



HAL
open science

Ecological and molecular bases of adaptive diversification in *Escherichia coli*

Jessika Consuegra Bonilla

► **To cite this version:**

Jessika Consuegra Bonilla. Ecological and molecular bases of adaptive diversification in *Escherichia coli*. Microbiology and Parasitology. Université Grenoble Alpes, 2016. English. NNT: 2016GREAV090 . tel-01626007

HAL Id: tel-01626007

<https://theses.hal.science/tel-01626007>

Submitted on 30 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Virologie - Microbiologie - Immunologie**

Arrêté ministériel : 25 mai 2016

Présentée par

Jessika CONSUEGRA BONILLA

Thèse dirigée par **Joël GAFFE** et
codirigée par **Dominique SCHNEIDER**

préparée au sein du **Laboratoire Techniques de l'Ingénierie
Médicale et de la Complexité - Informatique, Mathématiques et
Applications**
dans l'**École Doctorale Chimie et Sciences du Vivant**

BASES ECOLOGIQUES ET MOLECULAIRES DE LA DIVERSIFICATION ADAPTATIVE CHEZ *ESCHERICHIA COLI*

Thèse soutenue publiquement le « **13/12/2016** », devant le jury
composé de :

Mme Laurence DESPRÉS

PU, Grenoble – France. Président.

Mr Ivan MATIC

DR, Paris – France. Rapporteur.

Mr François LEULIER

CR, Lyon – France. Rapporteur.

Mr Daniel ROZEN

PU, Leiden – Pays-Bas. Examineur.

Mr Joël GAFFE

MCU, Grenoble – France. Directeur de thèse.

Mr Dominique SCHNEIDER

PU, Grenoble – France. Co-directeur de thèse.



*“Pongo estos seis versos en mi botella al mar
con el secreto designio de que algún día
llegue a una playa casi desierta
y un niño la encuentre y la destape
y en lugar de versos extraiga piedritas
y socorros y alertas y caracoles.”*

Mario Benedetti

To Mariela & Sophie...

Acknowledgements

I would first like to thank the French Ministry of Higher Education and Research for the attribution of the PhD grant that allowed me to pursue my doctorate.

I also wish to thank Ivan Matic, François Leulier, Daniel Rozen and Laurence Després for accepting to be part of my defense committee.

I would like to specially thank Dominique Schneider, first, for opening me the doors of the lab without even know me, and also for your trust, constant support and guidance during all these years. Your passion for research is of great inspiration.

Many thanks to Joël Gaffé, your kindness made much easier all these years. Thanks for hearing all my complains, for all the encouraging and nice words, for your trust. Thanks for your supervision, for always encourage me to see every result from a different point of view.

Thanks to Thomas Hindré, I think you become, (without knowing) my *ad honorem* supervisor. Thanks for all the scientific discussions, your guidance, support and your valuable comments on my manuscript.

Orkun S. Soyer, Tobias Großkopf, Guillaume Beslon, Daniel Rozen and John C. Willison thanks for all the discussions. Your insights on ecology, system biology, bacteria metabolism and *in-silico* evolution, as well as our collaboration valuably enriched my doctorate.

Thanks to JessiCa Plucain for introduce me with the Ara-2 population. Without your help, guidance and tricks it would not have been possible to pursuit the challenge of working with the S and L lineages.

Thanks to Amandine, not only for all your technical support in the lab during the days you worked with us but also for your friendship. It made much easier my new life in France.

The preparation of this doctorate also includes sharing great moments with lots of people in the lab. Thanks to the members of the old LAPM, specially to the people of our team, GEM. Thanks for all the discussions (scientific or not), the laughs, the “BANG” sessions, the cheese days, the “On descend”... the list is too long to name you all, you know who you are, you are in my heart.

Thanks to my family and friends in Colombia. Knowing that you are there for me despite the distance, that you trust me and that you are proud of me no matter what encourages me to go further.

Most importantly, thanks to Mehdi, my loving husband. My life would not be the same without you. You are an unending inspiration.

Abstract

Diversification events are central issues in evolution since they generate phenotypic innovation such as colonization of novel ecological niches and, ultimately, speciation. To study the ecological and molecular drivers of adaptive diversification, we used the longest still-running evolution experiment. Twelve independent populations are propagated in a glucose limited minimal medium from a common ancestor of *Escherichia coli* by serial daily transfers since 1988 for more than 60,000 generations. In one of the twelve populations, called Ara-2, a unique diversification event occurred: two phenotypically-differentiated lineages, named S (Small) and L (Large) according to their cell size, emerged from a common ancestor at ~ 6500 generations. The two lineages co-exist ever since, owing to negative frequency-dependent selection whereby each lineage is favored and invades the other when rare, such that no lineage gets extinct. Moreover, and before the split between the two S and L lineages, the population Ara-2 evolved a hypermutator phenotype, owing to a defect in a DNA repair gene. The objective of this thesis is to characterize the ecological, physiological and molecular mechanisms that allowed the emergence and stable co-existence of the S and L lineages.

First, we used a combination of *in vivo* and *in silico* experimental evolution to determine the ecological and physiological drivers of the emergence of the polymorphism. Several ecological mechanisms including tradeoff, seasonality and character displacement are involved in the emergence and long-term persistence of diversity. In particular, we showed that the L lineage secretes acetate which generates a new ecological opportunity that the S lineage exploited. In addition, the S and L lineages became fitter and fitter over time in their respective ecological niches, respectively acetate and glucose. Second, we propagated S and L clones separately to remove competition between the two lineages. In these conditions, frequency-dependent interactions between the S and L clones evolved separately were completely abolished, revealing the importance of competition in the maintenance of the polymorphism.

Third, we combined genetic, physiological and biochemical approaches to determine the role of an S-specific mutation that was previously found in *arcA*, encoding a global regulator, in the emergence of the S and L polymorphism. We showed that the evolved *arcA* allele conferred to the S lineage the capacity to grow on acetate by increasing the transcription of target genes involved in acetate consumption. During this study, we found an additional mutation, in the *acs* gene involved in acetate metabolism, that was also involved in the emergence of the S lineage. We further showed that these two mutations were favorable to the S lineage early during its emergence, and that other mutations occurred later that interacted epistatically with the *acs* and *arcA* evolved alleles. Therefore, these data showed that the establishment and further maintenance of the S and L polymorphism was a multi-step process involving epistatic interactions between several mutations. Fourth, we identified the long-term dynamics of mutation rates in this divergent population. A first early rise of a hypermutator was followed by a full reversion of this mutator state twice independently in each of the two S and L lineages.

The emergence of a long-term bacterial polymorphism reflects a complex restructuring of the metabolic and regulatory networks in the co-existing lineages, resulting in the generation and exploitation of a new ecological opportunity. Competition and evolution of divergent resource consumption were the selective forces driving the maintenance of the polymorphism.

Résumé

Les événements de diversification adaptative sont des éléments primordiaux de l'évolution. En effet, ils engendrent des innovations phénotypiques telles que la colonisation de nouvelles niches écologiques et au final, la spéciation. Afin d'étudier les ressorts écologiques et moléculaires de la diversification adaptative, nous utilisons la plus longue des expériences d'évolution en cours. Depuis 1988, soit plus de 60 000 générations, douze populations indépendantes issues d'un ancêtre commun d'*Escherