metabolites that are required [13]. Spatial structure is a likely solution to this problem and has also been frequently suggested to favor the evolution of mutualisms [7,291,292]: by being in close spatial proximities of provisioning genotypes, dependent strains can derive the metabolites that have been released into the external environment by other genotypes [5,7,11,136,291-293]. Since a prominent mode of growth in the microbial world is as biofilm communities [294-298], such spatial structure be can easy to achieve in nature and helps dependent genotypes to be in close vicinities of each other. Estrella and coworkers have also addressed the factors that determine the transition towards mutual independencies (2015) [111]. They found that the level of privatization i.e. the extent of the retention of metabolic functions, by the providing genotype (e.g. a prototroph) of a BQ function (such as a metabolite) determines the feasibility of such a transition. When the level of privatization by a provider genotype is high, coexistence and mutual interdependencies are strongly disfavored whereas low to intermediate levels of metabolite privatization strongly promote the evolution of metabolic interdependencies [111]. The aforementioned spatial structure is one way how cells of a provider genotype can privatize the supply of metabolites. For instance, it has been shown that growth of cooperating genotypes of yeast [299] and bacteria [300] that reciprocally exchange amino acids in a spatial structured environment, is often restricted to regions where cooperating genotypes are present. In the bacterial consortia it was also observed that cooperating genotypes were localized in metabolite-rich regions [300]. The argument of spatial structure has also been used to explain enhanced metabolic activity at syntrophic interfaces in a consortium of methanogens [27] and sulfate reducing bacteria isolated from nature [301]. Cooperating genotypes can also exchange metabolites through *nanotubes* [253,254]. For instance, it was recently discovered in two different studies that strains can alleviate their metabolic needs by deriving nutrients from producer genotypes through membrane derived extensions which not only mediate cell-cell contact but also facilitate the transport of nutrients across cells [253,254]. Such metabolite exchange through

nanotubes could be another way how strains can create spatially structured environments and potentially privatize metabolites [254].

An important question that arises at this point is how can metabolic complementarities emerge in natural communities of bacteria? Consider a hypothetical scenario of a microbial community in nature existing as a biofilm (Figure 4). This community has access to several utilizable metabolites and catabolic resources to enable the growth and survival of the genotypes that comprise this community in nature [302,303]. In addition, community members also produce metabolites and these are released into the external environment. As a result of the adaptive loss of metabolic autonomy and specialization in the production of metabolites, intricate metabolic complementarities can emerge between members of the community (Figure 4). Furthermore, genotypes can also modify their metabolic environment resulting in diversification into newer metabolically distinct species. Nonetheless, repeated bouts of coevolution, and increased investments by each partner into the cooperative exchange, should result in loss of functions in both partners, due the enormous fitness benefits of losing biosynthetic functions [51,100]. Over time, this BQH-like process should result in biosynthetic losses and give rise to complementarities at the level of metabolic functions in the coevolving genotypes similar to those observed in bacteria isolated from nature [26]. Thus, a simple process of adaptive gene loss can bring about remarkable changes in the structure, and possibly function of microbial communities in nature.

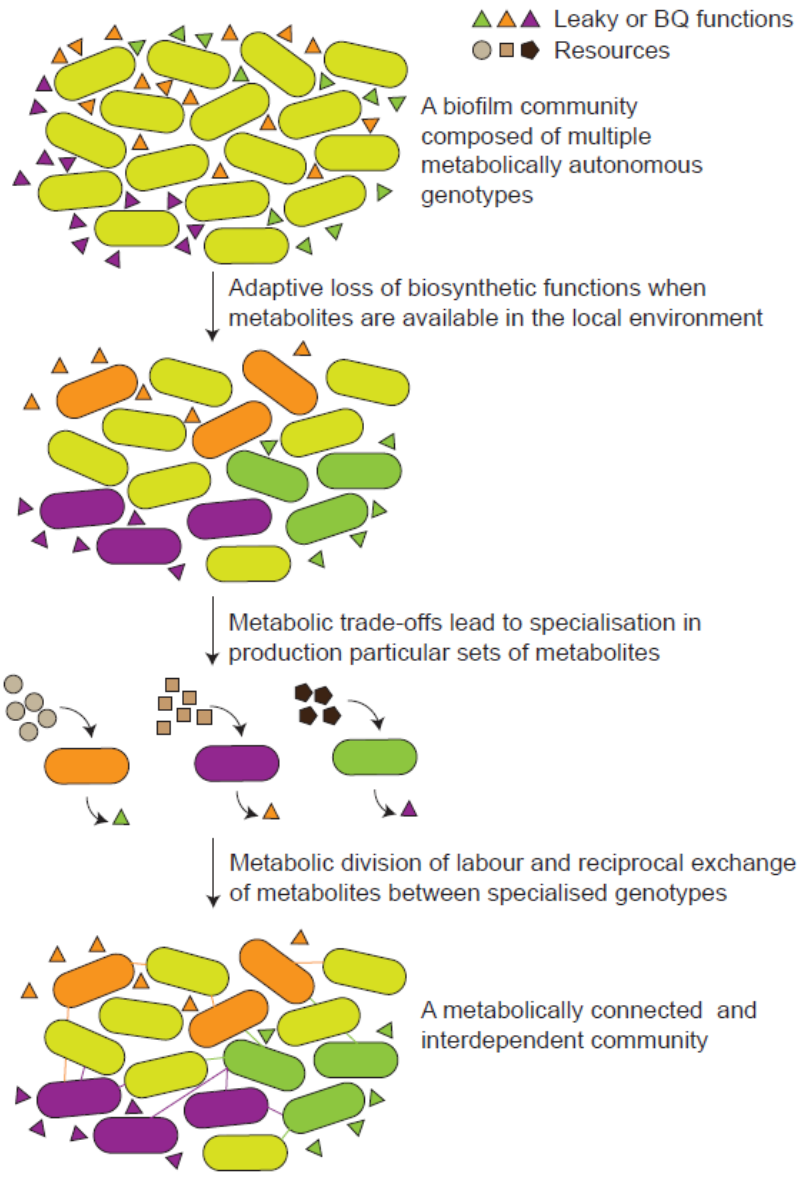


Figure 4: Evolution of inter-bacterial metabolic complementarities in microbial communities. Natural selection favors the loss of biosynthetic autonomy from initially prototrophic genotypes (yellow cells) in the presence of anabolites (colored triangles) in the environment, thus causing these genotypes to instead depend on the metabolic environment or other genotypes to compensate for a loss of these functions. However, prototrophic genotypes will not be selected against but are maintained in the population. The dependent genotypes can then diversify to utilize specific subsets of the available catabolic resources (circles, squares or pentagons) to maximize their fitness, while metabolic trade-offs will lead to the specialization in production of specific metabolites. This functional specialization can result in the evolution of interdependencies (colored lines) between different specialized genotypes, which could be connected through *nanotubes*, resulting in the formation of a single connected functional *metabolic unit*. Orange, purple and green cells represent different auxotrophic genotypes. Colored lines represent physical cell-cell connections that mediate metabolic exchanges between diverse cell types.

Can such processes be implicated in the evolution of natural communities? A recent metagenomic and empirical survey of ecologically relevant methanogenic communities in a controlled laboratory environment [304] strongly indicates that the answer is yes. The authors observed that these communities are tightly integrated in terms of catabolic interactions between different members [304]. However, most members in the community occupied similar metabolic *niches* but still coexisted with each other, contrary to expectations [304]. For instance, in one of the communities which catabolizes hexadecane, *Desulfovibrio* was present, which was surprising since sulfate reduction, a core attribute of this genus, was not detected in this community [304]. Interestingly, based on an analysis of transcriptional activity, members of the community were found to have distinct patterns of amino acid auxotrophies, such that many members selective express only certain sets of biosynthesis pathways when catabolizing different catabolic compounds (Figure 5) [304]. For instance, *Desulfovibrio* is the only member of the hexadecane community which can provision methionine and proline to other genotypes, in community catabolizing caprylate and hexadecane respectively (Figure 5) [304]. Interestingly, only certain members of the community, e.g. *Smithella* and *Syntrophomonas* provide the more biosynthetically expensive amino acids to other community members [304]. The differential provisioning of different amino acids by distinct genotypes in different catabolic conditions strengthens the key predictions made in chapter 3 as well as in the BGET that the type of catabolic resource and metabolic trade-offs shape the specialization in the type of metabolites that can be produced. The authors posit that the factors influencing community *structure* and *function* could extend beyond catabolism to include the importance of amino acid dependencies in nature [304]. A similar distribution of biosynthetic functions has also been

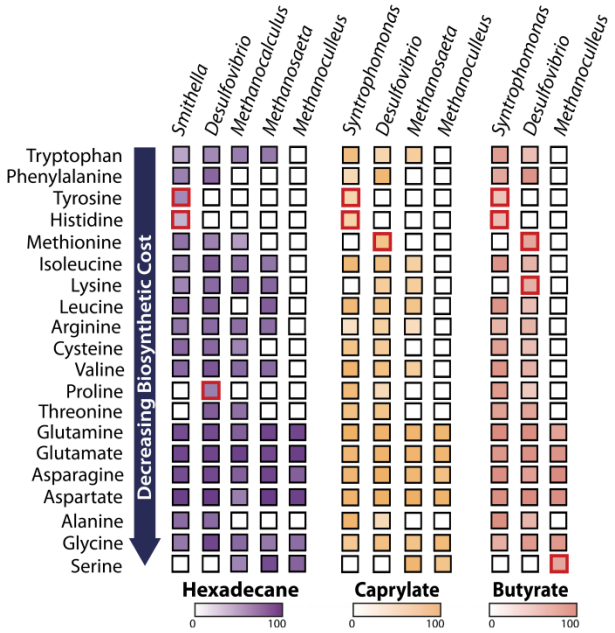


Figure 5: Biosynthetic functions are distributed between species in methanogenic communities. The amino acids are ordered by their biosynthetic cost (arrow). A colored square indicates the amino acid that a species can produce, and the intensity of color represents the relative expression of the corresponding amino acid biosynthesis pathway. Those amino acids exclusively produced by one species are highlighted in red. Figure from Embree *et al.*, (2015) [304].

observed in a freshwater microbial community [255]. The aforementioned multiple cases of genome reduced insect endosymbionts are another striking example of separate organisms that function as a single *metabolic unit* [305]. Dependencies have likely driven the extreme integration of genomes and metabolism of host and endosymbionts [63,73] or of multiple endosymbionts [69,177], even causing the diversification of endosymbionts in certain cases [305,306]. For instance, based on the genomic capabilities, the endosymbionts of Glassy-Winged Sharpshooter, *Sulcia muelleri* and *Baumannia cicadellinicola*, have been predicted to have evolved multiple complementarities in their biosynthetic functions (Figure 6) [69]. The predictions of such interdependencies in these organisms fit nicely with the outcomes from this thesis.

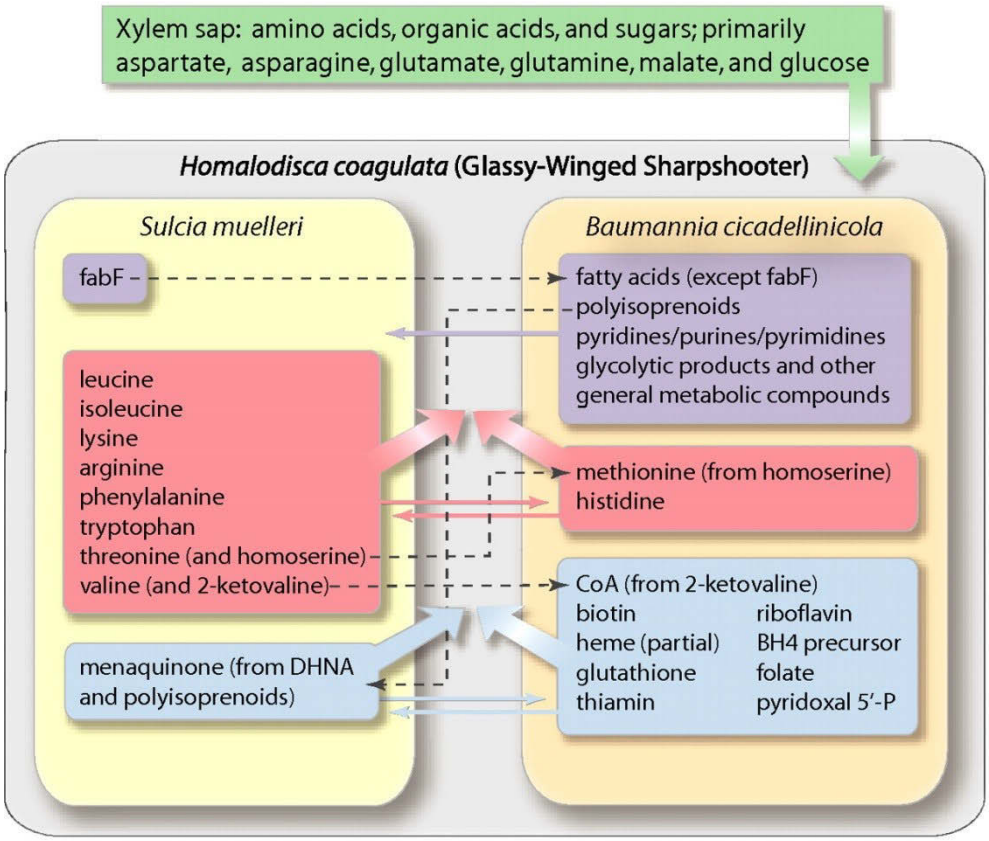


Figure 6: Metabolic interdependencies predicted in *Sulcia* and *Baumannia*. Some metabolites can be gained from the xylem sap (green boxes). Metabolites produced by endosymbionts which can be used by the host are indicated using large arrows while those that are shared between endosymbionts are indicate with small arrows. Metabolites, pathways, and genes in red and blue boxes are involved in amino acid and vitamin biosynthesis, respectively. Purple boxes indicate various other metabolic functions whereas dashed arrows indicate processes and genes shared between the symbionts. Figure from McCutcheon and Moran (2007) [69].

It is well known that cross-feeding interactions can stabilize genetic polymorphisms in bacteria [227,307-309] but as evidenced in the above example [304] and also in this study, adaptive metabolic interdependencies can in fact also have tremendous impacts on the

potential to drive the evolution of structure and function in microbial communities. These findings strongly makes the case for the microbial community to act as one integrated *metabolic unit*, whereby the entire set of metabolic functions may be distributed in parts amongst different community members, to enhance the productivity and stability of such communities. In the future, a targeted sampling of natural communities for their metabolic abilities and empirical studies of their metabolic activities should reveal the actual scale and function of metabolic interdependencies in nature.

1.5 Concluding remarks

The insights achieved from the theoretical and empirical findings of this dissertation strongly suggest that the natural selection strongly favors the loss of biosynthetic functions in bacteria across diverse lifestyles. These losses can potentiate the evolution of interdependencies between members of a microbial community because the adaptive benefits associated with being interdependent strongly outweigh the costs of employing a metabolically autonomous strategy. In addition, bacteria can also induce metabolic changes in environments causing *adaptive radiations* and the evolution of new metabolic phenotypes, and ecological interactions. Such intricate metabolic interdependencies could be a general feature of microbial life within communities whereby the entire bacterial collective functions as one *metabolic unit* (Figure 3 and 4). A great number of bacteria are simply uncultivable in laboratory environments unless co-cultured with strains which complement their metabolic needs [125,310]. This difficulty with cultivating most bacteria could simply be a general reflection of the extent of metabolic losses and interdependencies in nature. Perhaps, joining forces and accomplishing tasks that are otherwise energetically unfavorable if performed individually [1,311] is the norm in microbial communities in nature.

References:

1. Schloss PD, Handelsman J (2005) Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome biology 6: 229.
2. Little AE, Robinson CJ, Peterson SB, Raffa KF, Handelsman J (2008) Rules of engagement: interspecies interactions that regulate microbial communities. Annu Rev Microbiol 62: 375-401.
3. Faust K, Raes J (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10: 538-550.
4. Phelan VV, Liu WT, Pogliano K, Dorrestein PC (2012) Microbial metabolic exchange--the chemotype-to-phenotype link. Nat Chem Biol 8: 26-35.
5. Estrela S, Trisos CH, Brown SP (2012) From Metabolism to Ecology: Cross-Feeding Interactions Shape the Balance between Polymicrobial Conflict and Mutualism. American Naturalist 180: 566-576.
6. Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, et al. (2010) The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res 38: 3857-3868.
7. Sachs JL, Mueller UG, Wilcox TP, Bull JJ (2004) The evolution of cooperation. The Quarterly Review of Biology 79: 135-160.
8. Seth EC, Taga ME (2014) Nutrient cross-feeding in the microbial world. Front Microbiol 5: 350.
9. Cremer J, Melbinger A, Frey E (2012) Growth dynamics and the evolution of cooperation in microbial populations. Sci Rep 2: 281.
10. Nadell CD, Drescher K, Wingreen NS, Bassler BL (2015) Extracellular matrix structure governs invasion resistance in bacterial biofilms. ISME J 9: 1700-1709.
11. Nadell CD, Bucci V, Drescher K, Levin SA, Bassler BL, et al. (2013) Cutting through the complexity of cell collectives. Proceedings of the Royal Society of London B: Biological Sciences 280: 20122770.
12. Tan J, Zuniga C, Zengler K (2015) Unraveling interactions in microbial communities - from co-cultures to microbiomes. J Microbiol 53: 295-305.
13. Oliveira NM, Niehus R, Foster KR (2014) Evolutionary limits to cooperation in microbial communities. Proc Natl Acad Sci U S A 111: 17941-17946.
14. Hibbing ME, Fuqua C, Parsek MR, Peterson SB (2010) Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology 8: 15-25.
15. Foster KR, Bell T (2012) Competition, not cooperation, dominates interactions among culturable microbial species. Curr Biol 22: 1845-1850.

16. Fuhrman JA (2009) Microbial community structure and its functional implications. Nature 459: 193-199.
17. Marx CJ (2009) Getting in Touch with Your Friends. Science 324: 1150-1151.
18. Lotem A, Fishman MA, Stone L (1999) Evolution of cooperation between individuals. Nature 400: 226-227.
19. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, et al. (2011) Competitive and cooperative metabolic interactions in bacterial communities. Nature communications 2: 589.
20. Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Mol Syst Biol 6: 407.
21. Biggs MB, Medlock GL, Kolling GL, Papin JA (2015) Metabolic network modeling of microbial communities. Wiley Interdiscip Rev Syst Biol Med 7: 317-334.
22. Mee MT, Collins JJ, Church GM, Wang HH (2014) Syntrophic exchange in synthetic microbial communities. Proceedings of the National Academy of Sciences 111: E2149–E2156.
23. Harcombe W (2010) Novel cooperation experimentally evolved between species. Evolution 64: 2166-2172.
24. Pande S, Merker H, Bohl K, Reichelt M, Schuster S, et al. (2014) Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME J 8: 953–962.
25. Estrela S, Morris JJ, Kerr B (2015) Private benefits and metabolic conflicts shape the emergence of microbial interdependencies. Environ Microbiol.
26. Morris JJ (2015) Black Queen evolution: the role of leakiness in structuring microbial communities. Trends in Genetics 31(8): 475 - 482
27. McGlynn SE, Chadwick GL, Kempes CP, Orphan VJ (2015) Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature 526: 531-535.
28. Campbell K, Vowinckel J, Muelleder M, Malmsheimer S, Lawrence N, et al. (2015) Self-establishing communities enable cooperative metabolite exchange in a eukaryote. eLife.
29. Tasoff J, Mee MT, Wang HH (2015) An Economic Framework of Microbial Trade. PloS one 10: e0132907.
30. Doolittle WF (1999) Phylogenetic Classification and the Universal Tree. Science 284: 2124-2128.
31. Hacker J, Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity. EMBO reports 2: 376-381.
32. McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10: 13-26.

33. Han K, Li Z-f, Peng R, Zhu L-p, Zhou T, et al. (2013) Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. Scientific Reports 3: 2101.
34. Snel B, Bork P, Huynen MA (2002) Genomes in flux: the evolution of archaeal and proteobacterial gene content. Genome research 12: 17-25.
35. Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification 1. Annual Reviews in Microbiology 55: 709-742.
36. Puigbò P, Lobkovsky A, Kristensen D, Wolf Y, Koonin E (2014) Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. BMC Biology 12: 1-19.
37. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, et al. (2012) Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade. Mbio 3.
38. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. Current Opinion in Genetics & Development 15: 589-594.
39. Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36: 6688-6719.
40. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, et al. (2006) Comparative genomics of the lactic acid bacteria. Proceedings of the National Academy of Sciences 103: 15611-15616.
41. Batut B, Knibbe C, Marais G, Daubin V (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. Nat Rev Microbiol 12: 841-850.
42. Bolotin E, Hershberg R (2015) Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. Genome Biology and Evolution.
43. Olson MV (1999) When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet 64: 18-23.
44. Batut B, Knibbe C, Marais G, Daubin V (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. Nat Rev Micro 12: 841-850.
45. Ochman H, Davalos LM (2006) The Nature and Dynamics of Bacterial Genomes. Science 311: 1730-1733.
46. Martínez-Cano DJ, Reyes-Prieto M, Martínez-Romero E, Partida-Martínez LP, Latorre A, et al. (2014) Evolution of small prokaryotic genomes. Frontiers in Microbiology 5: 742.
47. Ochman H, Moran NA (2001) Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. Science 292: 1096-1099.
48. Moran NA (2002) Microbial Minimalism: Genome Reduction in Bacterial Pathogens. Cell 108: 583-586.

49. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, et al. (2006) The 160-Kilobase Genome of the Bacterial Endosymbiont Carsonella. Science 314: 267.
50. Andersson SGE, Kurland CG (1998) Reductive evolution of resident genomes. Trends in Microbiology 6: 263-268.
51. Morris JJ, Lenski RE, Zinser ER (2012) The black queen hypothesis: Evolution of dependencies through adaptive gene loss. Mbio 3: e00036-00012.
52. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science 309: 1242-1245.
53. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, et al. (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 103: 15611-15616.
54. van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, et al. (2006) The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution. Proceedings of the National Academy of Sciences of the United States of America 103: 9274-9279.
55. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D (2009) Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4: 13.
56. Renesto P, Ogata H, Audic S, Claverie J-M, Raoult D (2005) Some lessons from Rickettsia genomics. FEMS microbiology reviews 29: 99-117.
57. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp. APS*. Nature 407: 81-86.
58. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, et al. (2004) Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. Nature Biotechnology 22: 1554-1558.
59. Chubukov V, Gerosa L, Kochanowski K, Sauer U (2014) Coordination of microbial metabolism. Nat Rev Micro 12: 327-340.
60. Wagner A (2012) Metabolic networks and their evolution. Adv Exp Med Biol 751: 29-52.
61. Gottschalk G (2012) Bacterial metabolism: Springer Science & Business Media.
62. Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. Nature 440: 667-670.
63. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp* APS. Nature 407: 81-86.
64. Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, et al. (2009) A fragile metabolic network adapted for cooperation in the symbiotic bacterium Buchnera aphidicola. BMC systems biology 3: 24.
65. Parter M, Kashtan N, Alon U (2007) Environmental variability and modularity of bacterial metabolic networks. BMC Evolutionary Biology 7: 169.

66. McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biology and Evolution 2: 708-718.

67. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, et al. (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences 110: 20338–20343.

68. Cases I, de Lorenzo V, Ouzounis CA (2003) Transcription regulation and environmental adaptation in bacteria. Trends in Microbiology 11: 248-253.

69. McCutcheon JP, Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proceedings of the National Academy of Sciences of the United States of America 104: 19392-19397.

70. Martínez-Cano DJ, Reyes-Prieto M, Martinez-Romero E, Partida-Martinez LP, Latorre A, et al. (2015) Evolution of small prokaryotic genomes. Frontiers in Microbiology 5.

71. Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences 93: 2873-2878.

72. Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3: 850-861.

73. Wernegreen JJ (2015) Endosymbiont evolution: predictions from theory and surprises from genomes. Annals of the New York Academy of Sciences:

74. Kuo C-H, Ochman H (2009) Deletional Bias across the Three Domains of Life. Genome Biology and Evolution 1: 145-152.

75. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. Trends in Genetics 17: 589-596.

76. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, et al. (2005) Bacterial genome size reduction by experimental evolution. Proceedings of the National Academy of Sciences of the United States of America 102: 12112-12116.

77. van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, et al. (2003) Reductive genome evolution in *Buchnera aphidicola*. Proceedings of the National Academy of Sciences 100: 581-586.

78. Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, et al. (2009) A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. BMC Syst Biol 3: 24.

79. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. Annual Review of Genetics. pp. 165-190.

80. Gillespie JH (2010) Population genetics: a concise guide: JHU Press.

81. Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. PLoS Genet 4: e1000304.
82. Dufresne A, Garczarek L, Partensky F (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biology 6.
83. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A 110: 11463-11468.
84. Mendonça AG, Alves RJ, Pereira-Leal JB (2011) Loss of genetic redundancy in reductive genome evolution. PLoS Comput Biol 7: e1001082.
85. Zamenhof S, Eichhorn HH (1967) Study of microbial evolution through loss of biosynthetic functions- Establishment of defective mutants. Nature 216: 456-458
86. Dykhuizen D (1978) Selection for tryptophan auxotrophs of *Escherichia coli* in glucose-limited chemostats as a test of energy-conservation hypothesis of evolution. Evolution 32: 125-150.
87. Kaleta C, Schäuble S, Rinas U, Schuster S (2013) Metabolic costs of amino acid and protein production in *Escherichia coli*. Biotechnology journal 8: 1105-1114.
88. Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proceedings of the National Academy of Sciences 99: 3695-3700.
89. Wessely F, Bartl M, Guthke R, Li P, Schuster S, et al. (2011) Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. Molecular Systems Biology 7.
90. Shachrai I, Zaslaver A, Alon U, Dekel E (2010) Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. Molecular cell 38: 758-767.
91. Lee MC, Marx CJ (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet 8: e1002651.
92. Koskiniemi S, Sun S, Berg OG, Andersson DI (2012) Selection-driven gene loss in bacteria. Plos Genetics 8.
93. Handel A, Rozen DE (2009) The impact of population size on the evolution of asexual microbes on smooth versus rugged fitness landscapes. BMC Evol Biol 9: 236.
94. Arber W (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiology Reviews 24: 1-7.
95. Kreimer A, Borenstein E, Gophna U, Ruppin E (2008) The evolution of modularity in bacterial metabolic networks. Proceedings of the National Academy of Sciences 105: 6976-6981.
96. Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332: 1190-1192.

97. de Visser J, Cooper TF, Elena SF (2011) The causes of epistasis. Proceedings of the Royal Society B-Biological Sciences 278: 3617-3624.

98. Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF, et al. (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Molecular Ecology 21: 2058-2077.

99. Remold SK, Lenski RE (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nat Genet 36: 423-426.

100. Sachs JL, Hollowell AC (2012) The Origins of Cooperative Bacterial Communities. Mbio 3.

101. Cao H, Butler K, Hossain M, Lewis JD (2014) Variation in the fitness effects of mutations with population density and size in *Escherichia coli*. PloS one 9: e105369.

102. Savage DC (1977) Microbial ecology of the gastrointestinal tract. Annual Reviews in Microbiology 31: 107-133.

103. Kaltenpoth M, Goettler W, Koehler S, Strohm E (2010) Life cycle and population dynamics of a protective insect symbiont reveal severe bottlenecks during vertical transmission. Evolutionary Ecology 24: 463-477.

104. Wahl LM, Gerrish PJ, Saika-Voivod I (2002) Evaluating the Impact of Population Bottlenecks in Experimental Evolution. Genetics 162: 961-971.

105. Feldgarden M, Stoebel DM, Brisson D, Dykhuizen DE (2003) Size Doesn't Matter: Microbial Selection Experiments Address Ecological Phenomena. Ecology 84: 1679-1687.

106. Bosdriesz E (2015) Darwin's invisible hand.

107. Berkhout J, Bosdriesz E, Nikerel E, Molenaar D, de Ridder D, et al. (2013) How Biochemical Constraints of Cellular Growth Shape Evolutionary Adaptations in Metabolism. Genetics 194: 505-512.

108. Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. Molecular Systems Biology 5.

109. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436: 588-592.

110. Gonzalez-Cabaleiro R, Ofiteru ID, Lema JM, Rodriguez J (2015) Microbial catabolic activities are naturally selected by metabolic energy harvest rate. ISME J.

111. Estrela S, Morris JJ, Kerr B (2015) Private benefits and metabolic conflicts shape the emergence of microbial interdependencies. Environmental Microbiology: DOI: 10.1111/1462-2920.13028

112. Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER (2011) Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. PLoS One 6: e16805.

113. Morris JJ, Papoulis SE, Lenski RE (2014) Coexistence of evolving bacteria stabilized by a shared black queen function. Evolution 68: 2960-2971

114. Klasson L (2004) Evolution of minimal-gene-sets in host-dependent bacteria. Trends in Microbiology 12: 37-43.

115. Snitkin ES, Segrè D (2011) Epistatic Interaction Maps Relative to Multiple Metabolic Phenotypes. PLoS Genet 7: e1001294.

116. Jagdishchandra Joshi C, Prasad A (2014) Epistatic interactions among metabolic genes depend upon environmental conditions. Molecular BioSystems.

117. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332: 1193-1196.

118. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445: 383-386.

119. da Silva J, Coetzer M, Nedellec R, Pastore C, Mosier DE (2010) Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics 185: 293-303.

120. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, et al. (2009) Positive Epistasis Drives the Acquisition of Multidrug Resistance. PLoS Genet 5: e1000578.

121. Segre D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. Nat Genet 37: 77-83.

122. He XL, Qian WF, Wang Z, Li Y, Zhang JZ (2010) Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. Nature Genetics 42: 272-U120.

123. Benner SA, Sismour AM (2005) Synthetic biology. Nature Reviews Genetics 6: 533-543.

124. Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. Molecular systems biology 2.

125. Zomorrodi AR, Segre D (2015) Synthetic Ecology of Microbes: Mathematical Models and Applications. J Mol Biol.

126. Shou W, Ram S, Vilar JM (2007) Synthetic cooperation in engineered yeast populations. Proceedings of the National Academy of Sciences 104: 1877-1882.

127. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, et al. (2012) Experimental evolution. Trends in Ecology & Evolution 27: 547-560.

128. Vasi F, Travisano M, Lenski RE (1994) Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. American Naturalist: 432-456.

129. Garland T, Rose MR (2009) Experimental evolution: concepts, methods, and applications of selection experiments: University of California Press Berkeley; Los Angeles.
130. Travisano M (2001) Evolution: Towards a genetical theory of adaptation. Current Biology 11: R440-R442.
131. Fox JW, Lenski RE (2015) From Here to Eternity—The Theory and Practice of a Really Long Experiment. PLoS Biol 13: e1002185.
132. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407: 736-739.
133. Maughan H, Callicotte V, Hancock A, Birky CW, Nicholson WL, et al. (2006) The population genetics of phenotypic deterioration in experimental populations of *Bacillus subtilis*. Evolution 60: 686-695.
134. Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Molecular Systems Biology 5: 320-320.
135. Baart GJ, Martens DE (2012) Genome-scale metabolic models: reconstruction and analysis. Methods Mol Biol 799: 107-126.
136. Kim HU, Kim TY, Lee SY (2008) Metabolic flux analysis and metabolic engineering of microorganisms. Mol Biosyst 4: 113-120.
137. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Applied and Environmental Microbiology 60: 3724-3731.
138. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, et al. (2007) Metabolic modeling of a mutualistic microbial community. Molecular Systems Biology 3: 92-92.
139. Mazumdar V, Amar S, Segrè D (2013) Metabolic Proximity in the Order of Colonization of a Microbial Community. PLoS ONE 8: e77617.
140. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nature biotechnology 28: 245-248.
141. Schilling CH, Edwards JS, Letscher D, Palsson BØ (2000) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnology and bioengineering 71: 286-306.
142. Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Briefings in bioinformatics 10: 435-449.
143. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophysical journal 83: 1331-1340.
144. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. Journal of theoretical biology 213: 73-88.

145. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial 'pan-genome' (vol 102, pg 13950, 2005). Proceedings of the National Academy of Sciences of the United States of America 102: 16530-16530.

146. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, et al. (2008) The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E.coli* commensal and pathogenic isolates. Journal of Bacteriology 190: 6881-6893.

147. Karberg KA, Olsen GJ, Davis JJ (2011) Similarity of genes horizontally acquired by *Escherichia coli* and *Salmonella enterica* is evidence of a supraspecies pangenome. Proceedings of the National Academy of Sciences of the United States of America 108: 20154-20159.

148. Lukjancenko O, Wassenaar T, Ussery D (2010) Comparison of 61 Sequenced *Escherichia coli* Genomes. Microbial Ecology 60: 708-720.

149. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology 2.

150. Joyce AR, Reed JL, White A, Edwards R, Osterman A, et al. (2006) Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. Journal of Bacteriology 188: 8259-8271.

151. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, et al. (2009) Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. PLoS Genet 5: e1000344.

152. Moran NA, Plague GR (2004) Genornic changes following host restriction in bacteria. Current Opinion in Genetics & Development 14: 627-633.

153. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, et al. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461: 1243-U1274.

154. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, et al. (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 40: D742-D753.

155. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, et al. (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. Journal of Biological Databases and Curation.

156. Gómez-Valero L, Rocha EPC, Latorre A, Silva FJ (2007) Reconstructing the ancestor of *Mycobacterium leprae*: The dynamics of gene loss and genome reduction. Genome Research 17: 1178-1185.

157. Pagani I, Liolios K, Jansson J, Chen IMA, Smirnova T, et al. (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Research 40: D571-D579.

158. Vanstockem M, Michiels K, Vanderleyden J, Vangool AP (1987) Transposon mutagenesis of *Azospirillum brasilense* and *Azospirillum lipoferum* - Physical analysis of TN5 and TN5-MOB insertion mutants. Applied and Environmental Microbiology 53: 410-415.

159. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 27: 29-34.

160. Keseler I, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res 39: D583 - D590.

161. Thomason LC, Costantino N, Court DL (2007) *E. coli* genome manipulation by P1 transduction. Current protocols in molecular biology Chapter 1: Unit 1.17-Unit 11.17.

162. Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. American Naturalist 138: 1315-1341.

163. de Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A, et al. (2008) A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Molecular Systems Biology 4.

164. Tepper N, Shlomi T (2011) Computational Design of Auxotrophy-Dependent Microbial Biosensors for Combinatorial Metabolic Engineering Experiments. PLoS ONE 6: e16274.

165. Benjamini Y, Krieger AM, Yekutieli D (2006) Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93: 491-507.

166. Team RDC (2008) R: A Language and Environment for Statistical Computing.

167. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. Science 330: 1099-1102.

168. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. Proceedings of the National Academy of Sciences 110: 10039-10044.

169. Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, et al. Bacterial adaptation through loss of function.

170. Münster U (1993) Concentrations and fluxes of organic carbon substrates in the aquatic environment. Antonie van Leeuwenhoek 63: 243-274.

171. Ihssen J, Grasselli E, Bassin C, François P, Piffaretti J-C, et al. (2007) Comparative genomic hybridization and physiological characterization of environmental isolates indicate that significant (eco-)physiological properties are highly conserved in the species *Escherichia coli*. Microbiology 153: 2052-2066.

172. Boles BR, Thoendel M, Singh PK (2004) Self-generated diversity produces "insurance effects" in biofilm communities. Proceedings of the National Academy of Sciences of the United States of America 101: 16630-16635.

173. Giraud A, Matic I, Tenaillon O, Clara A, Radman M, et al. (2001) Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut. Science 291: 2606-2608.

174. Ensminger AW, Yassin Y, Miron A, Isberg RR (2012) Experimental evolution of *Legionella pneumophila* in mouse macrophages leads to strains with altered determinants of environmental survival. Plos Pathogens 8.

175. Ellers J, Kiers ET, Currie CR, McDonald BR, Visser B (2012) Ecological interactions drive evolutionary loss of traits. Ecology Letters 15: 1071-1082.

176. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270: 397-403.

177. McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Current Biology 21: 1366-1372.

178. Bernstein HC, Paulson SD, Carlson RP (2012) Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. Journal of Biotechnology 157: 159-166.

179. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological Reviews 59: 143-169.

180. Hugenholtz P, Goebel BM, Pace NR (1998) Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. Journal of Bacteriology 180: 4765-4774.

181. D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, et al. (2014) Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution 68: 2559–2570.

182. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, et al. (2014) Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. Genome Biol Evol 6: 741-753.

183. Luo H, Csűros M, Hughes AL, Moran MA (2013) Evolution of Divergent Life History Strategies in Marine Alphaproteobacteria. mBio 4.

184. Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabasi AL (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. Nature 427: 839-843.

185. Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science 312: 111-114.

186. Schenk MF, Szendro IG, Salverda MLM, Krug J, de Visser JAGM (2013) Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. Molecular Biology and Evolution 30: 1779-1787.

187. Flynn KM, Cooper TF, Moore FBG, Cooper VS (2013) The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. PLoS Genet 9: e1003426.

188. Verhelst N, Hatzinger R, Mair P (2007) The Rasch sampler. Journal of Statistical Software 20: 1-14.

189. Moura A, Savageau MA, Alves R (2013) Relative Amino Acid Composition Signatures of Organisms and Environments. PLoS ONE 8: e77319.

190. Bertels F, Merker H, Kost C (2012) Design and characterization of auxotrophy-based amino acid biosensors. Plos One 7.

191. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97: 6640-6645.

192. Zar JH (1999) Biostatistical analysis: Pearson Education India.

193. Berger D, Postma E (2014) Biased Estimates of Diminishing Returns Epistasis? Empirical Evidence Revisited. Genetics.

194. Kim W, Levy SB (2008) Increased Fitness of Pseudomonas fluorescens Pf0-1 Leucine Auxotrophs in Soil. Applied and Environmental Microbiology 74: 3644-3651.

195. Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, et al. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. Molecular microbiology 68: 1128-1148.

196. Paul BJ, Berkmen MB, Gourse RL (2005) DksA potentiates direct activation of amino acid promoters by ppGpp. Proceedings of the National Academy of Sciences 102: 7823-7828.

197. Braun PR, Al-Younes H, Gussmann J, Klein J, Schneider E, et al. (2008) Competitive inhibition of amino acid uptake suppresses chlamydial growth: involvement of the chlamydial amino acid transporter BrnQ. Journal of bacteriology 190: 1822-1830.

198. Zhuang K, Vemuri GN, Mahadevan R (2011) Economics of membrane occupancy and respiro-fermentation. Molecular systems biology 7: 500.

199. Leavitt RI, Umbarger H (1962) Isoleucine and valine metabolism in *Escherichia coli* xi. K-12 Valine Inhibition of the Growth of Escherichia coli Strain. Journal of bacteriology 83: 624-630.

200. Hosoda K, Suzuki S, Yamauchi Y, Shiroguchi Y, Kashiwagi A, et al. (2011) Cooperative adaptation to establishment of a synthetic bacterial mutualism.

201. Douglas A (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. Annual review of entomology 43: 17-37.

202. Moran NA (2007) Symbiosis as an adaptive process and source of phenotypic complexity. Proceedings of the National Academy of Sciences 104: 8627-8633.

203. Reddy T, Thomas AD, Stamatis D, Bertsch J, Isbandi M, et al. (2014) The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. Nucleic acids research: gku950.

204. Yoon SH, Han M-J, Jeong H, Lee CH, Xia X-X, et al. (2012) Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. Genome Biol 13: R37.

205. Marr AG (1991) Growth rate of *Escherichia coli*. Microbiological reviews 55: 316-333.

206. Koch AL (1997) Microbial physiology and ecology of slow growth. Microbiology and Molecular Biology Reviews 61: 305-318.

207. Goelzer A, Fromion V (2011) Bacterial growth rate reflects a bottleneck in resource allocation. Biochimica et Biophysica Acta (BBA)-General Subjects 1810: 978-988.

208. Zhao J, Shimizu K (2003) Metabolic flux analysis of *Escherichia coli* K12 grown on 13 C-labeled acetate and glucose using GC-MS and powerful flux calculation method. Journal of Biotechnology 101: 101-117.

209. Neidhardt FC, Ingraham JL, Schaechter M (1990) Physiology of the bacterial cell: a molecular approach.

210. Richmond RC (1970) Non-Darwinian evolution: a critique. Nature 225: 1025-1028.

211. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, et al. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Molecular systems biology 7: 535.

212. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protocols 6: 1290-1307.

213. D'Souza G, Waschina S, Kaleta C, Kost C (2015) Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes. Evolution 69: 1244-1254.

214. Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T (2012) Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. PLoS Comput Biol 8: e1002575-e1002575.

215. Snell EE (2009) Tryptophanase structure: Advances in enzymology and related areas of molecular biology 42: 287.

216. Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823.

217. Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4: 133-142.

218. Hartigan JA, Wong MA (1979) Algorithm AS 136: A k-means clustering algorithm. Applied statistics: 100-108.

219. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, et al. (2013) EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Research 41: D605-D612.

220. Swire J (2007) Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. Journal of molecular evolution 64: 558-571.

221. Savageau MA (1983) *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. American Naturalist: 732-744.

222. van Elsas JD, Semenov AV, Costa R, Trevors JT (2011) Survival of *Escherichia coli* in the environment: fundamental and public health aspects. The ISME journal 5: 173-183.

223. Pittard J, Yang J (2008) Biosynthesis of the aromatic amino acids. EcoSal Plus 3.

224. Im S, Davidson H, Pittard J (1971) Phenylalanine and tyrosine biosynthesis in *Escherichia coli* K-12: mutants derepressed for 3-deoxy-D-arabinoheptulosonic acid 7-phosphate synthetase (phe), 3-deoxy-D-arabinoheptulosonic acid 7-phosphate synthetase (tyr), chorismate mutase T-prephenate dehydrogenase, and transaminase A. Journal of bacteriology 108: 400-409.

225. Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K, et al. (2015) Proteome reallocation in *Escherichia coli* with increasing specific growth rate. Molecular BioSystems 11: 1184-1193.

226. Johnson DR, Goldschmidt F, Lilja EE, Ackermann M (2012) Metabolic specialization and the assembly of microbial communities. The ISME journal 6: 1985-1991.

227. Treves D, Manning S, Adams J (1998) Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. Mol Biol Evol 15: 789 - 797.

228. Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proceedings of the National Academy of Sciences 105: 7899-7906.

229. Hua Q, Joyce AR, Palsson BØ, Fong SS (2007) Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. Applied and environmental microbiology 73: 4639-4647.

230. Sabarly V, Bouvet O, Glodt J, Clermont O, Skurnik D, et al. (2011) The decoupling between genetic structure and metabolic phenotypes in *Escherichia coli* leads to continuous phenotypic diversity. J Evol Biol 24: 1559-1571.

231. Bailey JE (1991) Toward a science of metabolic engineering. Science 252: 1668-1675.

232. Masurekar PS (2008) Nutritional and engineering aspects of microbial process development. Natural Compounds as Drugs Volume I: Springer. pp. 291-328.

233. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, et al. (2012) Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade. mBio 3.

234. Kuo CH, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome complexity. Genome Research 19: 1450-1454.

235. Lee M-C, Marx CJ (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS genetics 8: e1002651-e1002651.

236. Dykhuizen DE (1990) Experimental Studies of Natural Selection in Bacteria. Annual Review of Ecology and Systematics 21: 373-398.

237. Grenier F, Matteau D, Baby V, Rodrigue S (2014) Complete Genome Sequence of *Escherichia coli* BW25113. Genome Announcements 2.

238. Deatherage DE, Barrick JE (2014) Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. Methods in molecular biology (Clifton, NJ) 1151: 165-188.

239. Howell D (2012) Statistical methods for psychology: Cengage Learning.

240. Fukuda R, Nagasawa-Fujimori H (1983) Mechanism of the rifampicin induction of RNA polymerase beta and beta' subunit synthesis in *Escherichia coli*. Journal of Biological Chemistry 258: 2720-2728.

241. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7: e46688.

242. Conrad TM, Frazier M, Joyce AR, Cho B-K, Knight EM, et al. (2010) RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. Proceedings of the National Academy of Sciences 107: 20500-20505.

243. Nishino K, Honda T, Yamaguchi A (2005) Genome-wide analyses of *Escherichia coli* gene expression responsive to the BaeSR two-component regulatory system. Journal of bacteriology 187: 1763-1772.

244. Williams MD, Ouyang TX, Flickinger MC (1994) Starvation-induced expression of SspA and SspB: the effects of a null mutation in sspA on *Escherichia coli* protein synthesis and survival during growth and prolonged starvation. Molecular microbiology 11: 1029-1043.

245. Hansen AM, Qiu Y, Yeh N, Blattner FR, Durfee T, et al. (2005) SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli*. Molecular microbiology 56: 719-734.

246. Atlung T, Ingmer H (1997) H-NS: a modulator of environmentally regulated gene expression. Molecular microbiology 24: 7-17.

247. Cho B-K, Federowicz S, Park Y-S, Zengler K, Palsson BØ (2012) Deciphering the transcriptional regulatory logic of amino acid metabolism. Nature chemical biology 8: 65-71.

248. Deutch AH, Rushlow KE, Smith CJ (1984) Analysis of the *Escherichia coli* proBA locus by DNA and protein sequencing. Nucleic Acids Research 12: 6337-6355.

249. Fujita AI, Abe T (2002) Amino acid concentration and distribution of lysozyme and protease activities in the guts of higher termites. Physiological Entomology 27: 76-78.

250. Werdin-Pfisterer NR, Kielland K, Boone RD (2012) Buried organic horizons represent amino acid reservoirs in boreal forest soils. Soil Biology & Biochemistry 55: 122-131.

251. Thomas SR, Ray A, Hodson ME, Pitt TL (2000) Increased sputum amino acid concentrations and auxotrophy ofPseudomonas aeruginosa in severe cystic fibrosis lung disease. Thorax 55: 795-797.

252. Barth A and Pitt T (1996) The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic Pseudomonas aeruginosa. Journal of Medical Microbiology 45: 110-119.

253. Benomar S, Ranava D, Cárdenas ML, Trably E, Rafrafi Y, et al. (2015) Nutritional stress induces exchange of cell material and energetic coupling between bacterial species. Nat Commun 6.

254. Pande S, Shitut S, Freund L, Westermann M, Bertels F, et al. (2015) Metabolic cross-feeding via intercellular nanotubes among bacteria. Nat Commun 6.

255. Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, et al. (2013) Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. ISME J 7: 137-147.

256. Martínez-Núñez MA, Rodríguez-Vázquez K, Pérez-Rueda E (2015) The lifestyle of prokaryotic organisms influences the repertoire of promiscuous enzymes. Proteins: Structure, Function, and Bioinformatics 83: 1625-1631.

257. Lochhead AG (1958) Soil bacteria and growth-promoting substances. Bacteriological Reviews 22: 145-153.

258. Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, et al. (2009) Legumes regulate Rhizobium bacteroid development and persistence by the supply of branched-chain amino acids. Proceedings of the National Academy of Sciences 106: 12477-12482.

259. Kerppola TK, Kahn ML (1988) Symbiotic Phenotypes of Auxotrophic Mutants of *Rhizobium meliloti*104A14. Microbiology 134: 913-919.

260. Frishman D, Valencia A (2009) Modern genome annotation: the BioSapiens Network: Springer Science & Business Media.

261. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499: 431-437.

262. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nature Reviews Genetics 8: 206-216.

263. Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, et al. (2008) Evolvability and hierarchy in rewired bacterial gene networks. Nature 452: 840-845.

264. Avin-Wittenberg T, Galili G (2012) Metabolism: Amino acid regulatory wisdom. Nat Chem Biol 8: 23-24.

265. Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muñiz-Rascado L, et al. (2011) RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). Nucleic Acids Research 39: D98-D105.

266. Bringel F, Hubert J-C (2003) Extent of Genetic Lesions of the Arginine and Pyrimidine Biosynthetic Pathways in *Lactobacillus plantarum*, *L. paraplantarum*, *L. pentosus*, and *L. casei*: Prevalence of CO2-Dependent Auxotrophs and Characterization of Deficient arg Genes in L. plantarum. Applied and Environmental Microbiology 69: 2674-2683.

267. Gerosa L, Kochanowski K, Heinemann M, Sauer U (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. Molecular Systems Biology 9.

268. Giovannoni SJ, Cameron Thrash J, Temperton B (2014) Implications of streamlining theory for microbial ecology. ISME J 8: 1553-1565.

269. Baydoun AR, Emery PW, Pearson JD, Mann GE (1990) Substrate-dependent regulation of intracellular amino acid concentrations in cultured bovine aortic endothelial cells. Biochemical and Biophysical Research Communications 173: 940-948.

270. Sarwar G, Botting H, Peace R (1987) Complete amino acid analysis in hydrolysates of foods and feces by liquid chromatography of precolumn phenylisothiocyanate derivatives. Journal-Association of Official Analytical Chemists 71: 1172-1175.

271. Hosokawa T, Kikuchi Y, Fukatsu T (2007) How many symbionts are provided by mothers, acquired by offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism? Mol Ecol 16: 5316-5325.

272. Price CT, Richards AM, Von Dwingelo JE, Samara HA, Abu Kwaik Y (2014) Amoeba host-*Legionella* synchronization of amino acid auxotrophy and its role in bacterial adaptation and pathogenic evolution. Environ Microbiol 16: 350-358.

273. Visser B, Le Lann C, den Blanken FJ, Harvey JA, van Alphen JJM, et al. (2010) Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. Proceedings of the National Academy of Sciences 107: 8677-8682.

274. Hillesland KL, Lim S, Flowers JJ, Turkarslan S, Pinel N, et al. (2014) Erosion of functional independence early in the evolution of a microbial mutualism. Proceedings of the National Academy of Sciences 111: 14822-14827.

275. Hillesland KL, Stahl DA (2010) Rapid evolution of stability and productivity at the origin of a microbial mutualism. Proceedings of the National Academy of Sciences 107: 2124-2129.

276. Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature 394: 69-72.

277. Gómez P, Buckling A (2013) Real-time microbial adaptive diversification in soil. Ecology letters 16: 650-655.

278. Schluter D (2000) The ecology of adaptive radiation: Oxford University Press.

279. Givnish TJ (2015) Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. New Phytologist 207: 297-303.

280. Dieckmann U, Doebeli M (2004) Adaptive dynamics of speciation: sexual populations. Adaptive speciation: 76-111.

281. Spencer CC, Bertrand M, Travisano M, Doebeli M (2007) Adaptive diversification in genes that regulate resource use in *Escherichia coli*.

282. Le Gac M, Brazas MD, Bertrand M, Tyerman JG, Spencer CC, et al. (2008) Metabolic changes associated with adaptive diversification in *Escherichia coli*. Genetics 178: 1049-1060.

283. Herron MD, Doebeli M (2013) Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. PLoS Biol 11: e1001490.

284. Friesen ML, Saxer G, Travisano M, Doebeli M, Elena S (2004) Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. Evolution 58: 245-260.

285. Turner CB, Blount ZD, Mitchell DH, Lenski RE (2015) Evolution and coexistence in response to a key innovation in a long-term evolution experiment with *Escherichia coli*. bioRxiv doi: http://dx.doi.org/10.1101/020958

286. Dykhuizen DE (1998) Santa Rosalia revisited: why are there so many species of bacteria? Antonie van Leeuwenhoek 73: 25-33.

287. Basan M, Hui S, Okano H, Zhang Z, Shen Y, et al. (2015) Overflow metabolism in Escherichia coli results from efficient proteome allocation. Nature 528: 99-104.

288. Qin X, Zerr DM, McNutt MA, Berry JE, Burns JL, et al. (2012) Pseudomonas aeruginosa Syntrophy in Chronically Colonized Airways of Cystic Fibrosis Patients. Antimicrobial Agents and Chemotherapy 56: 5971-5981.

289. Kost C (2015) Gemeinsam stärker: metabolische Arbeitsteilung bei Bakterien. BIOspektrum 21: 592-596.

290. Zhang X, Reed JL (2014) Adaptive Evolution of Synthetic Cooperating Communities Improves Growth Performance. PLoS ONE 9: e108297.

291. Fletcher JA, Doebeli M (2009) A simple and general explanation for the evolution of altruism. Proceedings of the Royal Society of London B: Biological Sciences 276: 13-19.

292. Bever JD, Richardson SC, Lawrence BM, Holmes J, Watson M (2009) Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. Ecology letters 12: 13-21.

293. Herre E, Knowlton N, Mueller U, Rehner S (1999) The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends in Ecology & Evolution 14: 49-53.

294. Donlan RM, Costerton JW (2002) Biofilms: survival mechanisms of clinically relevant microorganisms. Clinical microbiology reviews 15: 167-193.

295. Costerton JW, Cheng K, Geesey GG, Ladd TI, Nickel JC, et al. (1987) Bacterial biofilms in nature and disease. Annual Reviews in Microbiology 41: 435-464.

296. Marrie T, Costerton J (1985) Mode of growth of bacterial pathogens in chronic polymicrobial human osteomyelitis. Journal of clinical microbiology 22: 924-933.

297. Patel R (2005) Biofilms and antimicrobial resistance. Clinical orthopaedics and related research 437: 41-47.

298. Boles BR, Thoendel M, Singh PK (2004) Self-generated diversity produces "insurance effects" in biofilm communities. Proceedings of the National Academy of Sciences of the United States of America 101: 16630-16635.

299. Momeni B, Waite AJ, Shou W (2013) Spatial self-organization favors heterotypic cooperation over cheating. Elife 2: e00960.

300. Pande S, Kaftan F, Lang S, Svatoš A, Germerodt S, et al. (2015) Privatization of cooperative benefits stabilizes mutualistic cross-feeding interactions in spatially structured environments. The ISME Journal. doi: 10.1038/ismej.2015.212
301. Schink B, Stams AJ (2013) Syntrophism among prokaryotes: Springer.
302. Wang J, Atolia E, Hua B, Savir Y, Escalante-Chong R (2015) Natural variation in preparation for nutrient depletion reveals a cost-benefit tradeoff. PLoS Biol 13: e1002041.
303. Venturelli OS, Zuleta I, Murray RM, El-Samad H (2015) Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. PLoS Biol 13: e1002042.
304. Embree M, Liu JK, Al-Bassam MM, Zengler K (2015) Networks of energetic and metabolic interactions define dynamics in microbial communities. Proceedings of the National Academy of Sciences. doi:10.1073/pnas.1506034112
305. Kiers ET, West SA (2015) Evolving new organisms via symbiosis. Science 348: 392-394.
306. Campbell MA, Van Leuven JT, Meister RC, Carey KM, Simon C, et al. (2015) Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia. Proceedings of the National Academy of Sciences 112: 10192-10199.
307. Rosenzweig R, Sharp R, Treves D, Adams J (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics 137: 903 - 917.
308. Helling RB, Vargas CN, Adams J (1987) Evolution of *Escherichia coli* During Growth in a Constant Environment. Genetics 116: 349-358.
309. Pfeiffer T, Bonhoeffer S (2004) Evolution of cross-feeding in microbial populations. Am Nat 163: E126-135.
310. Ge Z, Girguis PR, Buie CR (2015) Nanoporous Microscale Microbial Incubators. Lab on a Chip. DOI**:** 10.1039/C5LC00978B
311. Chiu H-C, Levy R, Borenstein E (2014) Emergent biosynthetic capacity in simple microbial communities. PLoS Comput Biol 10(7): e1003695.

Supplemental Information

Supporting Information for chapter 1

SUPPORTING METHODS

Computation of protein and DNA sequence biosynthetic costs

Biosynthetic costs were estimated as the amount of carbon source that is required to produce 1 the amount of a certain protein per cell and 2 the DNA sequence of a certain gene. We used flux balance analysis within the Cobra toolbox v2.0 Schellenberger et al. 2011 in a genome-scale metabolic network of *Escherichia coli* K12 Orth et al. 2011. For each protein, the artificial reactions for protein synthesis,

$$(\mathbf{1})\ l\left(\sum_{\mathrm{a}_j \in \mathrm{AA}} n_j^a\ \mathrm{a}_j + \left(q * m_p\right)\mathrm{ATP}\right) \rightarrow l * m_p\ \mathrm{H_2O} + q\ \mathrm{ADP} + q\ \mathrm{Phosphate} + q\ \mathrm{H^+}$$

or for the synthesis of the corresponding DNA sequence

$$(\mathbf{2})\ k * \sum_{\mathrm{d}_j \in NA} n_j^d\ \mathrm{d}_j \rightarrow k * m_{DNA}\ \mathrm{Pyrophosphate}$$

have been included into the model. AA is the set of all 20 proteinogenic amino acids, NA is the set of the four desoxynucleoside triphosphates dATP, dCTP, dGTP, and dTTP. n^a_j represents the number of occurrences of amino acid a_j in the amino acid sequence of the protein and n^d_j the number of occurrences of the desoxynucleoside triphosphate d_j in the DNA sequence of the gene. m_p is the length of the amino acid sequence of the protein and m_{DNA} the length of the corresponding DNA sequence. The abundance of the protein per cell has been incorporated in the calculations by the parameter *l* and the number of DNA sequence copies by parameter k. Protein abundance data were taken from Wessely et al. 2011 and a maximum of 6.54 DNA sequence copies k=6.54 were assumed for all sequences, which corresponds to the maximal chromosomal copy-number of a single-locus gene near the origin of replication in an *E. coli* cell at 2.5 doublings per hour Klumpp et al. 2009. The parameter q represents the ATP requirement per amino acid residue during the polymerization process of translation. A previously reported value of q=4.2 was used Kaleta et al. 2013.

The lower bound for the flux of these reactions was set to a value of 1. The consumption of fructose as sole carbon source was minimized by linear programming to determine the minimal amount of fructose required to produce the corresponding proteins and DNA sequence. DNA and protein sequences were retrieved from the EcoCyc database Keseler et al. 2013.

Construction of auxotrophic strains of Acinetobacter baylyi

Linear constructs of the kanamycin cassette with 5'-overhangs homologous to the insertion site were produced by PCR. To this end, plasmid pKD4 Datsenko and Wanner 2000 DNA was used as a template to amplify the kanamycin resistance cassette. Upstream and downstream regions homologous to *argH, hisD, leuB*, and *trpB* were amplified using primers with a 5'-extension that was complementary to the primers used to amplify the kanamycin cassette Table S4. The three resulting products were combined by PCR to finally obtain the kanamycin cassette fused to the upstream and downstream homologous overhangs. Natural competence of *A. baylyi* was utilized to transform the linear fragments into the WT strain. Transformation was done by diluting 20 µl of a 16 h old culture grown in LB medium. This diluted culture was again incubated at 30 °C with shaking. 50 µl PCR mix containing the deletion cassette was added to this culture and again incubated at 30 °C with shaking for 2 h. Finally, the whole culture volume was concentrated to 100 µl and plated on LB agar plates containing kanamycin and incubated at 30 °C for colonies to appear.

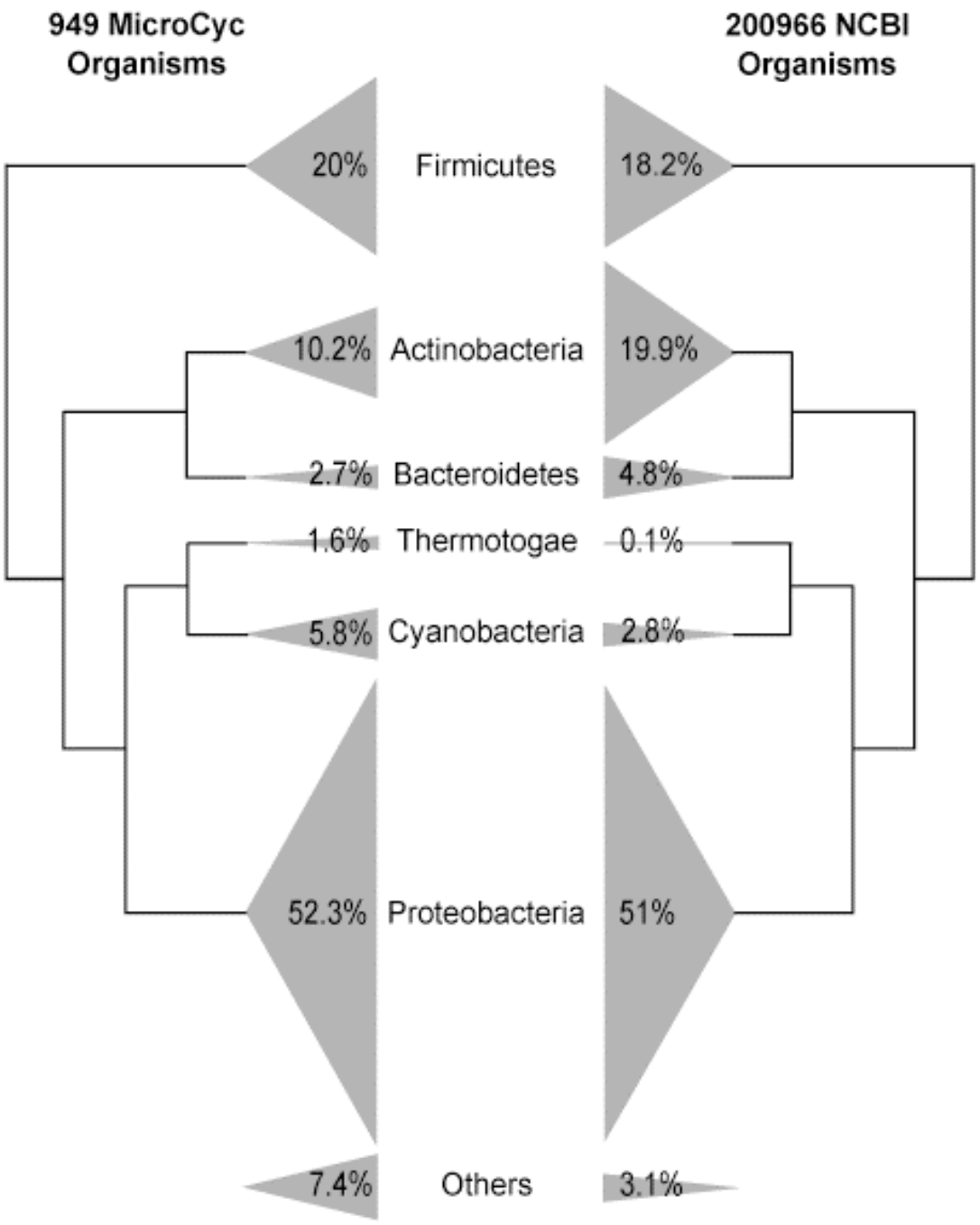


Figure S2. Taxonomic distribution of eubacterial strains used for *in silico* prediction of auxotrophies. Triangle size indicates the proportion of the phylum in the sample of 949 bacterial species from the MicroCyc database, which were used for auxotrophy prediction left cladogram and, for comparison, the proportion of each phylum in the National Center for Biotechnology Information NCBI

taxonomy database of all phylum-classified Eubacteria status: March 2013. Phylogeny adapted from Ciccarelli et al. 2006.

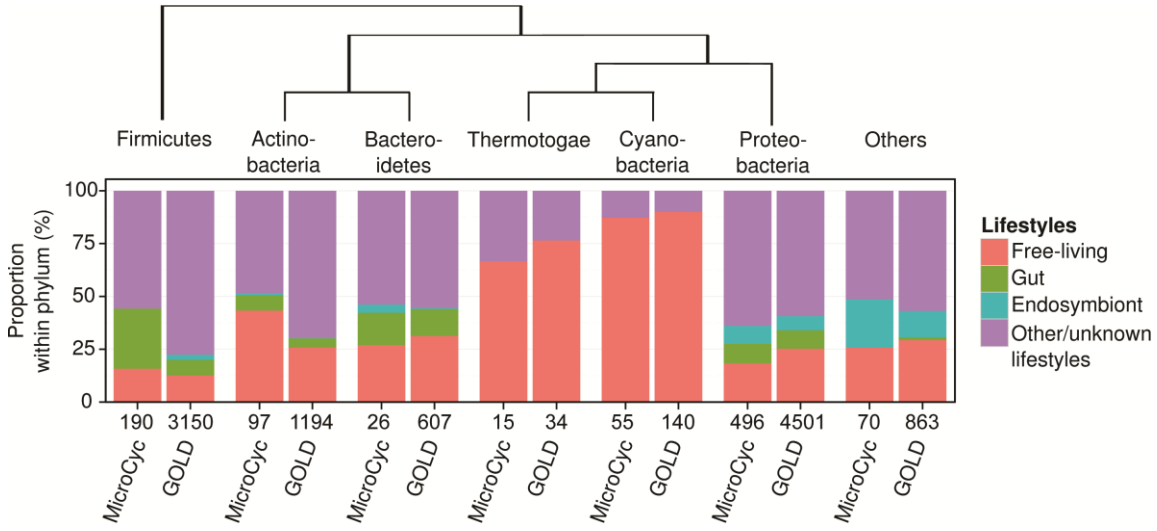
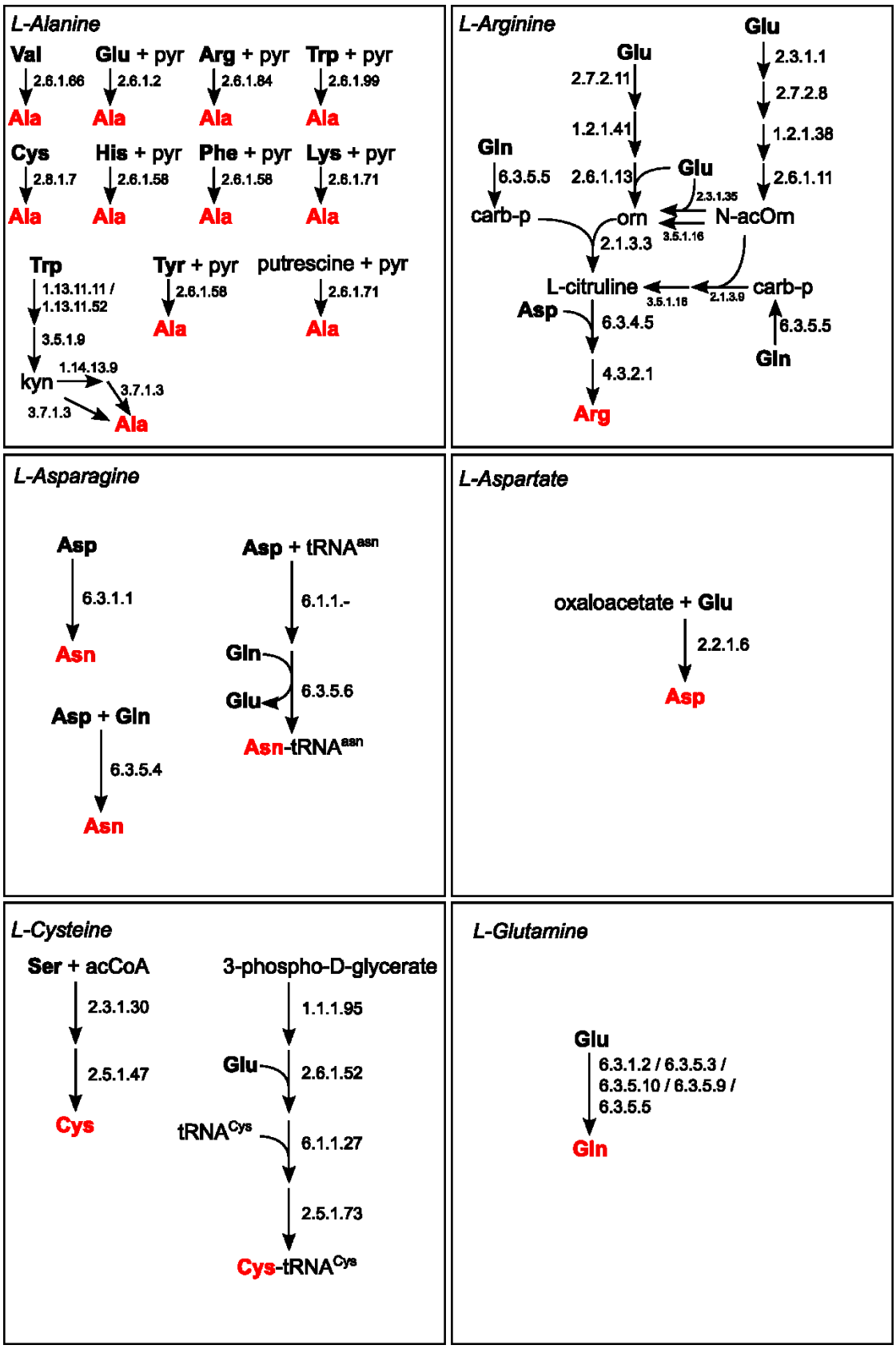
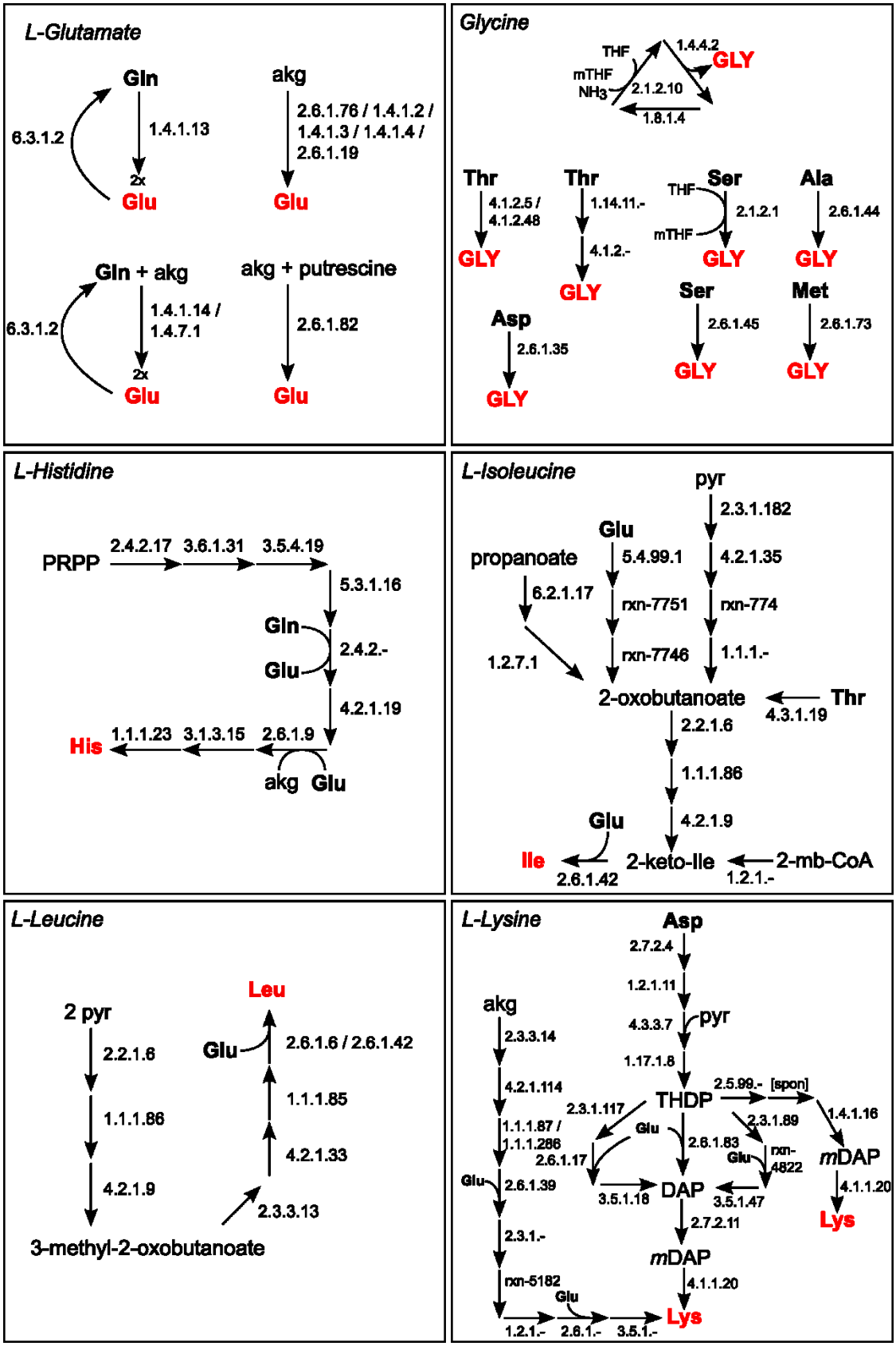
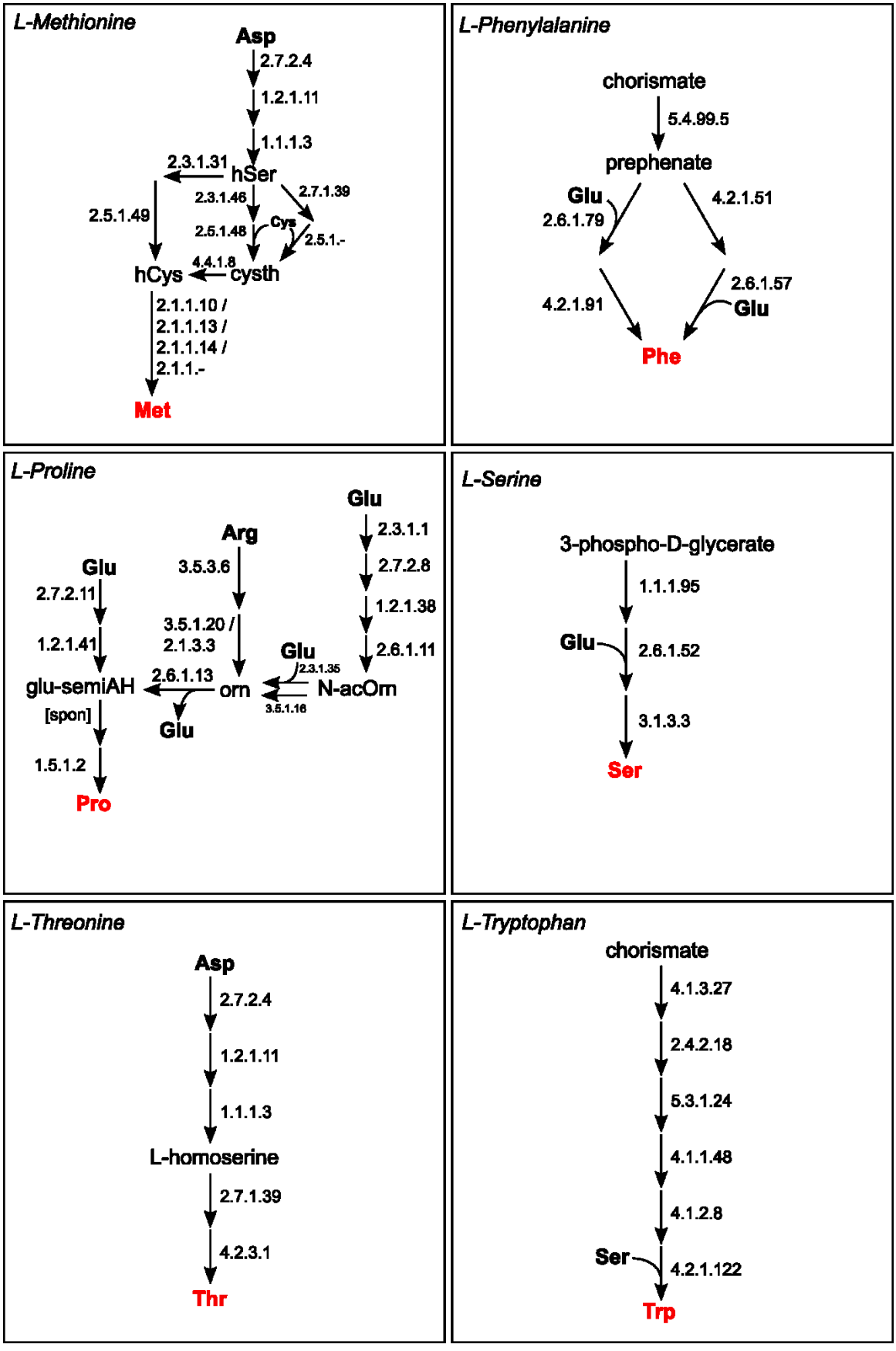
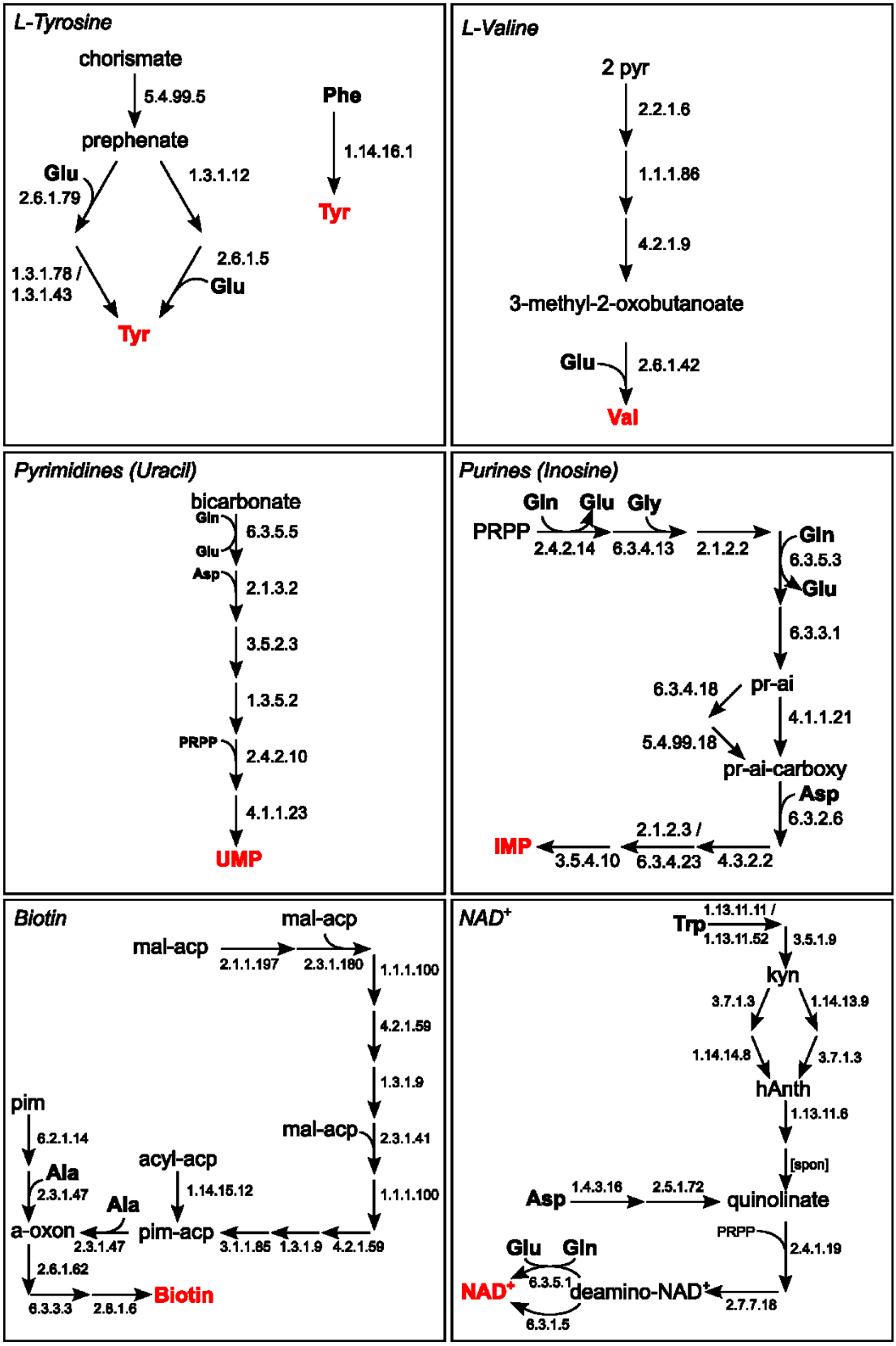


Figure S4. Phylogenetic distribution of free-living, gut-inhabiting, and endosymbiotic bacteria within the MicroCyc database i.e. 949 organisms; Vallenet et al. 2009 and the Genomes OnLine Database GOLD; 10,489 organisms; Pagani et al. 2012. Only organisms for which the whole genome sequence as well as its lifestyle as listed in the Genomes OnLine Database were known were included. Numbers below bars indicate the number of organisms within the corresponding phylum and database. Phylogeny adapted from Ciccarelli et al. 2006.


L-Alanine
Val
Glu + pyr
Arg + pyr
Trp + pyr
2.6.1.66
2.6.1.2
2.6.1.84
2.6.1.99
Ala
Ala
Ala
Ala
Cys
His + pyr
Phe + pyr
Lys + pyr
2.8.1.7
2.6.1.58
2.6.1.58
2.6.1.71
Ala
Ala
Ala
Ala
Trp
Tyr + pyr
putrescine + pyr
1.13.11.11 / 1.13.11.52
2.6.1.58
2.6.1.71
3.5.1.9
Ala
Ala
kyn
1.14.13.9
3.7.1.3
3.7.1.3
Ala
L-Arginine
Glu
Glu
2.3.1.1
2.7.2.11
2.7.2.8
1.2.1.41
1.2.1.38
Gln
Glu
6.3.5.5
2.6.1.13
2.6.1.11
2.3.1.35
carb-p
orn
N-acOrn
3.5.1.16
2.1.3.3
L-citruline
carb-p
3.5.1.18
2.1.3.9
Asp
6.3.4.5
6.3.5.5
Gln
4.3.2.1
Arg
L-Asparagine
Asp
Asp + tRNAasn
6.3.1.1
6.1.1.-
Asn
Gln
Glu
6.3.5.6
Asp + Gln
Asn-tRNAasn
6.3.5.4
Asn
L-Aspartate
oxaloacetate + Glu
2.2.1.6
Asp
L-Cysteine
Ser + acCoA
3-phospho-D-glycerate
2.3.1.30
1.1.1.95
2.5.1.47
Glu
2.6.1.52
Cys
tRNACys
6.1.1.27
2.5.1.73
Cys-tRNACys
L-Glutamine
Glu
6.3.1.2 / 6.3.5.3 / 6.3.5.10 / 6.3.5.9 / 6.3.5.5
Gln


L-Glutamate
Gln
6.3.1.2
1.4.1.13
2x
Glu
akg
2.6.1.76 / 1.4.1.2 / 1.4.1.3 / 1.4.1.4 / 2.6.1.19
Glu
Gln + akg
6.3.1.2
1.4.1.14 / 1.4.7.1
2x
Glu
akg + putrescine
2.6.1.82
Glu
Glycine
THF
mTHF
NH3
1.4.4.2
GLY
2.1.2.10
1.8.1.4
Thr
4.1.2.5 / 4.1.2.48
GLY
Thr
1.14.11.-
4.1.2.-
GLY
Ser
THF
mTHF
2.1.2.1
GLY
Ala
2.6.1.44
GLY
Asp
2.6.1.35
GLY
Ser
2.6.1.45
GLY
Met
2.6.1.73
GLY
L-Histidine
PRPP
2.4.2.17
3.6.1.31
3.5.4.19
5.3.1.16
Gln
Glu
2.4.2.-
4.2.1.19
His
1.1.1.23
3.1.3.15
2.6.1.9
akg
Glu
L-Isoleucine
pyr
2.3.1.182
Glu
propanoate
5.4.99.1
4.2.1.35
6.2.1.17
rxn-7751
rxn-774
1.2.7.1
rxn-7746
1.1.1.-
2-oxobutanoate
Thr
4.3.1.19
2.2.1.6
1.1.1.86
Glu
4.2.1.9
Ile
2.6.1.42
2-keto-Ile
2-mb-CoA
1.2.1.-
L-Leucine
2 pyr
2.2.1.6
1.1.1.86
4.2.1.9
3-methyl-2-oxobutanoate
Leu
Glu
2.6.1.6 / 2.6.1.42
1.1.1.85
4.2.1.33
2.3.3.13
L-Lysine
Asp
2.7.2.4
1.2.1.11
akg
2.3.3.14
4.3.3.7
pyr
1.17.1.8
4.2.1.114
THDP
2.5.99.- [spon]
1.4.1.16
2.3.1.117
1.1.1.87 / 1.1.1.286
Glu
2.3.1.89
2.6.1.83
2.6.1.17
rxn-4822
mDAP
Glu
2.6.1.39
3.5.1.18
DAP
3.5.1.47
4.1.1.20
Lys
2.7.2.11
2.3.1.-
mDAP
rxn-5182
4.1.1.20
Glu
Lys
1.2.1.-
2.6.1.-
3.5.1.-


L-Methionine
Asp
2.7.2.4
1.2.1.11
1.1.1.3
2.3.1.31
hSer
2.3.1.46
2.7.1.39
2.5.1.49
Cys
2.5.1.48
2.5.1.-
4.4.1.8
hCys
cysth
2.1.1.10 /
2.1.1.13 /
2.1.1.14 /
2.1.1.-
Met
L-Phenylalanine
chorismate
5.4.99.5
prephenate
Glu
2.6.1.79
4.2.1.51
4.2.1.91
2.6.1.57
Glu
Phe
L-Proline
Glu
Arg
2.3.1.1
Glu
3.5.3.6
2.7.2.8
2.7.2.11
1.2.1.38
3.5.1.20 /
2.1.3.3
1.2.1.41
Glu
2.3.1.35
2.6.1.11
glu-semiAH
2.6.1.13
orn
N-acOrn
[spon]
3.5.1.16
Glu
1.5.1.2
Pro
L-Serine
3-phospho-D-glycerate
1.1.1.95
Glu
2.6.1.52
3.1.3.3
Ser
L-Threonine
Asp
2.7.2.4
1.2.1.11
1.1.1.3
L-homoserine
2.7.1.39
4.2.3.1
Thr
L-Tryptophan
chorismate
4.1.3.27
2.4.2.18
5.3.1.24
4.1.1.48
4.1.2.8
Ser
4.2.1.122
Trp


L-Tyrosine
chorismate
5.4.99.5
prephenate
Glu
2.6.1.79
1.3.1.12
Phe
1.14.16.1
Tyr
1.3.1.78 /
1.3.1.43
2.6.1.5
Glu
Tyr
L-Valine
2 pyr
2.2.1.6
1.1.1.86
4.2.1.9
3-methyl-2-oxobutanoate
Glu
2.6.1.42
Val
Pyrimidines (Uracil)
bicarbonate
Gln
Glu
6.3.5.5
Asp
2.1.3.2
3.5.2.3
1.3.5.2
PRPP
2.4.2.10
4.1.1.23
UMP
Purines (Inosine)
Gln Glu Gly
PRPP
2.4.2.14
6.3.4.13
2.1.2.2
Gln
6.3.5.3
Glu
6.3.3.1
6.3.4.18
pr-ai
4.1.1.21
5.4.99.18
pr-ai-carboxy
Asp
6.3.2.6
2.1.2.3 /
IMP
3.5.4.10
6.3.4.23
4.3.2.2
Biotin
mal-acp
mal-acp
2.1.1.197
2.3.1.180
1.1.1.100
4.2.1.59
1.3.1.9
mal-acp
2.3.1.41
1.1.1.100
pim
6.2.1.14
Ala
2.3.1.47
acyl-acp
1.14.15.12
Ala
a-oxon
2.3.1.47
pim-acp
3.1.1.85
1.3.1.9
4.2.1.59
2.6.1.62
6.3.3.3
2.8.1.6
Biotin
NAD+
Trp
1.13.11.11 /
1.13.11.52
3.5.1.9
kyn
3.7.1.3
1.14.13.9
1.14.14.8
3.7.1.3
hAnth
1.13.11.6
[spon]
Asp
1.4.3.16
2.5.1.72
quinolinate
PRPP
2.4.1.19
Glu Gln
6.3.5.1
NAD+
deamino-NAD+
2.7.7.18
6.3.1.5

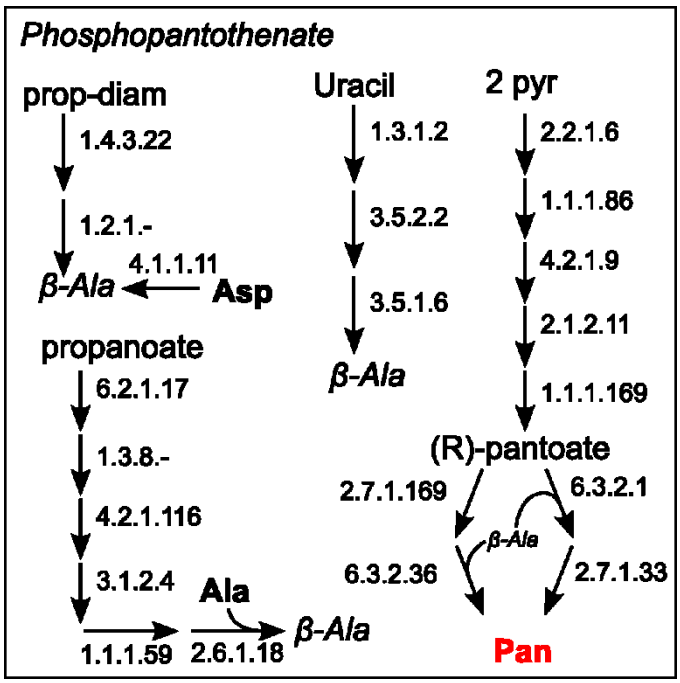


Figure S1. Metabolic pathways that were considered for the prediction of auxotrophies. Pathways including EC numbers were collected from the MetaCyc database Caspi et al. 2012. Target compounds of each metabolic route are written in red. Metabolites written in black bold type are indicating dependencies on other biosynthetic pathways. All reactions are named by the corresponding EC number or the MetaCyc reaction ID if no EC number is assigned to the reaction in MetaCyc. UMP uridine monophosphate is the precursor for cytosine and IMP inosine monophosphate is the precursor for guanosine. Abbreviations: pyr: pyruvate, acCoA: acetyl CoA, $tRNA^{Cys}$: uncharged tRNA for L-cysteine, Cys-$tRNA^{Cys}$: L-cysteine-charged tRNA for L-cysteine, 2-keto-Ile: 2-keto-isoleucine, 2-mb-CoA: 2-methylbutanoyl-CoA, carb-p: carbamyl-phosphate, N-acOrn: N-acetyl-L-ornithine, orn: L-ornithine, akg: α-keto-glutarat, glu-semiAH: L-glutamate-5-semialdehyd, THDP: S-2,3,4,5-tetrahydrodipicolinate, DAP: L,L-diaminopimelate, mDAP: meso-diaminopimelate, kyn: L-kynurenine, cysth: L-cystathionine, hSer: L-homoserine, hCys: L-homocysteine, THF: tetrahydrofolate, mTHF: 5,10-methylenetetrahydrofolate, acp: acyl carrier protein, mal-acp: a malonyl acp, acyl-acp: a long chain acyl-acp, pim: pimelate, pim-acp: pimelyl-acp, a-oxon: 8-amino-7-oxononanoate, hAnth: 3-hydroxyanthranilate, pr-ai: 5-amino-1-5-phospho-β-D-ribosylimidazole, pr-ai-carboxy: 5-amino-1-5-phospho-D-ribosylimidazole-4-carboxylate, prop-diam: propane-1,3-diamine, [spon]: spontaneous reaction.

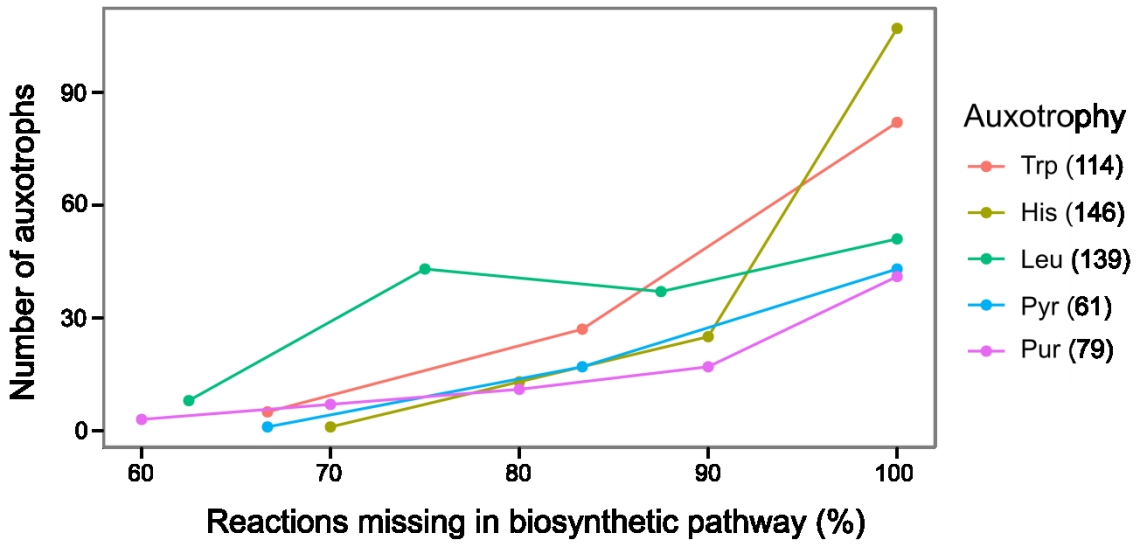


Figure S3: Incompleteness of the biosynthetic pathways forming tryptophan, histidine, leucine, pyrimidine, and purine within all Eubacteria predicted to be auxotrophic for these metabolites. The numbers behind the auxotrophy indicate the total number of strains, which are predicted to be auxotrophic for the corresponding compound. Predictions are based on analyses of the MicroCyc database i.e. 949 organisms; Vallenet et al. 2009. These five pathways were chosen, because they are the longest linear pathways in the data set see Fig. S1.

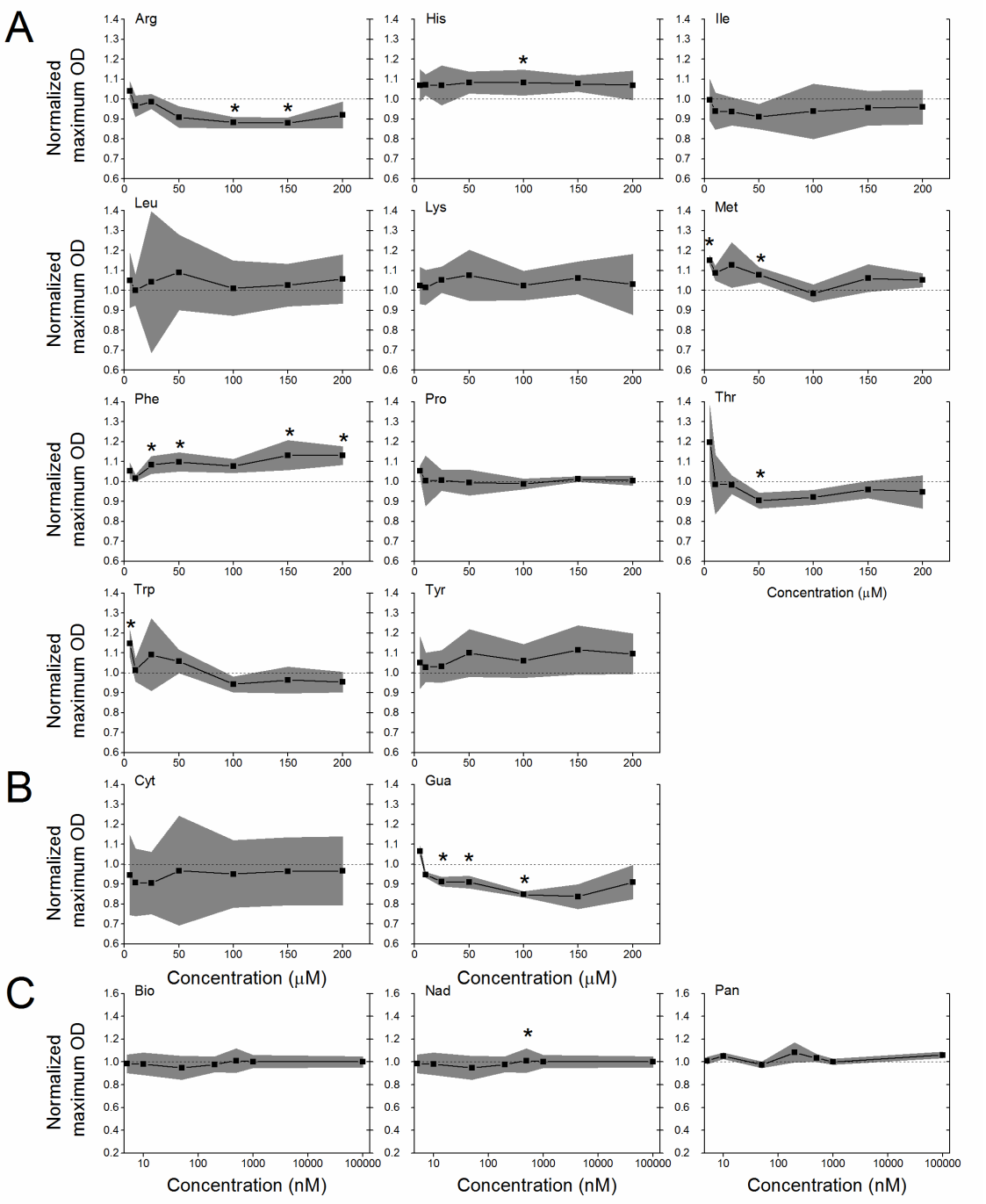


Figure S5: Growth response of *Escherichia coli* WT to increasing concentrations of the focal metabolites. Growth within 24 h was determined as optical density at 600 nm and is displayed as growth in minimal medium that contained a particular metabolite at a certain concentration relative to its growth in pure minimal medium. Each plot shows the concentration-dependent normalized growth response of WT in the presence of **A** an amino acid, **B** a nucleobase, or **C** a vitamin. All values are medians of four replicates and the grey-shaded area delimits the 95% confidence intervals. Asterisks mark significant differences from the growth of the WT in the absence of the focal compound i.e. dashed line;

FDR-corrected independent sample t-tests: *P<0.05, **P<0.01, and ***P<0.001, n=4. See Table S1 for abbreviations of metabolite names.

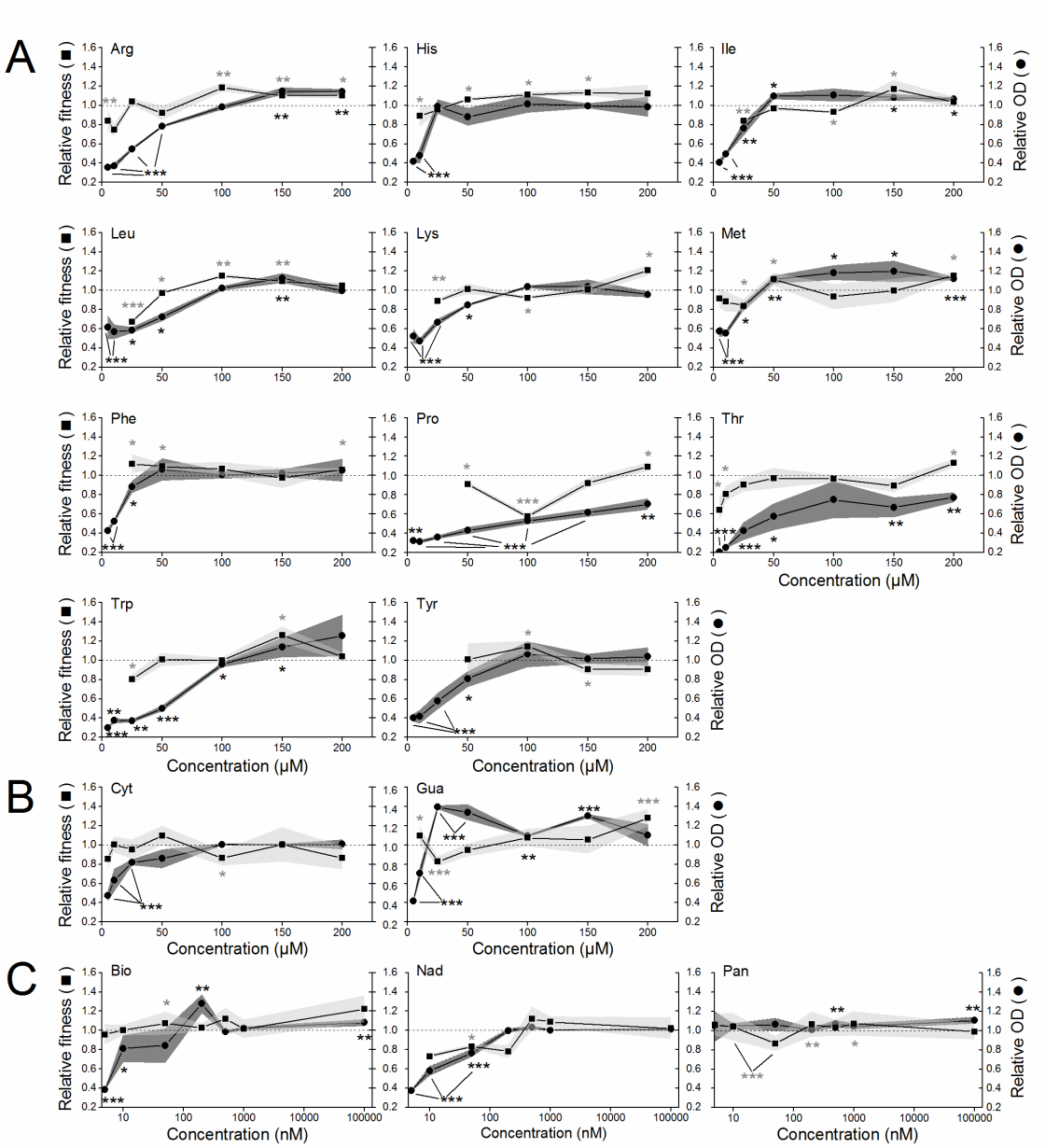


Figure S6. Productivity and competitive fitness of *Escherichia coli* auxotrophs relative to WT in increasing concentrations of the focal metabolites. Productivity i.e. OD and fitness of auxotrophic mutants within 24 h was determined relative to WT in both mono- circles or coculture squares using minimal medium that has been supplemented with **A** an amino acid, **B** a nucleobase, or **C** a vitamin in increasing concentrations. Relative OD of monocultures was determined as the ratio of the auxotroph's and the WT's optical densities measured at 600 nm and the relative fitness of cocultures is expressed as the ratio of their Malthusian parameters. Medians of four replicates are displayed. The dark and light grey regions mark the 95% confidence intervals for mono- and cocultures, respectively. Black and light grey asterisks mark significant differences of auxotrophs to WT levels i.e. dashed line in mono- and cocultures, respectively

monocultures: FDR-corrected independent sample t-tests, cocultures: FDR-corrected paired sample t-tests: *P<0.05, **P<0.01, and ***P<0.001; n=4. See Table S1 for abbreviations of metabolite names.

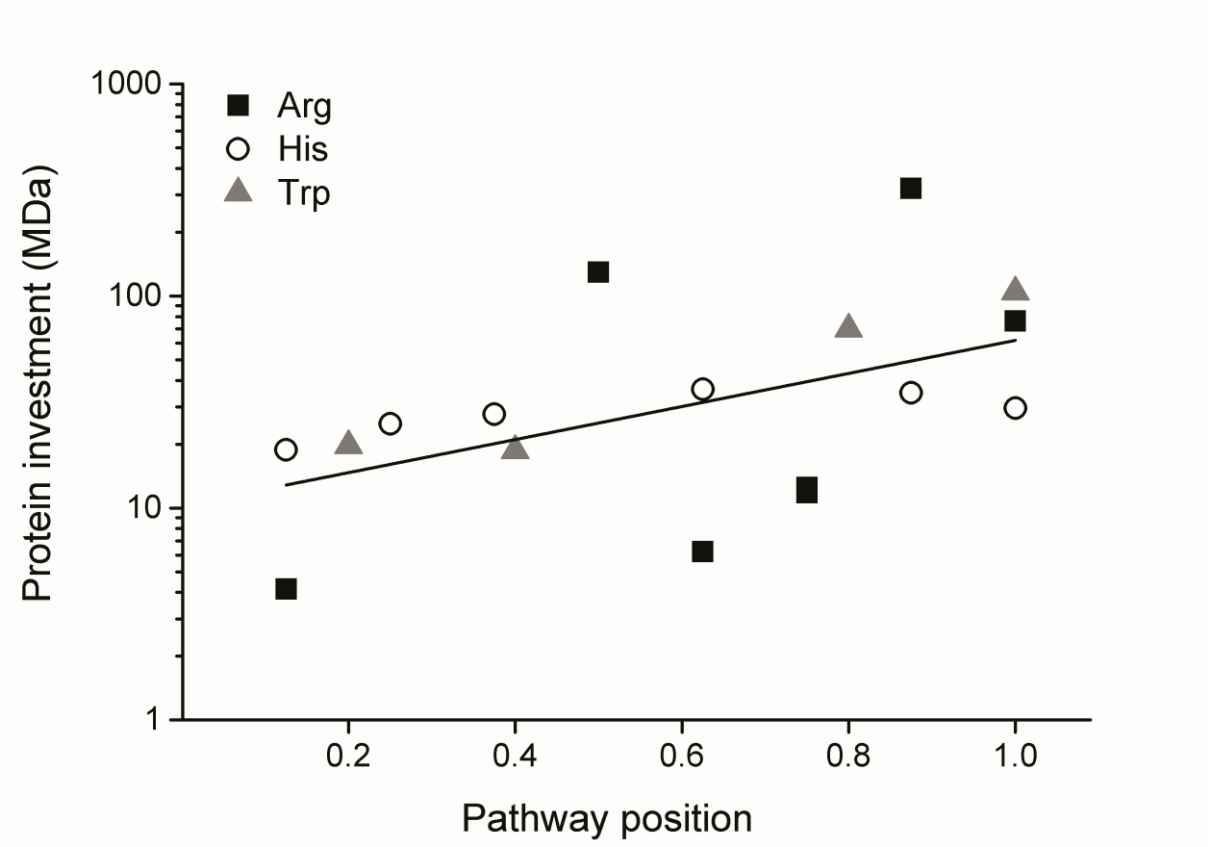


Figure S7. Relationship between the amount of protein invested by *Escherichia coli* into a certain biosynthetic step and the position of the gene within the biosynthetic pathways of arginine Arg, histidine His, and tryptophan Trp. Protein investment in Mega Dalton is the mass of the individual protein multiplied with the abundance of protein copies per cell. Data was obtained from Wessely et al. 2011. Pathway position is the normalised localisation of each gene between the start 0.1 and the end 1.0 of the pathway. The line is the linear fit line between both variables.

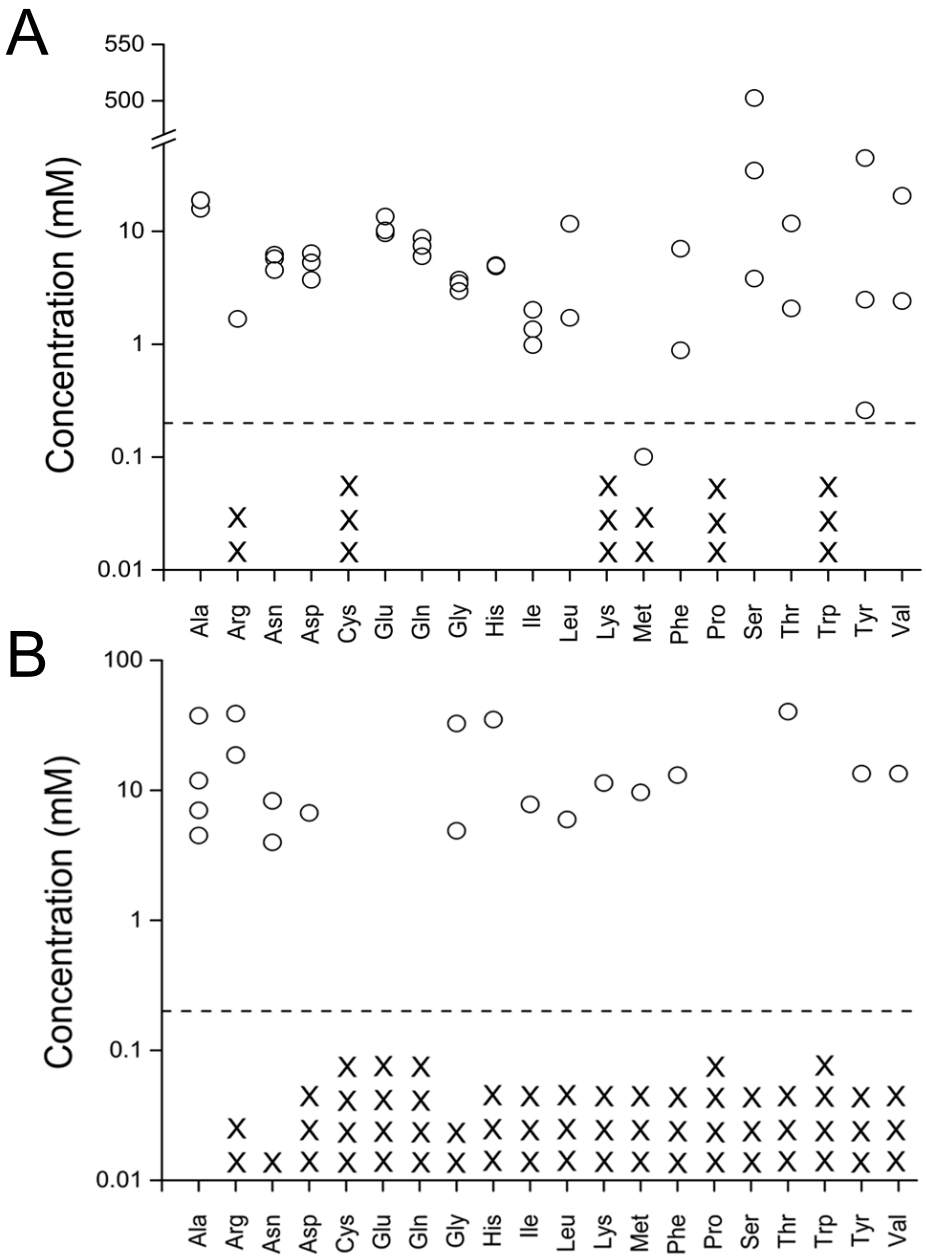


Figure S8. Amino acid concentrations in natural habitats of bacteria. Concentrations of individual amino acids found in **A** three different soil samples mM kg^{-1} soil Werdin-Pfisterer et al. 2012 and **B** the gut of four different termite species mM gut^{-1} Fujita and Abe 2002. Each circle indicates the amount of amino acid quantified in either a single soil sample or termite species. The dashed line represents the upper limit of amino acid concentrations i.e. 200 µM used in this study to determine the fitness of auxotrophic mutants. Crosses X signify instances in which the corresponding amino acid was not detected.

Table S1. Overview over the different auxotrophies analysed and the abbreviations used.

			Auxotrophy analysed in	
Class	Metabolite	Abbreviation	*Escherichia coli*	*Acinetobacter baylyi*
Amino acid	Alanine	Ala	●	
	Arginine	Arg	●	●
	Asparagine	Asn	●	
	Aspartic acid	Asp	●	
	Cysteine	Cys	●	
	Glutamine	Gln	●	
	Glutamic acid	Glu	●	
	Glycine	Gly	●	
	Histidine	His	●	●
	Isoleucine	Ile	●	
	Leucine	Leu	●	●
	Lysine	Lys	●	
	Methionine	Met	●	
	Phenylalanine	Phe	●	
	Proline	Pro	●	
	Serine	Ser	●	
	Threonine	Thr	●	
	Tryptophan	Trp	●	●
	Tyrosine	Tyr	●	
	Valine	Val	●	
Nucleobase	Cytosine	Cyt	●	
	Guanine	Gua	●	
Vitamin	Biotin	Bio	●	
	Nicotinamide adenine dinucleotide	Nad	●	
	Pantothenate	Pan	●	

Table S2. Strains used in this study. Abbreviations: ara$^{+/-}$ = ability to use arabinose as a C-source absent/ present, AT = auxotroph, WT = wild type.

Strain	Genotype	Phenotype	Reference
Escherichia coli BW25113 ara$^-$	F-, *ΔaraD-araB567*, *ΔlacZ4787*::rrnB-3, *λ*$^-$, *rph-1*, *ΔrhaD-rhaB568*, *hsdR514*	WT Red	Baba et al. 2006
Escherichia coli BW25113 ara$^+$	F-, *ΔaraD-araB567*, *ΔlacZ4787*::rrnB-3, *λ*$^-$, *rph-1*, *ΔrhaD-rhaB568*, *hsdR514, araA*	WT White	This study
Δ*argH* ara$^-$	WT ara$^-$, Δ*argH*::kan^R	AT	This study
Δ*hisD* ara$^-$	WT ara$^-$, Δ*hisD*::kan^R	AT	This study
Δ*ilvA* ara$^-$	WT ara$^-$, Δ*ilvA*::kan^R	AT	This study
Δ*leuB* ara$^-$	WT ara$^-$, Δ*leuB*::kan^R	AT	This study
Δ*lysA* ara$^-$	WT ara$^-$, Δ*lysA*::kan^R	AT	This study
Δ*metA* ara$^-$	WT ara$^-$, Δ*metA*::kan^R	AT	This study
Δ*pheA* ara$^-$	WT ara$^-$, Δ*pheA*::kan^R	AT	This study
Δ*proC* ara$^-$	WT ara$^-$, Δ*proC*::kan^R	AT	This study
Δ*thrC* ara$^-$	WT ara$^-$, Δ*thrC*::kan^R	AT	This study
Δ*trpB* ara$^-$	WT ara$^-$, Δ*trpB*::kan^R	AT	This study
Δ*tyrA* ara$^-$	WT ara$^-$, Δ*tyrA*::kan^R	AT	This study
Δ*pyrF* ara$^-$	WT ara$^-$, Δ*pyrF*::kan^R	AT	This study
Δ*guaB* ara$^-$	WT ara$^-$, Δ*guaB*::kan^R	AT	This study
Δ*bioF* ara$^-$	WT ara$^-$, Δ*bioH*::kan^R	AT	This study
Δ*nadA* ara$^-$	WT ara$^-$, Δ*nadA*::kan^R	AT	This study
Δ*panC* ara$^-$	WT ara$^-$, Δ*panC*::kan^R	AT	This study
Δ*argH* ara$^+$	WT ara$^+$, Δ*argH*::kan^R	AT	This study
Δ*hisD* ara$^+$	WT ara$^+$, Δ*hisD*::kan^R	AT	This study
Δ*ilvA* ara$^+$	WT ara$^+$, Δ*ilvA*::kan^R	AT	This study
Δ*leuB* ara$^+$	WT ara$^+$, Δ*leuB*::kan^R	AT	This study
Δ*lysA* ara$^+$	WT ara$^+$, Δ*lysA*::kan^R	AT	This study
Δ*metA* ara$^+$	WT ara$^+$, Δ*metA*::kan^R	AT	This study
Δ*pheA* ara$^+$	WT ara$^+$, Δ*pheA*::kan^R	AT	This study
Δ*proC* ara$^+$	WT ara$^+$, Δ*proC*::kan^R	AT	This study
Δ*thrC* ara$^+$	WT ara$^+$, Δ*thrC*::kan^R	AT	This study
Δ*trpB* ara$^+$	WT ara$^+$, Δ*trpB*::kan^R	AT	This study
Δ*tyrA* ara$^+$	WT ara$^+$, Δ*tyrA*::kan^R	AT	This study
Δ*pyrF* ara$^+$	WT ara$^+$, Δ*pyrF*::kan^R	AT	This study
Δ*guaB* ara$^+$	WT ara$^+$, Δ*guaB*::kan^R	AT	This study
Δ*bioH* ara$^+$	WT ara$^+$, Δ*bioH*::kan^R	AT	This study
Δ*nadA* ara$^+$	WT ara$^+$, Δ*nadA*::kan^R	AT	This study
Δ*panC* ara$^+$	WT ara$^+$, Δ*panC*::kan^R	AT	This study

Δ*argA* ara$^-$	WT ara$^-$, Δ*argA*::kan^R	AT	This study
Δ*argB* ara$^-$	WT ara$^-$, Δ*argB*::kan^R	AT	This study
Δ*argC* ara$^-$	WT ara$^-$, Δ*argC*::kan^R	AT	This study
Δ*argE* ara$^-$	WT ara$^-$, Δ*argE*::kan^R	AT	This study
Δ*argG* ara$^-$	WT ara$^-$, Δ*argG*::kan^R	AT	This study
Δ*argA* ara$^+$	WT ara$^+$, Δ*argA*::kan^R	AT	This study
Δ*argB* ara$^+$	WT ara$^+$, Δ*argB*::kan^R	AT	This study
Δ*argC* ara$^+$	WT ara$^+$, Δ*argC*::kan^R	AT	This study
Δ*argE* ara$^+$	WT ara$^+$, Δ*argE*::kan^R	AT	This study
Δ*argG* ara$^+$	WT ara$^+$, Δ*argG*::kan^R	AT	This study
Δ*trpA* ara$^-$	WT ara$^-$, Δ*trpA*::kan^R	AT	This study
Δ*trpD* ara$^-$	WT ara$^-$, Δ*trpD*::kan^R	AT	This study
Δ*trpE* ara$^-$	WT ara$^-$, Δ*trpE*::kan^R	AT	This study
Δ*trpA* ara$^+$	WT ara$^+$, Δ*trpA*::kan^R	AT	This study
Δ*trpD* ara$^+$	WT ara$^+$, Δ*trpD*::kan^R	AT	This study
Δ*trpE* ara$^+$	WT ara$^+$, Δ*trpE*::kan^R	AT	This study
Δ*hisA* ara$^-$	WT ara$^-$, Δ*hisA*::kan^R	AT	This study
Δ*hisB* ara$^-$	WT ara$^-$, Δ*hisB*::kan^R	AT	This study
Δ*hisC* ara$^-$	WT ara$^-$, Δ*hisC*::kan^R	AT	This study
Δ*hisA* ara$^+$	WT ara$^+$, Δ*hisA*::kan^R	AT	This study
Δ*hisB* ara$^+$	WT ara$^+$, Δ*hisB*::kan^R	AT	This study
Δ*hisC* ara$^+$	WT ara$^+$, Δ*hisC*::kan^R	AT	This study
Δ*argH*	WT, Δ*argH*::kan^R	AT	Baba et al. 2006
Δ*hisD*	WT, Δ*hisD*::kan^R	AT	Baba et al. 2006
Δ*ilvA*	WT, Δ*ilvA*::kan^R	AT	Baba et al. 2006
Δ*leuB*	WT, Δ*leuB*::kan^R	AT	Baba et al. 2006
Δ*lysA*	WT, Δ*lysA*::kan^R	AT	Baba et al. 2006
Δ*metA*	WT, Δ*metA*::kan^R	AT	Baba et al. 2006
Δ*pheA*	WT, Δ*pheA*::kan^R	AT	Baba et al. 2006
Δ*proC*	WT, Δ*proC*::kan^R	AT	Baba et al. 2006
Δ*thrC*	WT, Δ*thrC*::kan^R	AT	Baba et al. 2006
Δ*trpB*	WT, Δ*trpB*::kan^R	AT	Baba et al. 2006
Δ*tyrA*	WT , Δ*tyrA*::kan^R	AT	Baba et al. 2006
Δ*pyrF*	WT , Δ*pyrF*::kan^R	AT	Baba et al. 2006
Δ*guaB*	WT, Δ*guaB*::kan^R	AT	Baba et al. 2006
Δ*bioF*	WT, Δ*bioH*::kan^R	AT	Baba et al. 2006
Δ*nadA*	WT, Δ*nadA*::kan^R	AT	Baba et al. 2006
Δ*panC*	WT, Δ*panC*::kan^R	AT	Baba et al. 2006
Δ*argA*	WT ,Δ*argA*::kan^R	AT	Baba et al. 2006
Δ*argB*	WT ,Δ*argB*::kan^R	AT	Baba et al. 2006

Δ*argC*	WT ,Δ*argC*::kan^R	AT	Baba et al. 2006
Δ*argE*	WT, Δ*argE*::kan^R	AT	Baba et al. 2006
Δ*argG*	WT, Δ*argG*::kan^R	AT	Baba et al. 2006
Δ*trpA*	WT, Δ*trpA*::kan^R	AT	Baba et al. 2006
Δ*trpD*	WT, Δ*trpD*::kan^R	AT	Baba et al. 2006
Δ*trpE*	WT, Δ*trpE*::kan^R	AT	Baba et al. 2006
Δ*hisA*	WT, Δ*hisA*::kan^R	AT	Baba et al. 2006
Δ*hisB*	WT, Δ*hisB*::kan^R	AT	Baba et al. 2006
Δ*hisC*	WT, Δ*hisC*::kan^R	AT	Baba et al. 2006
REL 606	F-,*tsx-467Am*, *araA 92D*, *lon*, *rpsL227* strR, *hsdR*, *[mal+]*LamS		Studier et al. 2009
REL 607	F-, *tsx-467*Am, *araA* 92G, lon-, *rpsL*227 strR, hsdR-, [mal+]LamS		Lenski et al. 1991
Acinetobacter baylyi ADP1		WT	Vaneechoutte et al. 2006
A. baylyi Δ*argH*	WT, Δ*argH*::kan^R	AT	This study
A. baylyi Δ*hisD*	WT, Δ*hisD*::kan^R	AT	This study
A. baylyi Δl*euB*	WT, Δ*leuB*::kan^R	AT	This study
A. baylyi Δ*trpB*	WT, Δ*trpB*::kan^R	AT	This study

Table S3. Comparison of biosynthetic costs for DNA sequence and the corresponding protein. Cost are given as fructose molecules that are at least needed to produce the DNA sequence of the gene or the amount of the protein. NA=no data available.

Gene	Cost of DNA Sequence 10^4 fructose molecules	Cost of Protein 10^4 fructose molecules
argA	1.783502731	3.439609569
argB	1.088531764	NA
argC	1.373221287	NA
argE	1.555713634	5.507498182
argG	1.796792481	273.9335755
argH	1.8338215	64.23576185
hisA	1.042527654	NA
hisB	1.451117531	NA
hisC	1.452515028	30.78864344
hisD	1.739683547	24.86644536
trpA	1.126263614	61.31053206
trpB	1.607108823	89.72362679
trpD	2.103003962	16.01084019
trpE	2.061666192	15.98466297

Table S4. Primers used for the construction of *Acinetobacter baylyi* auxotrophs. UF = upstream forward, UR = Upstream reverse, DF = downstream forward, DR = downstream reverse.

Gene	Primer	Sequence 5'-3'
Kanamycin resistance cassette	UF	TGTAGGCTGGAGCTGCTTC
	UR	CATATGAATATCCTCCTTA
argH	UF	GAGGTCTGGGTTGAGGTTGG
	UR	GAAGCAGCTCCAGCCTACATAACGCTGCATTTGCAC
hisD	UF	TATGCAAGCCTTGGTGAGCA
	UR	GAAGCAGCTCCAGCCTACACAGCCTCTTCCACTTGA
leuB	UF	CCGTTTACAGGGCTCAGTGT
	UR	GAAGCAGCTCCAGCCTACATCACCCAATCCTGTCAC
trpB	UF	AACCACACACGCTTTTGCAG
	UR	GAAGCAGCTCCAGCCTACAGCTGATCCACATTGGACT
argH	DF	TAAGGAGGATATTCATATGTGCTTCTGGTTTCCAGC
	DR	GGATTTTGCGCCATTCCCTG
hisD	DF	TAAGGAGGATATTCATATGGTAACTGCTCTACGGGG
	DR	ATGCGTCTGCCTGATCTACC
leuB	DF	TAAGGAGGATATTCATATGTTGCCCGAACACCGATC
	DR	CGTTCACGAATCCATGCAAGT
trpB	DF	TAAGGAGGATATTCATATGACGTGATGTGGAAATGG
	DR	AGTTGGGGCTGGATGTCTTG

LITERATURE CITED

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2:0008.

Caspi, R., T. Altman, K. Dreher, C. A. Fulcher, P. Subhraveti, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, et al. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 40:D742-D753.

Ciccarelli, F. D., T. Doerks, C. von Mering, C. J. Creevey, B. Snel, and P. Bork. 2006. Toward automatic reconstruction of a highly resolved tree of life. Science 311:1283-1287.

Datsenko, K. A. and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97:6640-6645.

Fujita, A. I. and T. Abe. 2002. Amino acid concentration and distribution of lysozyme and protease activities in the guts of higher termites. Physiol. Entomol. 27:76-78.

Kaleta, C., S. Schäuble, U. Rinas, and S. Schuster. 2013. Metabolic costs of amino acid and protein production in *Escherichia coli*. Biotechnol. J. 8:1105-1114.

Keseler, I. M., A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martínez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, et al. 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. 41:D605-D612.

Klumpp, S., Z. Zhang, and T. Hwa. 2009. Growth rate-dependent global effects on gene expression in bacteria. Cell 139:1366-1375.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315-1341.

Orth, J. D., T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, and B. O. Palsson. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism - 2011. Mol. Syst. Biol. 7:535.

Pagani, I., K. Liolios, J. Jansson, I. M. A. Chen, T. Smirnova, B. Nosrat, V. M. Markowitz, and N. C. Kyrpides. 2012. The Genomes OnLine Database GOLD v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 40:D571-D579.

Schellenberger, J., R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, et al. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protocols 6:1290-1307.

Studier, F. W., P. Daegelen, R. E. Lenski, S. Maslov, and J. F. Kim. 2009. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21DE3 and comparison of the *E. coli* B and K-12 genomes. J. Mol. Biol. 94:653-680.

Vallenet, D., S. Engelen, D. Mornico, S. Cruveiller, L. Fleury, A. Lajus, Z. Rouy, D. Roche, G. Salvignol, C. Scarpelli, et al. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. J. Biol. Datab. Curation doi: 10.1093/database/bap021.

Vaneechoutte, M., D. M. Young, L. N. Ornston, T. De Baere, A. Nemec, T. Van Der Reijden, E. Carr, I. Tjernberg, and L. Dijkshoorn. 2006. Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly

described species *Acinetobacter baylyi*. Appl. Environ. Microbiol. 72:932-936.

Werdin-Pfisterer, N. R., K. Kielland, and R. D. Boone. 2012. Buried organic horizons represent amino acid reservoirs in boreal forest soils. Soil Biol. Biochem. 55:122-131.

Wessely, F., M. Bartl, R. Guthke, P. Li, S. Schuster, and C. Kaleta. 2011. Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. Mol. Syst. Biol. 7:515.

Supporting Information for chapter 2

SUPPORTING METHODS

Adjustment of fructose and succinate concentrations

We used Flux-Balance-Analysis and a genome-scale metabolic model of *E. coli* (Orth et al. 2011) to calculate how many mole of a carbon source are needed to produce '1 mole of biomass'. We refer to this value using q_x for a carbon source *x*. In detail, q_x was calculated by constraining the flux through the biomass reaction (growth associated maintenance (GAM) estimate: 53.95) of the model to a value equal 1 mmol x gDW^{-1} x h^{-1} and by minimizing the influx of the carbon source *x*. The optimization was performed within Matlab 7.14 (Mathworks) with the COBRA Toolbox version 2.0.5 (Schellenberger et al. 2011) and the TOMLAB v7.9 as linear programming solver. The final concentration of carbon source x was calculated as

$$c_x = c_{Fru} \cdot q_x / q_{Fru}$$

using 5 g l^{-1} fructose (c_{Fru}= 27.75 mM) as reference. The corresponding concentration of disodium succinate was 8.86 g l^{-1} (c_{Suc} = 54.68 mM).

This procedure is similar to the approach used by Adadi et al. (2012), where concentrations were adjusted to match the number of reducible carbon atoms. Using the genome-scale metabolic network of *E. coli* also takes the physiological capabilities of the cell to transform a certain carbon source into biomass into account.

SUPPORTING FIGURES

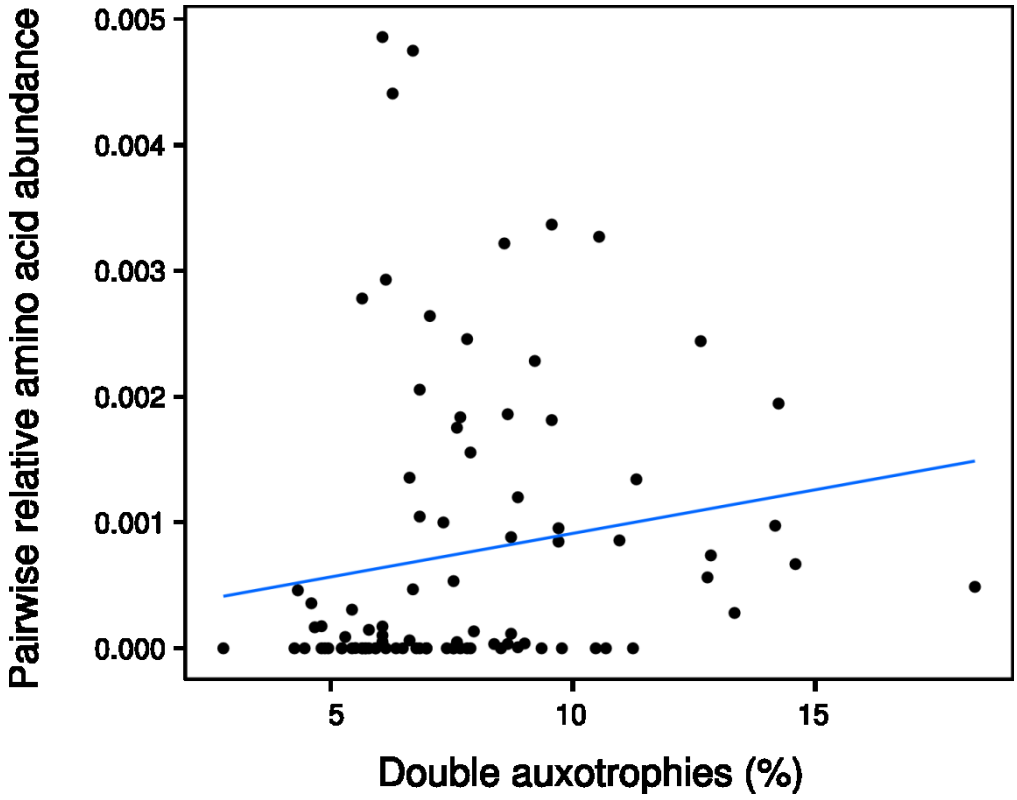


Figure S1. Correlation of the frequency of double auxotrophies among 1,432 eubacteria and the median of pairwise products of amino acid abundances in 69 natural environments (Moura et al. 2013). Kendall's rank correlation: $R_\tau = 0.22$, $P = 0.003$, $n = 91$.

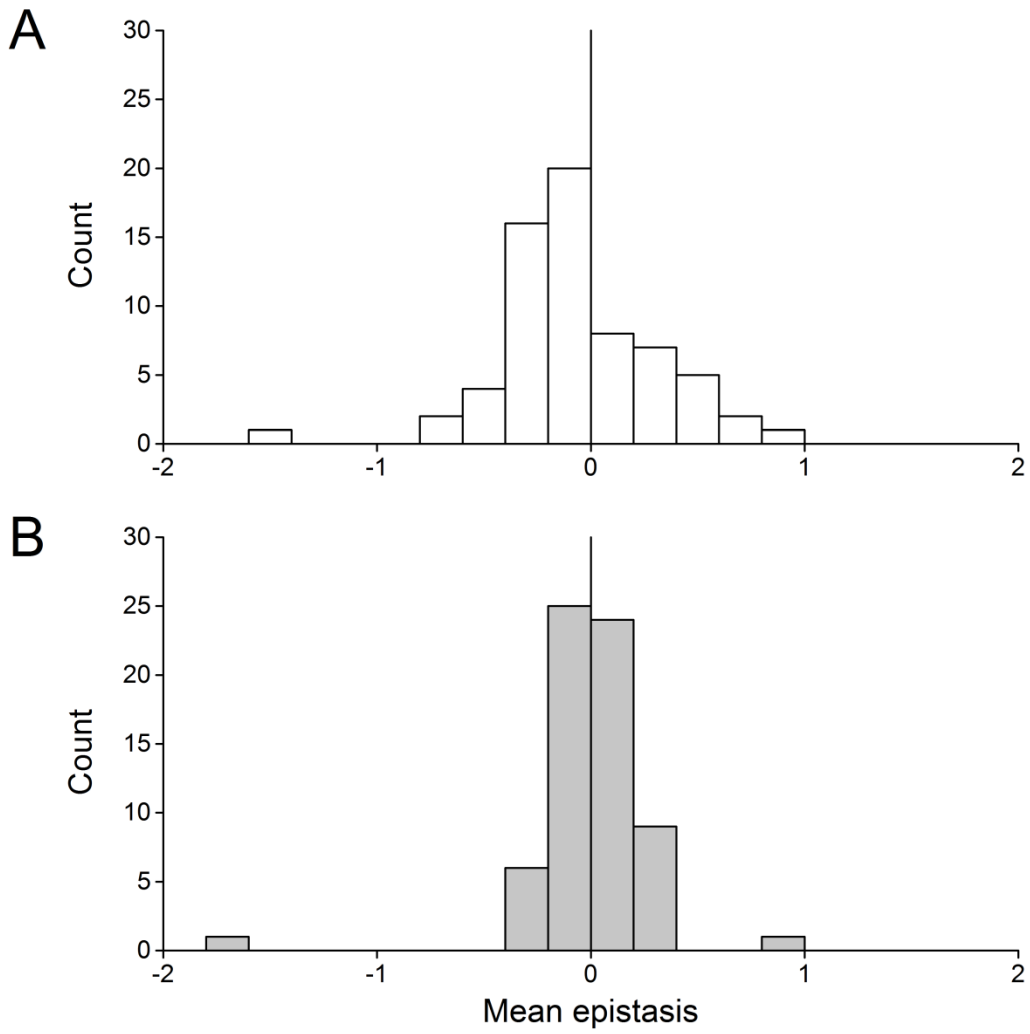


Figure S2. Frequency distribution of epistatic effects for 55 double- and 16 triple gene deletion mutants as determined in (A) the fructose- and (B) the succinate-containing environment.

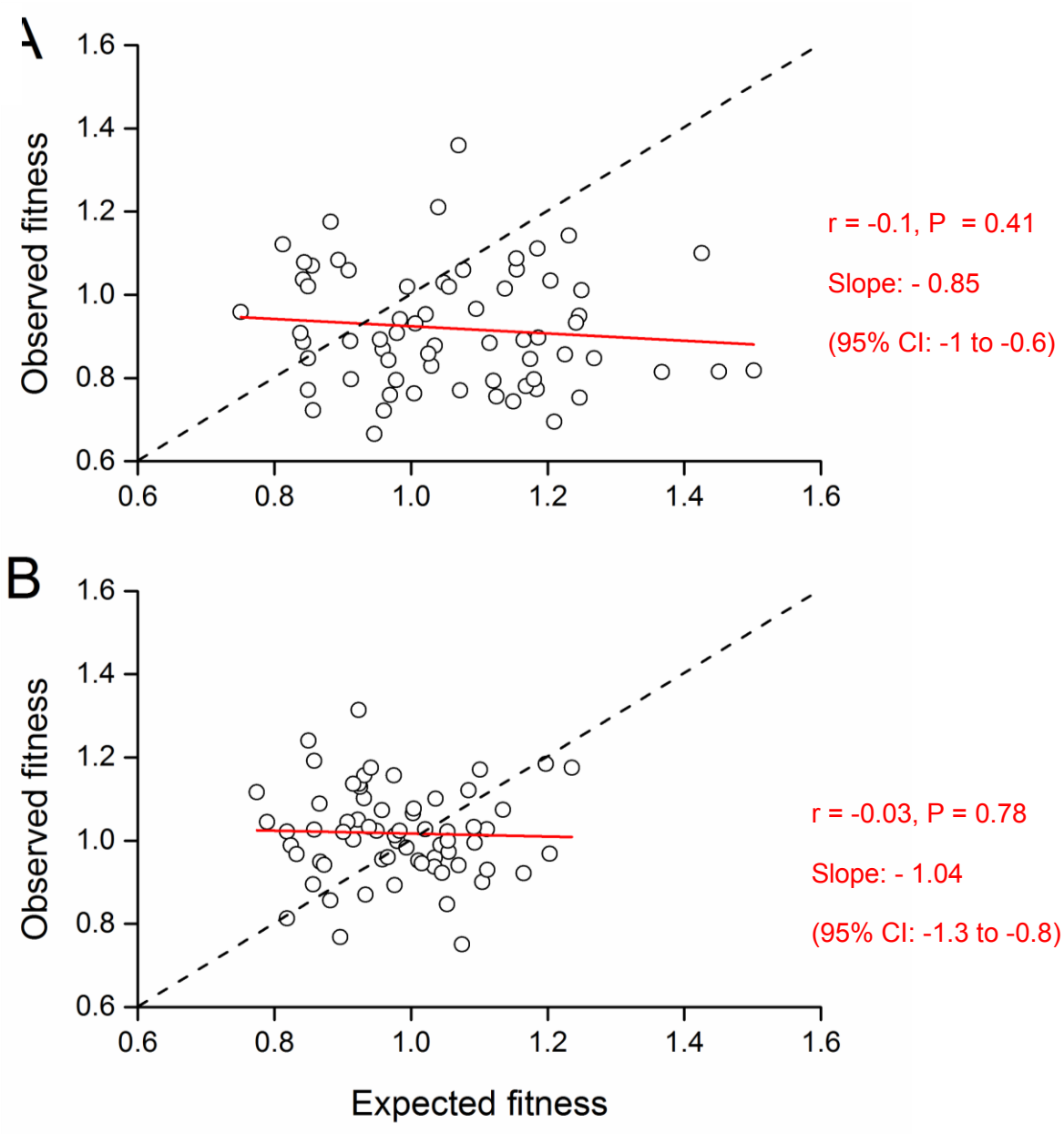


Figure S3. Type II Standard Major Axis (SMA) regression of observed and expected fitness as determined in (A) the fructose- and (B) the succinate-containing environment. The solid red line represents the regression ($P > 0.05$, $n = 66$), while the dotted black line indicates the null model assuming no epistasis.

SUPPORTING TABLES

Table S1. Strains used in this study. Abbreviations: ara$^{+/-}$ = ability to use arabinose as a carbon source present/ absent, AT = auxotroph, WT = wild type.

Strain	Genotype	Phenotype	Reference
Escherichia coli BW25113 ara$^-$	F-, *ΔaraD-araB567*, *ΔlacZ4787*::rrnB-3, *λ*$^-$, *rph-1*, *ΔrhaD-rhaB568*, *hsdR514*	WT Red	Baba et al. 2006
Escherichia coli BW25113 ara$^+$	F-, *ΔaraD-araB567*, *ΔlacZ4787*::rrnB-3, *λ*$^-$, *rph-1*, *ΔrhaD-rhaB568*, *hsdR514, araA*	WT White	D'Souza et al. 2014
Δ*argH*	WT ara$^-$, Δ*argH*::kan^R	AT	D'Souza et al. 2014
Δ*hisD*	WT ara$^-$, Δ*hisD*::kan^R	AT	D'Souza et al. 2014
Δ*ilvA*	WT ara$^-$, Δ*ilvA*::kan^R	AT	D'Souza et al. 2014
Δ*leuB*	WT ara$^-$, Δ*leuB*::kan^R	AT	D'Souza et al. 2014
Δ*lysA*	WT ara$^-$, Δ*lysA*::kan^R	AT	D'Souza et al. 2014
Δ*metA*	WT ara$^-$, Δ*metA*::kan^R	AT	D'Souza et al. 2014
Δ*pheA*	WT ara$^-$, Δ*pheA*::kan^R	AT	D'Souza et al. 2014
Δ*proC*	WT ara$^-$, Δ*proC*::kan^R	AT	D'Souza et al. 2014
Δ*thrC*	WT ara$^-$, Δ*thrC*::kan^R	AT	D'Souza et al. 2014
Δ*trpB*	WT ara$^-$, Δ*trpB*::kan^R	AT	D'Souza et al. 2014
Δ*tyrA*	WT ara$^-$, Δ*tyrA*::kan^R	AT	D'Souza et al. 2014
Δ*argH* Δ*ilvA*	WT ara$^-$, Δ*argH,* Δ*ilvA*::kan^R	AT	This study
Δ*argH* Δ*leuB*	WT ara$^-$, Δ*argH,* Δ*leuB*::kan^R	AT	This study
Δ*argH* Δ*lysA*	WT ara$^-$, Δ*argH*, Δ*lysA*::kan^R	AT	This study
Δ*metA* Δ*argH*	WT ara$^-$, Δ*metA,* Δ*argH*::kan^R	AT	This study
Δ*argH* Δ*pheA*	WT ara$^-$, Δ*argH,* Δ*pheA*::kan^R	AT	This study
Δ*proC* Δ*argH*	WT ara$^-$, Δ*proC,* Δ*argH*::kan^R	AT	This study
Δ*argH* Δ*thrC*	WT ara$^-$, Δ*argH,* Δ*thrC*::kan^R	AT	This study
Δ*argH* Δ*trpB*	WT ara$^-$, Δ*argH,* Δ*trpB*::kan^R	AT	This study
Δ*argH* Δ*tyrA*	WT ara$^-$, Δ*argH,* Δ*tyrA*::kan^R	AT	This study
Δ*ilvA* Δ*hisD*	WT ara$^-$, Δ*ilvA,* Δ*hisD*::kan^R	AT	This study
Δ*leuB* Δ*hisD*	WT ara$^-$, Δ*leuB,* Δ*hisD*::kan^R	AT	This study
Δ*lysA* Δ*hisD*	WT ara$^-$, Δ*lysA,* Δ*hisD*::kan^R	AT	This study
Δ*metA* Δ*hisD*	WT ara$^-$, Δ*metA,* Δ*hisD*::kan^R	AT	This study
Δ*hisD* Δ*pheA*	WT ara$^-$, Δ*hisD,* Δ*pheA*::kan^R	AT	This study
Δ*hisD* Δ*proC*	WT ara$^-$, Δ*hisD,* Δ*proC*::kan^R	AT	This study
Δ*hisD* Δ*thrC*	WT ara$^-$, Δ*hisD,* Δ*thrC*::kan^R	AT	This study
Δ*hisD* Δ*trpB*	WT ara$^-$, Δ*hisD,* Δ*trpB*::kan^R	AT	This study
Δ*hisD* Δ*tyrA*	WT ara$^-$, Δ*hisD,* Δ*tyrA*::kan^R	AT	This study
Δ*ilvA* Δ*leuB*	WT ara$^-$, Δ*ilvA,* Δ*leuB*::kan^R	AT	This study

Δ*ilvA* Δ*lysA*	WT ara$^-$, Δ*ilvA,* Δ*lysA*::kan^R	AT	This study
Δ*ilvA* Δ*metA*	WT ara$^-$, Δ*ilvA,* Δ*metA*::kan^R	AT	This study
Δ*ilvA* Δ*pheA*	WT ara$^-$, Δ*ilvA,* Δ*pheA*::kan^R	AT	This study
Δ*ilvA* Δ*proC*	WT ara$^-$, Δ*ilvA,* Δ*proC*::kan^R	AT	This study
Δ*ilvA* Δ*thrC*	WT ara$^-$, Δ*ilvA,* Δ*thrC*::kan^R	AT	This study
Δ*trpB* Δ*ilvA*	WT ara$^-$, Δ*trpB,* Δ*ilvA*::kan^R	AT	This study
Δ*ilvA* Δ*tyrA*	WT ara$^-$, Δ*ilvA,* Δ*tyrA*::kan^R	AT	This study
Δ*leuB* Δ*lysA*	WT ara$^-$, Δ*leuB,* Δ*lysA*::kan^R	AT	This study
Δ*metA* Δ*leuB*	WT ara$^-$, Δ*metA,* Δ*leuB*::kan^R	AT	This study
Δ*pheA* Δ*leuB*	WT ara$^-$, Δ*pheA,* Δ*leuB*::kan^R	AT	This study
Δ*proC* Δ*leuB*	WT ara$^-$, Δ*proC,* Δ*leuB*::kan^R	AT	This study
Δ*thrC* Δ*leuB*	WT ara$^-$, Δ*thrC,* Δ*leuB*::kan^R	AT	This study
Δ*leuB* Δ*trpB*	WT ara$^-$, Δ*leuB,* Δ*trpB*::kan^R	AT	This study
Δ*lysA* Δ*metA*	WT ara$^-$, Δ*lysA,* Δ*metA*::kan^R	AT	This study
Δ*lysA* Δ*pheA*	WT ara$^-$, Δ*lysA,* Δ*pheA* ::kan^R	AT	This study
Δ*lysA* Δ*proC*	WT ara$^-$, Δ*lysA,* Δ*proC*::kan^R	AT	This study
Δ*thrC* Δ*lysA*	WT ara$^-$, Δ*thrC,* Δ*lysA*::kan^R	AT	This study
Δ*lysA* Δ*trpB*	WT ara$^-$, Δ*lysA*, Δ*trpB*::kan^R	AT	This study
Δ*metA* Δ*pheA*	WT ara$^-$, Δ*metA,* Δ*pheA*::kan^R	AT	This study
Δ*proC* Δ*metA*	WT ara$^-$, Δ*proC,* Δ*metA*::kan^R	AT	This study
Δ*metA* Δ*thrC*	WT ara$^-$, Δ*metA,* Δ*thrC*::kan^R	AT	This study
Δ*metA* Δ*trpB*	WT ara$^-$, Δ*metA,* Δ*trpB*::kan^R	AT	This study
Δ*pheA* Δ*proC*	WT ara$^-$, Δ*pheA,* Δ*proC*::kan^R	AT	This study
Δ*pheA* Δ*thrC*	WT ara$^-$, Δ*pheA,* Δ*thrC*::kan^R	AT	This study
Δ*pheA* Δ*trpB*	WT ara$^-$, Δ*pheA,* Δ*trpB*::kan^R	AT	This study
Δ*pheA* Δ*tyrA*	WT ara$^-$, Δ*pheA,* Δ*tyrA*::kan^R	AT	This study
Δ*proC* Δ*thrC*	WT ara$^-$, Δ*proC,* Δ*thrC*::kan^R	AT	This study
Δ*trpB* Δ*proC*	WT ara$^-$, Δ*trpB* , Δ*proC*::kan^R	AT	This study
Δ*thrC* Δ*trpB*	WT ara$^-$, Δ*thrC,* Δ*trpB*::kan^R	AT	This study
Δ*thrC* Δ*tyrA*	WT ara$^-$, Δ*thrC,* Δ*tyrA*::kan^R	AT	This study
Δ*trpB* Δ*tyrA*	WT ara$^-$, Δ*trpB,* Δ*tyrA*::kan^R	AT	This study
Δ*trpB* Δ*pheA* Δ*tyrA*	WT ara$^-$, Δ*trpB,* Δ*pheA,* Δ*tyrA*::kan^R	AT	This study
Δ*lysA* Δ*metA* Δ*argH*	WT ara$^-$, Δ*lysA,* Δ*metA,* Δ*argH*::kan^R	AT	This study
Δ*lysA* Δ*metA* Δ*thrC*	WT ara$^-$, Δ *lysA* Δ*metA* Δ*thrC*::kan^R	AT	This study
Δ*trpB* Δ*pheA* Δ*metA*	WT ara$^-$, Δ*trpB,* Δ*pheA,* Δ*metA*::kan^R	AT	This study
Δ*trpB* Δ*pheA* Δ*leuB*	WT ara$^-$, Δ*trpB,* Δ*pheA,* Δ*leuB*::kan^R	AT	This study
Δ*metA* Δ*thrC* Δ*argH*	WT ara$^-$, Δ*metA,* Δ*thrC,* Δ*argH*::kan^R	AT	This study
Δ*metA* Δ*thrC* Δ*hisD*	WT ara$^-$, Δ*metA,* Δ*thrC,* Δ*hisD*::kan^R	AT	This study
Δ*thrC* Δ*lysA* Δ*hisD*	WT ara$^-$, Δ*thrC,* Δ*lysA,* Δ*hisD*::kan^R	AT	This study
Δ*trpB* Δ*pheA* Δ*hisD*	WT ara$^-$, Δ*trpB,* Δ*pheA,* Δ*hisD*::kan^R	AT	This study

Δ*proC* Δ*thrC* Δ*ilvA*	WT ara$^-$, Δ*proC,* Δ*thrC,* Δ*ilvA*::kan^R	AT	This study
Δ*trpB* Δ*leuB* Δ*thrC*	WT ara$^-$, Δ*trpB,* Δ*leuB,* Δ*thrC*::kan^R	AT	This study
Δ*metA* Δ*argH* Δ*hisD*	WT ara$^-$, Δ*metA,* Δ*argH,* Δ*hisD*::kan^R	AT	This study
Δ*proC* Δ*lysA* Δ*hisD*	WT ara$^-$, Δ*proC,* Δ*lysA,* Δ*hisD*::kan^R	AT	This study
Δ*proC* Δ*lysA* Δ*tyrA*	WT ara$^-$, Δ*proC,* Δ*lysA,* Δ*tyrA*::kan^R	AT	This study
Δ*ilvA* Δ*thrC* Δ*trpB*	WT ara$^-$, Δ*ilvA,* Δ*thrC,* Δ*trpB*::kan^R	AT	This study
Δ*trpB* Δ*pheA* Δ*thrC*	WT ara$^-$, Δ*trpB,* Δ*pheA,* Δ*thrC*::kan^R	AT	This study
Δ*argH*::kan^S	WT ara$^-$, Δ*argH*::kan^S	AT	This study
Δ*hisD*::kan^S	WT ara$^-$, Δ*hisD*::kan^S	AT	This study
Δ*ilvA*::kan^S	WT ara$^-$, Δ*ilvA*::kan^S	AT	This study
Δ*leuB*::kan^S	WT ara$^-$, Δ*leuB*::kan^S	AT	This study
Δ*lysA*::kan^S	WT ara$^-$, Δ*lysA*::kan^S	AT	This study
Δ*metA*::kan^S	WT ara$^-$, Δ*metA*::kan^S	AT	This study
Δ*pheA*::kan^S	WT ara$^-$, Δ*pheA*::kan^S	AT	This study
Δ*proC*::kan^S	WT ara$^-$, Δ*proC*::kan^S	AT	This study
Δ*thrC*::kan^S	WT ara$^-$, Δ*thrC*::kan^S	AT	This study
Δ*trpB*::kan^S	WT ara$^-$, Δ*trpB*::kan^S	AT	This study
Δ*ilvA* Δ*leuB*::kan^S	WT ara$^-$, Δ*ilvA,* Δ*leuB*::kan^S	AT	This study
Δ*ilvA* Δ*thrC*::kan^S	WT ara$^-$, Δ*ilvA,* Δ*thrC*::kan^S	AT	This study
Δ*thrC* Δ*trpB*::kan^S	WT ara$^-$, Δ*thrC,* Δ*trpB*::kan^S	AT	This study
Δ*thrC* Δ*trpB*::kan^S	WT ara$^-$, Δ*thrC,* Δ*trpB*::kan^S	AT	This study
Δ*ilvA* Δ*thrC* Δ*trpB*::kan^S	WT ara$^-$, Δ*ilvA,* Δ*thrC,* Δ*trpB*::kan^S	AT	This study
Δ*lysA* Δ*metA* Δ*thrC*::kan^S	WT ara$^-$, Δ*lysA,* Δ*metA,* Δ*thrC*::kan^S	AT	This study
Δ*trpB* Δ*pheA* Δ*thrC*::kan^S	WT ara$^-$, Δ*trpB,* Δ*pheA,* Δ*thrC*::kan^S	AT	This study
Δ*trpB* Δ*leuB* Δ*thrC*::kan^S	WT ara$^-$, Δ*trpB,* Δ*leuB,* Δ*thrC*::kan^S	AT	This study
Δ*argH*	WT, Δ*argH*::kan^R	AT	Baba et al. 2006
Δ*hisD*	WT, Δ*hisD*::kan^R	AT	Baba et al. 2006
Δ*ilvA*	WT, Δ*ilvA*::kan^R	AT	Baba et al. 2006
Δ*leuB*	WT, Δ*leuB*::kan^R	AT	Baba et al. 2006
Δ*lysA*	WT, Δ*lysA*::kan^R	AT	Baba et al. 2006
Δ*metA*	WT, Δ*metA*::kan^R	AT	Baba et al. 2006
Δ*pheA*	WT, Δ*pheA*::kan^R	AT	Baba et al. 2006
Δ*proC*	WT, Δ*proC*::kan^R	AT	Baba et al. 2006
Δ*thrC*	WT, Δ*thrC*::kan^R	AT	Baba et al. 2006
Δ*trpB*	WT, Δ*trpB*::kan^R	AT	Baba et al. 2006
Δ*tyrA*	WT , Δ*tyrA*::kan^R	AT	Baba et al. 2006

Table S2. Fitness cost of the kanamycin resistance marker. Mean Malthusian parameter (± 95% confidence interval (CI)) of kanamycin resistant (kan^R) and sensitive (kan^S) auxotrophic mutants was determined by coculturing both competitors in the fructose-containing environment for 24 h. Each comparison has been replicated 10 times. P values of independent sample t-tests are given. % MDD = minimum detectable difference calculated as described (Zar 1999).

Genotype	Malthusian parameter	± 95% CI	P value	% MDD
Δ*argH*::kan^R	5.59	0.24	0.95	2.80
Δ*argH*::kan^S	5.60	0.20		
Δ*hisD*::kan^R	4.64	0.10	0.10	0.24
Δ*hisD*::kan^S	4.76	0.09		
Δ*ilvA*::kan^R	5.93	0.30	0.73	2.24
Δ*ilvA*::kan^S	6.01	0.28		
Δ*leuB*::kan^R	5.48	0.13	0.31	0.92
Δ*leuB*::kan^S	5.55	0.11		
Δ*lysA*::kan^R	5.10	0.17	0.31	0.85
Δ*lysA*::kan^S	5.01	0.05		
Δ*metA*::kan^R	5.24	0.07	0.66	1.84
Δ*metA*::kan^S	5.27	0.07		
Δ*pheA*::kan^R	5.09	0.10	0.62	1.67
Δ*pheA*::kan^S	5.05	0.09		
Δ*proC*::kan^R	4.89	0.11	0.05	0.13
Δ*proC*::kan^S	4.75	0.06		
Δ*thrC*::kan^R	5.24	0.05	0.82	2.27
Δ*thrC*::kan^S	5.23	0.06		
Δ*trpB*::kan^R	6.29	0.07	0.65	2.15
Δ*trpB*::kan^S	6.27	0.06		
Δ*ilvA* Δ*leuB*::kan^R	4.82	0.24	0.91	2.31
Δ*ilvA* Δ*leuB*::kan^S	4.83	0.18		
Δ*ilvA* Δ*thrC*::kan^R	5.38	0.67	0.33	0.93
Δ*ilvA* Δ*thrC*::kan^S	4.88	0.73		
Δ*lysA* Δ*metA*::kan^R	5.04	0.07	0.67	1.77
Δ*lysA* Δ*metA*::kan^S	5.02	0.07		
Δ*thrC* Δ*trpB*::kan^R	4.92	0.58	0.42	1.07
Δ*thrC* Δ*trpB*::kan^S	4.59	0.52		
Δ*ilvA* Δ*thrC* Δ*trpB*::kan^R	6.19	0.19	0.41	1.35
Δ*ilvA* Δ*thrC* Δ*trpB*::kan^S	6.29	0.11		
Δ*lysA* Δ*metA* Δ*thrC*::kan^R	5.66	0.09	0.91	2.72
Δ*lysA* Δ*metA* Δ*thrC*::kan^S	5.67	0.14		
Δ*trpB* Δ*pheA* Δ*thrC*::kan^R	5.87	0.07	0.68	2.09
Δ*trpB* Δ*pheA* Δ*thrC*::kan^S	5.89	0.08		
Δ*trpB* Δ*leuB* Δ*thrC*::kan^R	5.78	0.07	0.49	1.51
Δ*trpB* Δ*leuB* Δ*thrC*::kan^S	5.83	0.12		

Table S3. Relative fitness and epistatic interactions among auxotrophy-causing mutations in the fructose-containing environment. Mean fitness of each mutant genotype relative to wild type (± 95% confidence interval (CI)) was calculated from 8 replicates. Epistasis was estimated by comparing estimated and observed fitness values using a multiplicative model. Instances of significant epistasis are depicted in bold. NA = not applicable.

Genotype	Relative fitness	± 95% CI	Epistasis
Δ*argH*	0.84	0.08	NA
Δ*hisD*	0.95	0.05	NA
Δ*ilvA*	1.23	0.13	NA
Δ*leuB*	1.02	0.08	NA
Δ*lysA*	0.88	0.05	NA
Δ*metA*	1.21	0.07	NA
Δ*pheA*	1.17	0.10	NA
Δ*proC*	1.00	0.04	NA
Δ*thrC*	0.99	0.07	NA
Δ*trpB*	0.95	0.08	NA
Δ*tyrA*	1.01	0.04	NA
Δ*argH* Δ*ilvA*	1.02	0.03	-0.01
Δ*argH* Δ*leuB*	1.02	0.03	-0.03
Δ*argH* Δ*lysA*	**0.95**	**0.06**	**0.20**
Δ*metA* Δ*argH*	**0.82**	**0.07**	**-0.20**
Δ*argH* Δ*pheA*	1.02	0.04	0.02
Δ*proC* Δ*argH*	**1.07**	**0.03**	**0.21**
Δ*argH* Δ*thrC*	**1.07**	**0.11**	**0.23**
Δ*argH* Δ*trpB*	**1.12**	**0.06**	**0.30**
Δ*argH* Δ*tyrA*	**0.72**	**0.09**	**-0.13**
Δ*ilvA* Δ*hisD*	**0.84**	**0.05**	**-0.32**
Δ*leuB* Δ*hisD*	0.88	0.09	0.04
Δ*lysA* Δ*hisD*	**1.03**	**0.05**	**0.19**
Δ*metA* Δ*hisD*	1.06	0.05	-0.09
Δ*hisD* Δ*pheA*	**0.88**	**0.06**	**-0.23**
Δ*hisD* Δ*proC*	0.86	0.03	-0.08
Δ*hisD* Δ*thrC*	**0.66**	**0.05**	**-0.28**
Δ*hisD* Δ*trpB*	0.88	0.08	-0.02
Δ*hisD* Δ*tyrA*	**0.72**	**0.03**	**-0.23**
Δ*ilvA* Δ*leuB*	**0.84**	**0.03**	**-0.42**
Δ*ilvA* Δ*lysA*	**0.96**	**0.11**	**-0.12**
Δ*ilvA* Δ*metA*	**0.81**	**0.08**	**-0.68**
Δ*ilvA* Δ*pheA*	**0.81**	**0.06**	**-0.63**
Δ*ilvA* Δ*proC*	**0.75**	**0.06**	**-0.49**
Δ*ilvA* Δ*thrC*	1.14	0.09	-0.08
Δ*trpB* Δ*ilvA*	1.11	0.22	-0.07
Δ*ilvA* Δ*tyrA*	**1.01**	**0.07**	**-0.23**

Δ*leuB* Δ*lysA*	**1.05**	**0.16**	**0.15**
Δ*metA* Δ*leuB*	**0.94**	**0.17**	**-0.29**
Δ*pheA* Δ*leuB*	**1.03**	**0.05**	**-0.17**
Δ*proC* Δ*leuB*	**0.87**	**0.06**	**-0.15**
Δ*thrC* Δ*leuB*	0.95	0.06	-0.06
Δ*leuB* Δ*trpB*	0.94	0.05	-0.04
Δ*lysA* Δ*metA*	1.06	0.04	-0.01
Δ*lysA* Δ*pheA*	**1.21**	**0.12**	**0.17**
Δ*lysA* Δ*proC*	**1.08**	**0.07**	**0.19**
Δ*thrC* Δ*lysA*	**1.17**	**0.08**	**0.29**
Δ*lysA* Δ*trpB*	**1.02**	**0.08**	**0.17**
Δ*metA* Δ*pheA*	**1.10**	**0.09**	**-0.32**
Δ*proC* Δ*metA*	**0.85**	**0.02**	**-0.36**
Δ*metA* Δ*thrC*	**0.69**	**0.04**	**-0.51**
Δ*metA* Δ*trpB*	**0.89**	**0.12**	**-0.27**
Δ*pheA* Δ*proC*	**0.77**	**0.06**	**-0.41**
Δ*pheA* Δ*thrC*	**0.78**	**0.10**	**-0.38**
Δ*pheA* Δ*trpB*	**0.75**	**0.09**	**-0.37**
Δ*pheA* Δ*tyrA*	**0.89**	**0.08**	**-0.28**
Δ*proC* Δ*thrC*	**0.76**	**0.08**	**-0.24**
Δ*trpB* Δ*proC*	**0.84**	**0.05**	**-0.12**
Δ*thrC* Δ*trpB*	0.89	0.10	-0.06
Δ*thrC* Δ*tyrA*	0.93	0.10	-0.07
Δ*trpB* Δ*tyrA*	**0.75**	**0.07**	**-0.21**
Δ*trpB* Δ*pheA* Δ*tyrA*	**1.01**	**0.13**	**0.74**
Δ*lysA* Δ*metA* Δ*argH*	**0.77**	**0.07**	**0.23**
Δ*lysA* Δ*metA* Δ*thrC*	**0.79**	**0.04**	**-0.10**
Δ*trpB* Δ*pheA* Δ*metA*	**0.81**	**0.04**	**0.41**
Δ*trpB* Δ*pheA* Δ*leuB*	**1.08**	**0.12**	**0.51**
Δ*metA* Δ*thrC* Δ*argH*	**0.85**	**0.10**	**0.31**
Δ*metA* Δ*thrC* Δ*hisD*	**0.74**	**0.04**	**0.48**
Δ*thrC* Δ*lysA* Δ*hisD*	**0.90**	**0.06**	**-0.13**
Δ*trpB* Δ*pheA* Δ*hisD*	**0.79**	**0.11**	**0.49**
Δ*proC* Δ*thrC* Δ*ilvA*	**1.35**	**0.12**	**0.91**
Δ*trpB* Δ*leuB* Δ*thrC*	**0.93**	**0.07**	**0.51**
Δ*metA* Δ*argH* Δ*hisD*	**0.90**	**0.08**	**0.10**
Δ*proC* Δ*lysA* Δ*hisD*	**0.79**	**0.07**	**-1.48**
Δ*proC* Δ*lysA* Δ*tyrA*	**0.77**	**0.07**	**-0.37**
Δ*ilvA* Δ*thrC* Δ*trpB*	**0.84**	**0.08**	**0.71**
Δ*trpB* Δ*pheA* Δ*thrC*	**0.79**	**0.07**	**-0.16**

Table S4. Relative fitness and epistatic interactions among auxotrophy-causing mutations in the succinate-containing environment. Mean fitness of each mutant genotype relative to wild type (± 95% confidence interval (CI)) was calculated from 8 replicates. Epistasis was estimated by comparing estimated and observed fitness values using a multiplicative model. Instances of significant epistasis are depicted in bold. NA = not applicable.

Genotype	Relative fitness	± 95% CI	Epistasis
Δ*argH*	0.95	0.15	NA
Δ*hisD*	0.85	0.07	NA
Δ*ilvA*	0.90	0.03	NA
Δ*leuB*	1.02	0.08	NA
Δ*lysA*	0.95	0.06	NA
Δ*metA*	1.14	0.05	NA
Δ*pheA*	1.08	0.05	NA
Δ*proC*	1.04	0.06	NA
Δ*thrC*	0.96	0.05	NA
Δ*trpB*	0.97	0.03	NA
Δ*tyrA*	1.00	0.03	NA
Δ*argH* Δ*ilvA*	**1.08**	**0.06**	**0.22**
Δ*argH* Δ*leuB*	1.02	0.08	-0.08
Δ*argH* Δ*lysA*	1.00	0.05	0.08
Δ*metA* Δ*argH*	**1.03**	**0.07**	**-0.05**
Δ*argH* Δ*pheA*	**0.95**	**0.03**	**-0.07**
Δ*proC* Δ*argH*	1.06	0.07	0.06
Δ*argH* Δ*thrC*	**1.05**	**0.05**	**0.12**
Δ*argH* Δ*trpB*	**1.10**	**0.10**	**0.17**
Δ*argH* Δ*tyrA*	0.95	0.04	-0.00
Δ*ilvA* Δ*hisD*	**1.11**	**0.05**	**0.34**
Δ*leuB* Δ*hisD*	**1.02**	**0.03**	**0.20**
Δ*lysA* Δ*hisD*	0.81	0.07	-0.00
Δ*metA* Δ*hisD*	**0.89**	**0.07**	**-0.08**
Δ*hisD* Δ*pheA*	**1.12**	**0.03**	**0.20**
Δ*hisD* Δ*proC*	**0.76**	**0.06**	**-0.12**
Δ*hisD* Δ*thrC*	**0.98**	**0.06**	**0.16**
Δ*hisD* Δ*trpB*	**0.96**	**0.04**	**0.13**
Δ*hisD* Δ*tyrA*	0.89	0.05	0.03
Δ*ilvA* Δ*leuB*	**1.13**	**0.04**	**0.21**
Δ*ilvA* Δ*lysA*	0.94	0.08	0.08
Δ*ilvA* Δ*metA*	**0.93**	**0.03**	**-0.09**
Δ*ilvA* Δ*pheA*	0.99	0.05	0.01
Δ*ilvA* Δ*proC*	1.02	0.05	0.07
Δ*ilvA* Δ*thrC*	0.94	0.06	0.06
Δ*trpB* Δ*ilvA*	0.85	0.04	-0.02
Δ*ilvA* Δ*tyrA*	**1.04**	**0.11**	**0.13**

Δ*leuB* Δ*lysA*	1.01	0.05	0.03
Δ*metA* Δ*leuB*	**0.92**	**0.07**	**-0.24**
Δ*pheA* Δ*leuB*	**0.90**	**0.03**	**-0.20**
Δ*proC* Δ*leuB*	**0.94**	**0.08**	**-0.12**
Δ*thrC* Δ*leuB*	1.02	0.04	0.04
Δ*leuB* Δ*trpB*	**0.98**	**0.03**	**-0.01**
Δ*lysA* Δ*metA*	**0.99**	**0.11**	**-0.09**
Δ*lysA* Δ*pheA*	1.10	0.07	0.06
Δ*lysA* Δ*proC*	1.07	0.07	0.07
Δ*thrC* Δ*lysA*	**1.31**	**0.13**	**0.39**
Δ*lysA* Δ*trpB*	**1.15**	**0.04**	**0.22**
Δ*metA* Δ*pheA*	**1.17**	**0.07**	**-0.06**
Δ*proC* Δ*metA*	1.18	0.05	-0.01
Δ*metA* Δ*thrC*	1.17	0.10	0.07
Δ*metA* Δ*trpB*	**0.93**	**0.06**	**-0.18**
Δ*pheA* Δ*proC*	**1.07**	**0.18**	**-0.06**
Δ*pheA* Δ*thrC*	0.98	0.08	-0.05
Δ*pheA* Δ*trpB*	1.02	0.10	-0.03
Δ*pheA* Δ*tyrA*	1.12	0.17	0.03
Δ*proC* Δ*thrC*	**0.95**	**0.09**	**-0.05**
Δ*trpB* Δ*proC*	1.02	0.04	0.00
Δ*thrC* Δ*trpB*	**1.03**	**0.05**	**0.09**
Δ*thrC* Δ*tyrA*	0.96	0.04	-0.00
Δ*trpB* Δ*tyrA*	**1.15**	**0.13**	**0.18**
Δ*trpB* Δ*pheA* Δ*tyrA*	**0.97**	**0.02**	**-0.26**
Δ*lysA* Δ*metA* Δ*argH*	**1.00**	**0.04**	**-0.36**
Δ*lysA* Δ*metA* Δ*thrC*	0.92	0.02	-0.05
Δ*trpB* Δ*pheA* Δ*metA*	0.96	0.08	0.04
Δ*trpB* Δ*pheA* Δ*leuB*	0.75	0.11	-0.07
Δ*metA* Δ*thrC* Δ*argH*	**0.84**	**0.05**	**-0.34**
Δ*metA* Δ*thrC* Δ*hisD*	1.17	0.14	0.08
Δ*thrC* Δ*lysA* Δ*hisD*	**1.04**	**0.11**	**-0.29**
Δ*trpB* Δ*pheA* Δ*hisD*	0.94	0.17	-0.07
Δ*proC* Δ*thrC* Δ*ilvA*	**1.02**	**0.06**	**-0.18**
Δ*trpB* Δ*leuB* Δ*thrC*	**1.13**	**0.05**	**0.13**
Δ*metA* Δ*argH* Δ*hisD*	1.07	0.06	-0.00
Δ*proC* Δ*lysA* Δ*hisD*	**0.87**	**0.04**	**-1.66**
Δ*proC* Δ*lysA* Δ*tyrA*	**1.02**	**0.07**	**0.22**
Δ*ilvA* Δ*thrC* Δ*trpB*	**1.19**	**0.08**	**0.93**
Δ*trpB* Δ*pheA* Δ*thrC*	**1.24**	**0.06**	**0.25**

SUPPORTING REFERENCES

Adadi, R., B. Volkmer, R. Milo, M. Heinemann, and T. Shlomi. 2012. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. PLoS Comp. Biol. 8:e1002575.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2:0008.

D'Souza, G., S. Waschina, S. Pande, K. Bohl, C. Kaleta, and C. Kost. 2014. Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution 68-9: 2559–2570.

Orth, J. D., T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, and B. Ø. Palsson. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Mol. Syst. Biol. 7:535.

Schellenberger, J., R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, and S. Rahmanian. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nat. Protoc. 6:1290-1307.

Zar, J. H. 1999. Biostatistical analysis. 4th ed. Prentice-Hall, Inc., New Jersey, USA.

Supporting information for chapter 3

Table S1. Final amino acid concentrations (μM) in the media used for the precultures of auxotrophs and for the growth kinetic assays.

Amino acid	Auxotroph precultures	Growth kinetic assays							
		Lvl 1	Lvl 2	Lvl 3	Lvl 4	Lvl 5	Lvl 6	Lvl 7	Lvl 8
His	15	0	2.5	3.75	5	6.25	7.5	8.75	10
Tyr	30	0	5	7.5	10	12.5	15	17.5	20
Phe	30	0	5	7.5	10	12.5	15	17.5	20
Trp	150	0	25	37.5	50	62.5	75	87.5	100
Leu	60	0	10	15	20	25	30	35	40
Lys	60	0	10	15	20	25	30	35	40
Ile	45	0	7.5	11.25	15	18.75	22.5	26.25	30

Table S2. Strains used in this study.

Strain	Genotype	Phenotype	Reference
Escherichia coli **BW25113 ara$^-$**	F-, ⊿*araD-araB567*, ⊿*lacZ4787*::rrnB-3, *λ*$^-$, *rph-1*, ⊿*rhaD-rhaB568*, *hsdR514*	WT	Baba *et al.* (2006)
Δ*hisD*	WT ara$^-$, Δ*hisD*::kan^R	AT	D'Souza et al. (2014)
Δ*pheA*	WT ara$^-$, Δ*pheA*::kan^R	AT	D'Souza et al. (2014)
Δ*tyrA*	WT ara$^-$, Δ*tyrA*::kan^R	AT	D'Souza et al. (2014)
Δ*trpB*	WT ara$^-$, Δ*trpB*::kan^R	AT	D'Souza et al. (2014)
Δ*leuB*	WT ara$^-$, Δ*leuB*::kan^R	AT	D'Souza et al. (2014)
Δ*lysA*	WT ara$^-$, Δ*lysA*::kan^R	AT	D'Souza et al. (2014)
Δ*ilvA*	WT ara$^-$, Δ*ilvA*::kan^R	AT	D'Souza et al. (2014)
Δ*hisD*	WT ara$^-$, Δ*hisD*	AT, kan^S	This study
Δ*pheA*	WT ara$^-$, Δ*pheA*	AT, kan^S	This study
Δ*trpB*	WT ara$^-$, Δ*trpB*	AT, kan^S	This study
Δ*leuB*	WT ara$^-$, Δ*leuB*	AT, kan^S	This study
Δ*lysA*	WT ara$^-$, Δ*lysA*	AT, kan^S	This study
Δ*ilvA*	WT ara$^-$, Δ*ilvA*	AT, kan^S	This study

Abbreviations: ara$^-$ – inability to use arabinose as carbon source, WT – wild type, AT – auxotroph, kan^S – kanamycine sensitive.

References

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2006, **2**:2006.0008.

D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, Kost C: Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution 2014, **68**:2559–2570.

Table S3. Carbon sources considered for biosynthetic cost estimation.

D-fructose, L-lactate, succinate, L-malate, α-ketoglutarate, D-galactose, maltose, D-glucsoe, pyruvate, acetate, L-arabinose, N-acetyl-D-glucosamine, D-glucarate, L-aspartate, D-alanine, threhalose, D-mannose, D-sorbitol, glycerol, L-fucose, D-glucuronate, D-gluconate, glycerol 3-phosphate, D-xylose, D-mannitol, L-glutamate, D-glucose 6-phosphate, D-malate, D-ribose, L-rhamnose, melibiose, thymidine, L-asparagine, octadecenoate, fumarate, butyrate, phenylacetaldehyde, 5-dehydro-D-gluconate, acetoacetate, adenosine, L-alanine, D-allose, D-fructose 6-phosphate, D-galactarate, galactitol, D-galacturonate, D-glucosamine, deoxyadenosine, dihydroxyacetone, L-glutamine, inosine, (S)-Propane-1,2-diol, L-tartrate, lactose, maltotriose, N-acetyl-D-mannosamine, N-acetylneuraminate, propionate, uridine, D-glucose 1-phosphate, and L-lyxose.

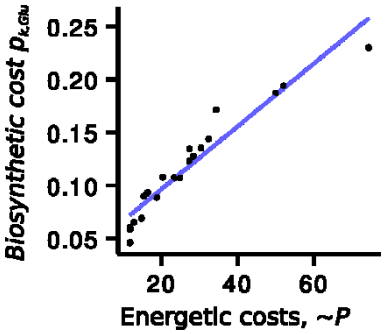


Figure S1. Biosynthetic cost estimations from this study are in line with previously reported estimations (Akashi and Gojobori 2002).

Reference

Akashi H, Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proc Natl Acad Sci USA*. 2002, **99**:3695–3700.

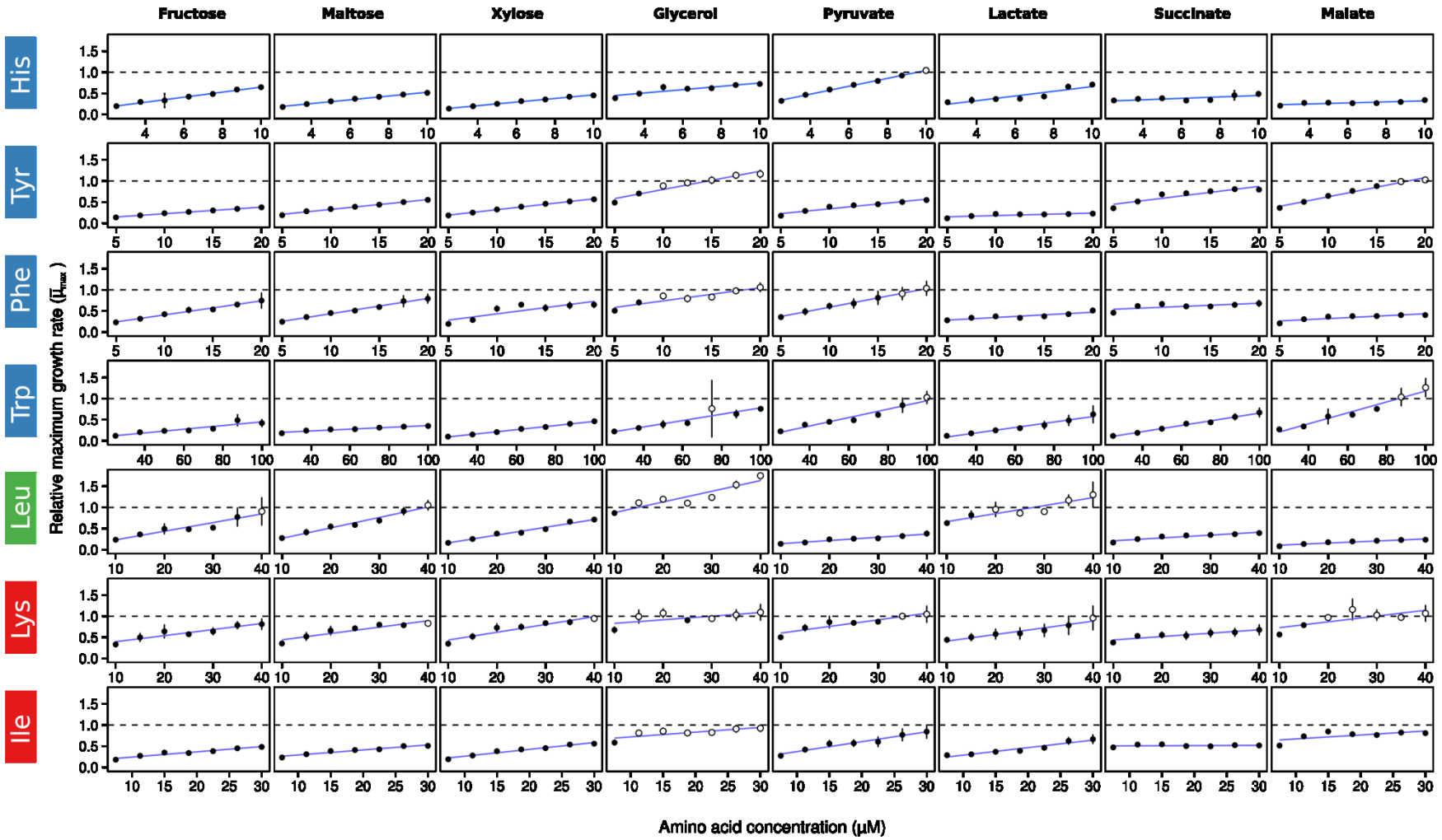


Figure S2. Maximum growth rates of auxotrophs under various carbon sources and amino acid concentrations relative to the maximum growth rate level of the wild type growing under the same carbon source and without amino acid supplementation (=1, dashed line). Error bars indicate the 95% confidence intervals. Filled circles denote the growth rates of the auxotrophs which are significantly lower than the WT strain growth rate (FDR-corrected Welch two sample t-tests, $P < 0.05$, $n = 6$), empty circles indicate no significant difference.

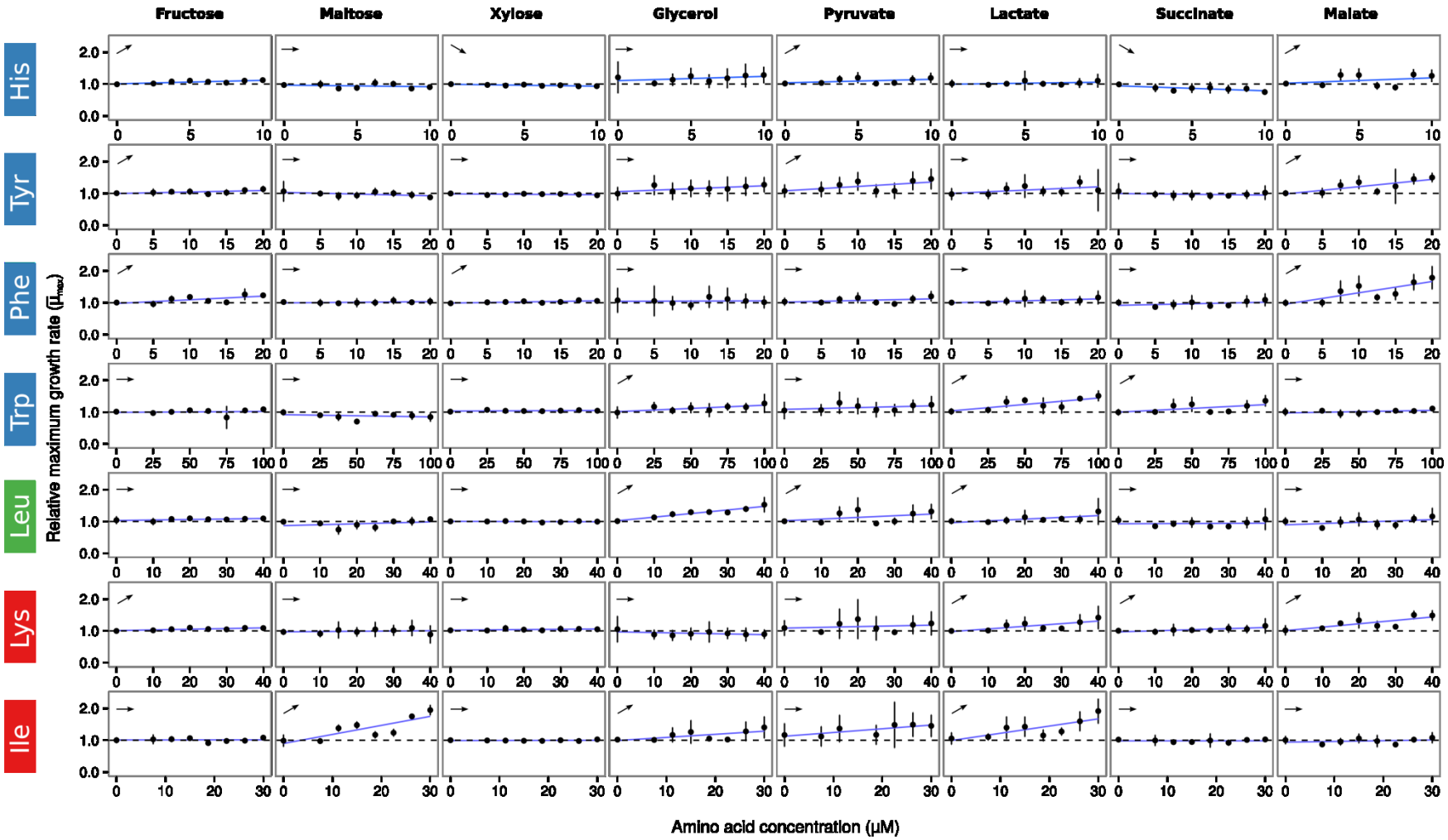
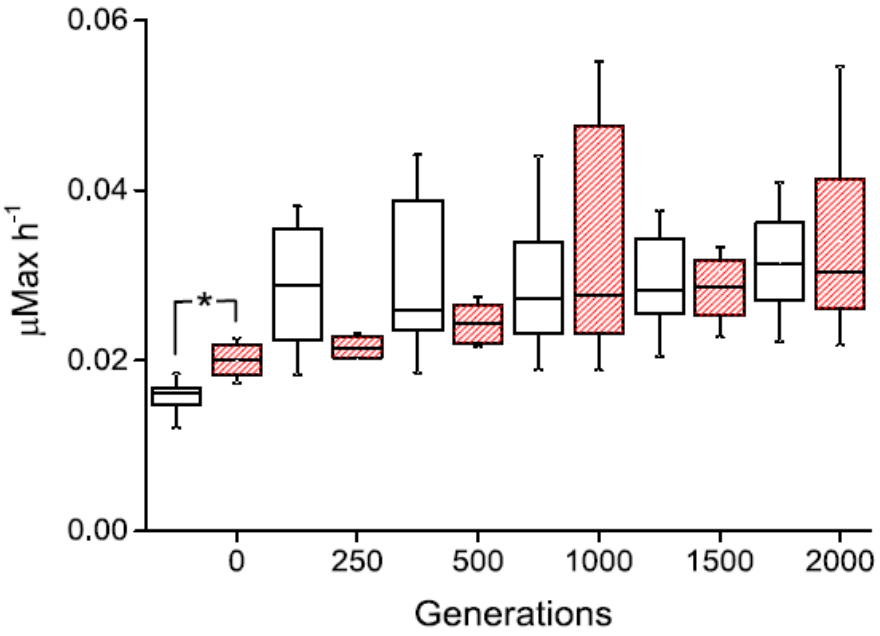


Figure S3. Maximum growth rates of the *E. coli* wild type strain under various carbon sources and amino acid concentrations relative to the maximum growth rate level of the wild type growing under the same carbon source and without amino acid supplementation (=1, dashed line). Error bars indicate the 95% confidence intervals. Arrows indicate significant correlation (up- or down arrows) or no significant correlation (horizontal arrows) of the two axes (FDR corrected linear mixed-model fit by maximizing the restricted log-likelihood, n=48).

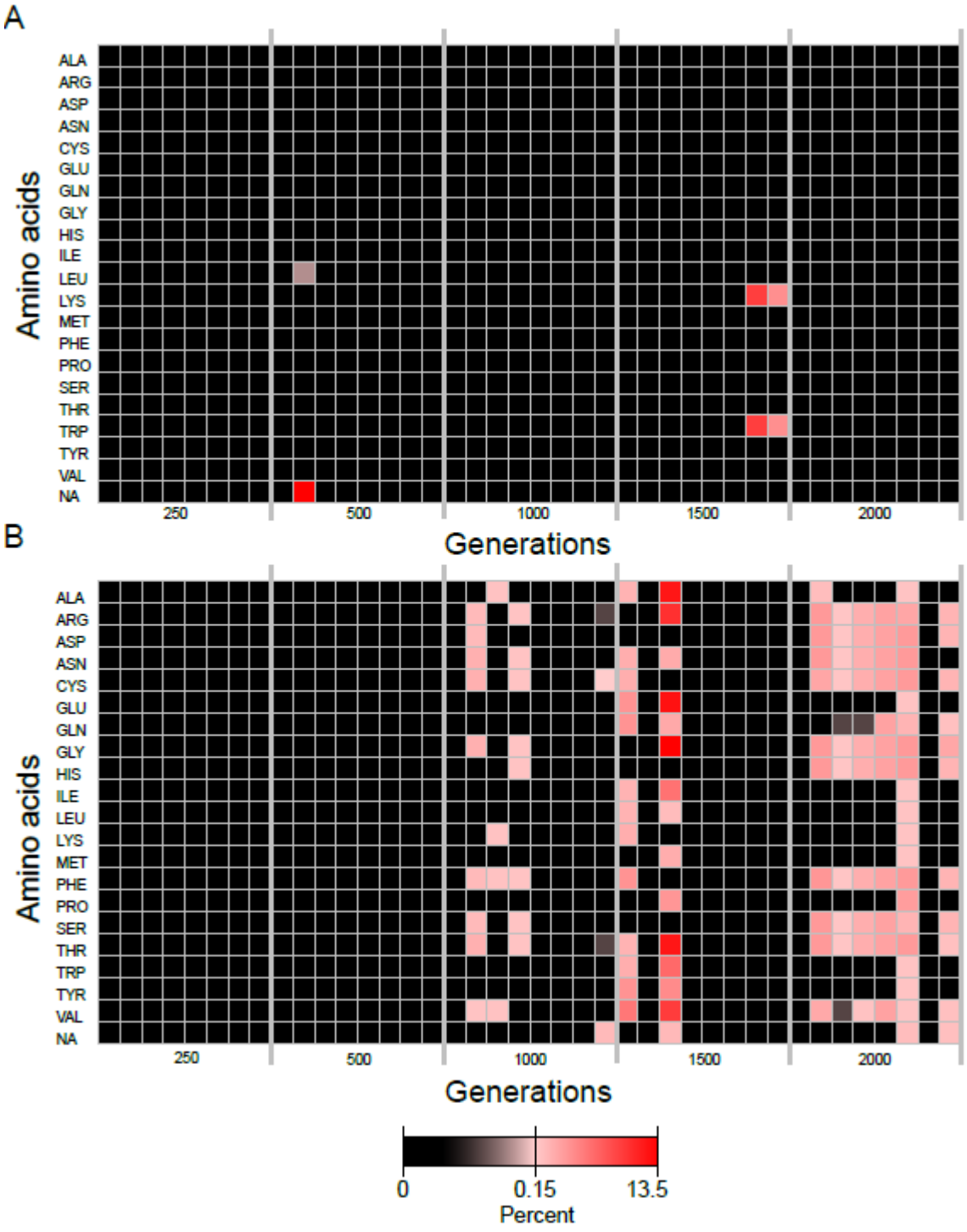
Supporting information for chapter 4

Figure S1: Growth rates relative to the ancestor of populations evolving under the two different regimes



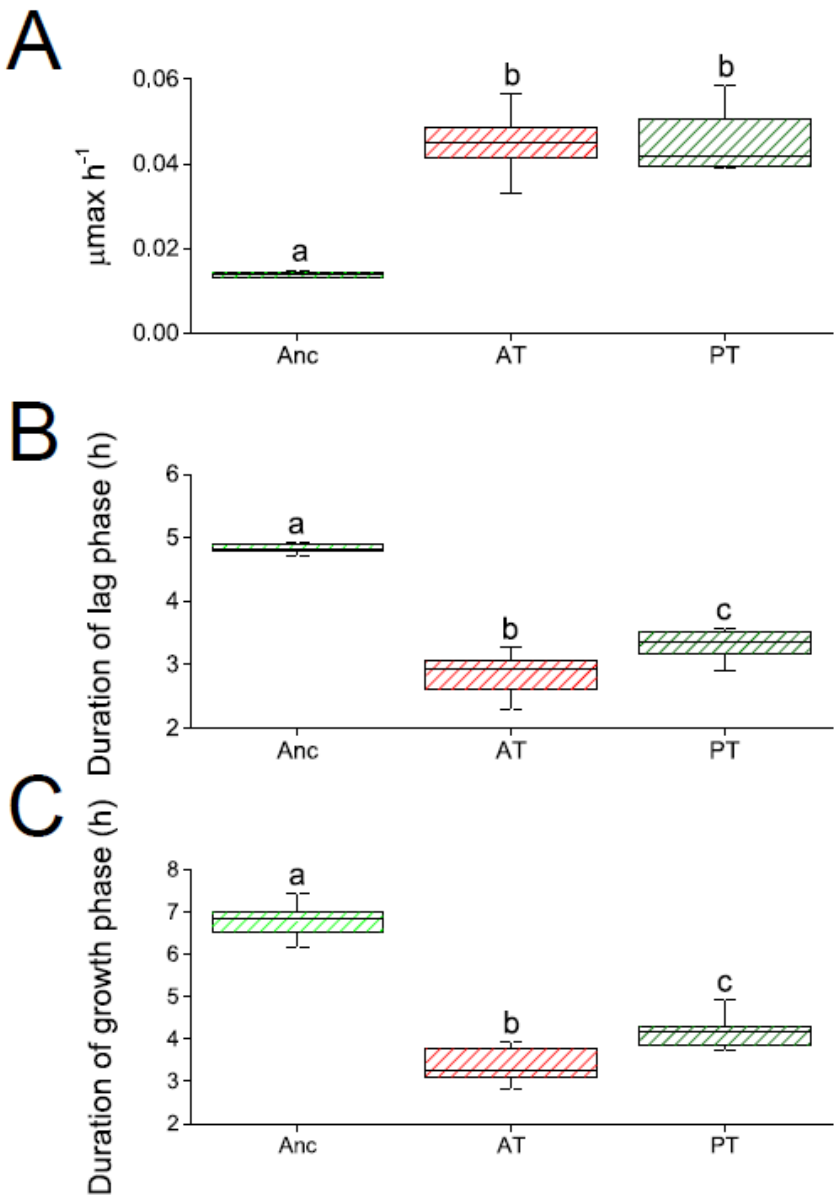
Box plots span the growth rates (μMax h^{-1}) of the replicate populations from the AA (empty boxes) and non-AA (dashed lines) regimes. Asterisks indicate significant differences in fitness ($P<0.05$) as determined by independent sample t-tests. Box plots consist of medians (horizontal lines within boxes), interquartile range (boxes) and 1.5x- interquartile range (whiskers).

Figure S2: Amino acid auxotrophy profiles in the replicate populations over time



The color grid depicts the fraction of the population which was auxotrophic for amino acids. The intensity of the colors reflects the percentage of specific amino acid auxotrophic strains (per 1000 colonies sampled) at different points that were found in (A) populations evolving in the non-AA and (B) AA regime. Black grids indicate completely prototrophic populations whereas red indicates a high frequency of auxotrophic strains (Percent range: 0.1-20%). For each color-grid, a column represents one of the replicate populations (R1 – R8) over different time points (separated by thick gray lines) in the course of the experiment for the two regimes.

Figure S3: Growth kinetic parameters of the ancestral, auxotrophic and prototrophic strains from the AA-regime



Box plots span the (A) µMax h^{-1}, (B) duration of lag phase (h), and (C) duration of the growth rate (h) of the ancestral (Anc), auxotrophic (AT) and prototrophic (PT) strains from the AA-regime. Letters (a,b or c) denote significant differences in the measures of the different strains. (One way ANOVA and LSD post hoc test, $p<0.05$). Box plots

consist of medians (horizontal lines within boxes), interquartile range (boxes) and 1.5x- interquartile range (whiskers).

Supporting Tables:

Table S1: Composition of different dropout media

Medium	AA excluded	AAs present (Concentration = 100µM each)	Function
MMAB -AA	All	None	Non-AA regime, Determination of auxotrophy
MMAB +AA	None	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	AA regime Allows growth of all auxotrophs
Alanine -	Ala	Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Ala
Arginine -	Arg	Ala , Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr. Val	Allows growth of all auxotrophs except those for Arg
Asparigine -	Asn	Ala, Arg, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Asn
Aspartate -	Asp	Ala, Arg, Asn, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Asp
Cysteine -	Cys	Ala, Arg, Asn, Asp, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Cys
Glutamine -	Gln	Ala, Arg, Asn, Asp, Cys, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Gln
Glutamate -	Glu	Ala, Arg, Asn, Asp, Cys, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Glu
Glycine -	Gly	Ala, Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Gly
Histidine -	His	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for His
Isoleucine -	Ile	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Ile
Leucine -	Leu	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Leu

Lysine -	Lys	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Lys
Methionine -	Met	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Phe, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Met
Phenylalanine -	Phe	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Pro, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Phe
Proline -	Pro	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Pro
Serine-	Ser	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Thr, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Ser
Threonine -	Thr	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Trp, Tyr, Val	Allows growth of all auxotrophs except those for Thr
Tryptophan -	Trp	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Tyr, Val	Allows growth of all auxotrophs except those for Trp
Tyrosine -	Tyr	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Val	Allows growth of all auxotrophs except those for Tyr
Valine -	Val	Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr	Allows growth of all auxotrophs except those for Val

Table S2: Genetic regions with mutations in the auxotrophic and prototrophic strains from the two regimes. Single nucleotide polymorphisms (SNPs) in coding or intergenic regions, insertions and deletions arising in the auxotrophic (AT) and prototrophic (PT) strains isolated from the AA or Non-AA regime are listed along with their occurrences in the all the strains sequenced

Genetic region	Mutation type	Regime/s	Strain	Instances
ptsP	SNP,Coding	AA	AT	1/14
yoaA	SNP,Coding	AA	AT	1/14
hemF	SNP,Coding	AA	AT	1/14
dhaM	SNP,Coding	AA	AT	1/14
sspA	SNP,Coding	AA	AT	1/14

stpA	SNP,Coding	AA	AT	1/14
yhdW	SNP,Coding	AA	AT	1/14
rpoB	SNP,Coding	Non-AA	AT	2/14
yqiB	SNP,Coding	Non-AA	AT	2/14
metG	SNP,Coding	Non-AA	AT	1/14
cra /ilvH	SNP, Intergenic	AA	AT	1/14
dxr/frr	SNP, Intergenic	Non-AA	AT	1/14
ynaE/pinR	SNP, Intergenic	Non-AA	AT	1/14
yaeF/nlpE	SNP, Intergenic	Non-AA	AT	1/14
yfjM /yfjL	Insertion	AA	AT	1/14
uspC/flhD	Insertion	AA	AT	1/14
ykfC–proB	Deletion	AA	AT	1/14
insF1–mdtB	Deletion	AA	AT	1/14
wcaN	Deletion	AA	AT	1/14
gltD	SNP,Coding	AA	AT,PT	4/14
crr/ptsI	SNP, Intergenic	AA	AT,PT	2/14
rcnR/thiM	SNP, Intergenic	AA	AT,PT	3/14
rph	Deletion	AA,	AT,PT	4/14
rph/pyrE	Deletion	AA, Non-AA	AT,PT	4/14
wzxC/wcaK	Insertion	AA, Non-AA	AT,PT	2/14
rpsK	SNP,Coding	AA	PT	1/14
cra	SNP,Coding	AA	PT	1/14
fecR	SNP,Coding	AA	PT	1/14
hflC	SNP,Coding	AA	PT	1/14
mppA/pgrR	Insertion	Non-AA	PT	1/14

Table S3: Mutations unique to auxotrophic strains and their predicted functional implications. *Growth: Y, No growth: N ; NA: Not applicable ; Growth data is based on previous studies

Gene	Replicate; regime	Function	Mutation	Predicted effect on protein activity [26]	Growth on minimal medium of deletion genotype	Ref.
ptsP	2,AA	Phosphoenol-pyruvate-protein phosphor-transferase	M604R	Neutral	NA	
yoaA	2,AA	conserved protein	D121Y	Neutral	NA	
hemF	3;AA	coproporphyrinogen III oxidase	G127V	Deleterious	Y	[56]
dhaM	3;AA	dihydroxyacetone kinase subunit M	V196A	Neutral	NA	
sspA	5;AA	stringent starvation protein A	Q24*	Deleterious	N	[29]
stpA	6;AA	H-NS-like DNA-binding protein with RNA chaperone activity	R49S	Neutral	NA	
yhdW	8;AA	putative transport protein	C→A, pseudogene	NA	NA	
rpoB	7,8; Non-AA	RNA polymerase, β subunit	T135P	Deleterious	N	[56]
yqiB	7,8; Non-AA	predicted dehydrogenase	Q117K	Deleterious	Y	
metG	7; Non-AA	methionyl-tRNA synthetase	R600P	Deleterious	N	
cra /ilvH	4;AA	NA	C→T, intergenic	NA	NA	
dxr/frr	8; Non-AA	NA	T→A, intergenic	NA	NA	
ynaE/pinR	8; Non-AA	NA	A→G, intergenic	NA	NA	
yaeF/nlpE	7; Non-AA	NA	A→C, intergenic	NA	NA	
yfjM /yfjL	4;AA	NA	(CATAGTGC)6 →7,intergenic	NA	NA	
uspC/flhD	6;AA	NA	+4 bp, intergenic	NA	NA	
ΔykfC	6;AA	CP4-6 prophage; conserved protein	Deletion	NA	Y	[56]
ΔinsN	6;AA	CP4-6 prophage; partial regulator of insertion element IS911A	Deletion	NA	N	
ΔinsI1	6;AA	transposase of IS30	Deletion	NA	N	
ΔperR	6;AA	PerR transcriptional regulator	Deletion	NA	Y	
ΔykfA	6;AA	predicted GTP-binding protein	Deletion	NA	Y	

ΔyafZ	6;AA	CP4-6 prophage; conserved protein	Deletion	NA	Y
ΔykfK	6;AA	Pseudogene	Deletion	NA	N
ΔykfL	6;AA	Pseudogene	Deletion	NA	N
ΔyafY	6;AA	CP4-6 prophage; inner membrane lipoprotein	Deletion	NA	Y
ΔykfB	6;AA	predicted protein	Deletion	NA	Y
ΔykfF	6;AA	predicted protein	Deletion	NA	Y
ΔyafX	6;AA	predicted protein	Deletion	NA	Y
ΔykfG	6;AA	predicted DNA repair protein	Deletion	NA	Y
ΔykfH	6;AA	predicted protein	Deletion	NA	Y
ΔyafW	6;AA	antitoxin of the YkfI-YafW toxin-antitoxin pair	Deletion	NA	Y
ΔykfI	6;AA	toxin of the YkfI-YafW toxin-antitoxin system	Deletion	NA	Y
ΔykfN	6;AA	Pseudogene	Deletion	NA	Y
ΔthrW	6;AA	tRNA: $tRNA^{thrW}$	Deletion	NA	Y
ΔproA	6;AA	component of glutamate-5-semialdehyde dehydrogenase	Deletion	NA	N
ΔproB	6;AA	component of γ-glutamyl kinase	Deletion	NA	N
ΔinsF1	2,AA	IS3 element protein InsF	Deletion	NA	Y
ΔinsE1	2,AA	IS3 element protein InsE	Deletion	NA	Y
ΔgatR	2,AA	negative DNA-binding transcriptional regulator of galactitol metabolism	Deletion	NA	Y
ΔyegS	2,AA	lipid kinase	Deletion	NA	Y
ΔyegR	2,AA	predicted protein	Deletion	NA	Y
ΔyegZ	2,AA	predicted protein fragment	Deletion	NA	Y
ΔogrK	2,AA	DNA-binding transcriptional regulator, prophage P2 remnant	Deletion	NA	Y
ΔyegQ	2,AA	predicted peptidase	Deletion	NA	Y
ΔyegP	2,AA	predicted protein	Deletion	NA	Y
ΔbaeR	2,AA	BaeR transcriptional regulator	Deletion	NA	Y
ΔbaeS	2,AA	BaeS sensory histidine kinase	Deletion	NA	Y
ΔmdtD	2,AA	putative transport protein, major facilitator superfamily (MFS)	Deletion	NA	Y
ΔmdtC	2,AA	MdtABC-TolC multidrug efflux system - membrane subunit	Deletion	NA	Y

ΔmdtB	2,AA	MdtABC-TolC multidrug efflux system - membrane subunit	Deletion	NA	Y
wcaN	6;AA	predicted uridylyltransferase subunit with GalU	-1 bp, coding	Deleterious	Y

Acknowledgements

A lot of people have contributed in some form or another in the course of this doctoral thesis. Without their help, the last 50 months or so would not have been possible. This is a small way of saying thank you to these wonderful people.

Thank you to my parents for making me understand the importance of the gift of education. Mum, for introducing me to science, especially biology, and teaching me to be brave. Papa, for making everything possible, every single time.

Thank you to Eisha for always being there, for your constant support, for your delicious pastas, for your maggi, and for enduring my philosophical rants.

Thank you to Christian for being a great mentor. Your insight and motivation has been crucial in driving the development of my doctoral thesis. Thank you for always encouraging me to think about "the bigger picture" whether it was experimental replicates, the scale of an experiment or the relevance of our findings. Thanks for expanding my view about science and for converting a microbiologist into a somewhat decent evolutionary microbiologist.

Thank you to Silvio for being a great friend, coauthor, pub crawler, and for always being 'splendid'. Thank you for bringing your tremendous insight into our papers and I am hoping we could write many more.

Thank you to Christoph for being part of my thesis committee, and your contributions and help to make our work manuscripts a great combination of multiple approaches.

Thank you to Prof. Kothe for being part of the thesis committee and for your helpful suggestions throughout the course of the thesis.

Thank you to Grit Winnefeld for always ensuring I could legally stay in Germany and all the administrative work in the institute. Your help was invaluable!

Thank you to the EEEs. Samay (for always helping out and guidance in the lab and out of it), Holger (for picking me up at the station and introducing me to Jena, for the cakes, for all the nasty but helpful questions), Anne-Kathrin (for the 'famous' and highly calorific dark chocolate cake and for being a great desk partner), Daniel (for the dry ice, for letting me say "I hate you", for the hihihi, and for the monster fun), Silvio (for always being 'splendid' and being lenient with the dry ice), Shraddha (for the fun, noise, and constant supply of bakarwadis), Felix (for some of the strains I

worked with), Linda and Tina, (for helping out with many strain constructions and verifications), Lisa (for the fun in the lab), Tobias and Florian (for those random chats in the lab), Ghada (for the 'baspussa' and constant supply of chatter and chips), Samir (for the fun in the lab and the kickass rajma), Muhammad (for the supply of food across the table), Ariane, and Esther (for the fun in the lab). All of you guys have been an incredible set of people to work with. Thank you once again!

Thank you to everyone in the BOL group for being great colleagues. Thank you especially to Angelika Berg for ensuring that tips were never an issue. Thank you to Anja and Andrea for helping out with the ordering several times.

Thank you to Arjan de Visser and all the reviewers of the three manuscripts that this thesis is based on, for their critique and comments which helped to improve them considerably.

Thank you to Kenneth Flynn, Vaughn Cooper, Tim Cooper and Ellie Harrison for their help on the different analyses, statistics and graphing styles used in the various manuscripts in this thesis.

Thank you to various people at MPI-CE. Mathias Stoll and Martin Niebergail for ensuring computer related stuff was always taken care of; Jan Kellman for helping me write the DAAD grant which helped me to attend the very enriching GRC and GRS; Karin for all her help with IMPRS related activities and Linda for always being eager in sourcing books and publications for me.

Thank you to Carsten, Katja and Jan at the JSMC for their help with conferences, grants and other organizational activities.

Thank you to Martin Ackermann, Roberto Kolter and Greg Challis for their helpful advice on life, science, wine and beyond.

Thank you to Srinandan for introducing me to the world of microbial ecology, especially biofilms and communities, and showing how cool they can be.

Thank you to the Indians: Samay, Abith, Tulsi, Latha, Sailendharan, Revathi, Eisha, Shraddha who made life at and after work fun.

Thank you to Sindy, Hella, Katha, Ivanna, Thomas, Govind, Ramses, Anna, Mathis, Daniel Braga, Ana-Carolina, Theresa, Vincent, Ben, and Pam, for all those wonderful evenings over pizza, alcohol, games and movies in Jena.

Thank you to friends back home and elsewhere who have always rooted for me. Ekta, Royston, Allwyn, Mansa, Kashif, Deepali, Rajesh, and Adrian, you been a great support system over the years. Thank you once again!

CURRICULUM VITAE

GLEN GERALD D'SOUZA
Max Planck Institute for Chemical Ecology, Phone: ++49-176-32789723
Hans-Knöll-Strasse 8, Email: glendsouza89@gmail.com
07745, Jena, Germany

CURRENT POSITION

Ph.D. student at the Experimental Evolution and Ecology Group, Max Planck Institute for Chemical Ecology, Jena, Germany (*Since September, 2011*)

EDUCATION

2009 - 2011 | M.SC. MICROBIOLOGY, The Maharaja Sayajirao University of Baroda, Vadodara, India

2006 - 2009 | B.SC. MICROBIOLOGY, Wilson College, The University of Mumbai, India

RESEARCH EXPERIENCE

Ph.D. Dissertation *(2011 – present, anticipated 03/2016)*
Title: **The evolution of metabolic dependency in bacteria**
Advisor: Dr. Christian Kost, Experimental Evolution and Ecology Research Group, Max Planck Institute for Chemical Ecology Jena, Germany

M.Sc. Dissertation *(2009 – 2011)*
Title: **Bacterial biofilm community structure and denitrification activity**
Advisor: Dr. Anuradha Nerurkar, Department of Microbiology and Biotechnology Centre, The Maharaja Sayajirao University of Baroda, Vadodara, India

PUBLICATIONS

In review

1. Waschina S, D'Souza G, Kaleta C, and Kost C
Metabolic network architecture and carbon source determine metabolite production costs. *The FEBS Journal*

Peer reviewed and published:

1. D'Souza G*, Waschina S*, Kaleta C, and Kost C (2015)
Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes. *Evolution* 69 (5):1244–1254; doi: 10.1111/evo.12640
** indicates equal contributions as first authors*

2. D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C and Kost C (2014)
Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution* 68 (9): 2559-2570; doi: 10.1111/evo.12468

3. Srinandan C, D'Souza G, Srivastava N, Nayak B, and Nerurkar A (2012)
Carbon sources influence the nitrate removal activity, community structure and biofilm architecture. *Bioresource Technology* 117:292-299; doi:10.1016/j.biortech.2012.04.079

COMPLEMENTARY EDUCATION

2014 | Statistics and statistical analysis with R
2014 | Scientific research funding
2014 | Ph.D. Summer School: Methods for Mathematical and Empirical Analysis of Microbial Communities
2013 | CUSO workshop: Bridging Theoretical and Experimental Evolution

RESEARCH PRESENTATIONS

Invited presentations

1. **Strong selection for loss of metabolic autonomy in bacteria.** (Molecular Microbial Ecology Group, EAWAG/ETH-Zurich, Zurich, Switzerland May 2015)

Oral presentations

1. **Experimental evolution of metabolic dependency in bacteria.** (15th Meeting of the European Society of Evolutionary Biology: Symposium on The evolution and ecology of trait loss and dependency, Lausanne, Switzerland, August 2015)
2. **Experimental evolution of metabolic dependency in bacteria.** (Gordon Research Seminar on Microbial Population Biology: Ecology, Evolution and Applications in Microbial Population Biology, Andover, USA, July 2015)
3. **Strong selection for loss of metabolic autonomy in bacteria.** (MiCom 2015, 5th International Students Conference on Microbial Communication, Jena, Germany, April 2015)
4. **Less is more: Selective advantages can explain the loss of biosynthetic functions in bacteria** (PhD Summer School: Methods for Mathematical and Empirical Analysis of Microbial Communities, Newton Institute, Cambridge, UK, October 2014)
5. **Lose to gain: Adaptive fitness benefits drive prevalent loss of biosynthetic genes in bacteria** (VW Status Symposium on Evolutionary Biology, Hannover, Germany, October 2013)
6. **Lose to gain: Adaptive fitness benefits drive prevalent loss of biosynthetic genes in bacteria** (Jena Center for Bioinformatics Workshop: Bioinformatics meets biodiversity, Jena, Germany, September 2013)
7. **Lose to gain: Adaptive fitness benefits drive prevalent loss of biosynthetic genes in bacteria** (MPI-CE Institute Symposium, Jena, Germany, September 2013)
8. **Lose to gain: Adaptive fitness benefits drive prevalent loss of biosynthetic genes in bacteria** (CUSO workshop: Bridging Theoretical and Experimental Evolution, La Fouly, Switzerland, June 2013)
9. **Lose to gain: Adaptive fitness benefits drive prevalent loss of biosynthetic genes in bacteria** (12th IMPRS-CE Symposium, Jena, Germany, April 2013)

10. **You lose some, you gain some: Trait loss confers a fitness advantage to bacteria** (MiCom 2012, 3rd International Students Conference on Microbial Communication, Jena, Germany, November 2012)

Poster presentations

1. **Experimental evolution of metabolic dependency in bacteria.** (Gordon Research Conference on Microbial Population Biology: Disentangling Causes and Consequences, Andover, USA, July 2015)
2. **Lose to gain: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria** (566th WE-Heraeus-Seminar: Mechanisms, Strategies, and Evolution of Microbial Systems, Bad Honnef, Germany, June 2014)
3. **Towards a mechanistic understanding of the selective benefit of auxotrophy in Escherichia coli** (4th International Student Conference on Microbial Communication – MiCom 2014, Jena, Germany, March 2014)
4. **The benefits of being auxotrophic: Can fitness advantages account for the loss of biosynthetic functions in bacteria?** (11th IMPRS-CE symposium Dornburg, Germany, February 2012)
5. **Influence of different carbon sources and high nitrate on biofilm community structure in denitrifying conditions** (MiCom 2011: 2nd International Student Conference on Microbial Communication, Jena, Germany, September 2011)
6. **Influence of different carbon sources and high nitrate on biofilm community structure in denitrifying conditions** (51st Annual Conference of Association of Microbiologists of India, Ranchi, India, November 2010)

GRANTS, AWARDS AND ACHIEVEMENTS

2015 | Travel grant from the International Max Planck Research School to attend the 15th Meeting of the European Society of Evolutionary Biology, Lausanne, Switzerland
(Amount: €200)
2015 | Grant towards registration fees from the conference chair to attend the Gordon Research Seminar on Microbial Population Biology 2015, Andover, USA (Amount: $240)
2015 | Travel grant from the German Academic Exchange Service (DAAD) to attend the Gordon Research Conference on Microbial Population Biology 2015, Andover, USA (Amount: €1948) **2015 |** Travel grant from the Jena School for Microbial Communication (JSMC) to attend the Gordon Research Seminar on Microbial Population Biology 2015, Andover, USA
(Amount: €250)
2014 | Newton Institute Grant for travel and registration to attend the PhD Summer School: Methods for Mathematical and Empirical Analysis of Microbial Communities, University of Cambridge (Amount: £420)
2012 | The Mr. A.N. Kacchy Memorial Gold Medal in Microbiology, Faculty of Science, The Maharaja Sayajirao University of Baroda *(The highest score achieved in the academic year)*

2012 | Mrs. Jayalaxmi Patwa Post Graduate Scholarship, Department of Microbiology and Biotechnology Centre, The Maharaja Sayajirao University of Baroda *(The highest score achieved in the academic year)*
2012 | Prof. S.D. Sabnis Post Graduate Scholarship, The Maharaja Sayajirao University of Baroda *(The highest score achieved in the academic year)*
2011 | Junior Research Fellowship (All India rank: 76 of ~10,000 applicants), Council of Scientific and Industrial Research, Government of India *(Not used)*
2010 | Junior Research Fellowship (All India rank: 265 of ~10,000 applicants), University Grants Commission, Government of India *(Not used)*
2010 | Dr. M.S. Patel Post Graduate Scholarship, The Maharaja Sayajirao University of Baroda *(The highest score achieved in the academic year)*
2010 | Shri. Kamal Roy Bhatt Memorial Prize for the highest score in M.Sc. Microbiology, Department of Microbiology and Biotechnology Centre, The Maharaja Sayajirao University of Baroda *(The highest score achieved in the academic year)*
2010 | Shri. Vitthal Bhika Patel Prize for the highest score in the academic year in M.Sc. Microbiology, Department of Microbiology and Biotechnology Centre, The Maharaja Sayajirao University of Baroda
2010 | Best Poster Award in Environmental Microbiology at the 51st Annual Conference of the Association of Microbiologists of India
2009 | The Vikrant Khanvilkar Memorial Prize for the highest score in B.Sc. Microbiology, Wilson College

SOCIETAL MEMBERSHIPS

1. European Society for Evolutionary Biology
2. Society of Biological Chemists of India, Baroda Chapter

ACTIVITIES AND SCIENTIFIC OUTREACH

1. Responsible for organization of the scientific program of the 4th International Student Conference on Microbial Communication – MiCom 2014 held in Jena in March 2014
2. Part of the presentation “Cooperation or conflict: Insights into the social lives of bacteria” for the general public at the 5th Long Night of Sciences in Jena (Attended by approximately 10,000 people)
3. Responsible for financial, logistical and scientific organization of ADNIKA: A nature conservation exposition held in Mumbai in January 2007 and September 2008 at Wilson College, Mumbai. (Attended by approximately 10,000 people)

Declaration of Independent Assignment

I declare in accordance with the conferral of the degree of doctor from the Faculty of Biology and Pharmacy of the Friedrich-Schiller-University of Jena that the submitted thesis was written only with the assistance of the literature cited in the text.

People who assisted in experiments, data analysis and writing of the manuscripts are listed as co-authors of the respective manuscripts.

I was not assisted by a consultant for doctorate theses.

The thesis has not been previously submitted elsewhere, whether to the Friedrich-Schiller-University of Jena or to any other university.

Jena, 14th December 2015

Glen Gerald D'Souza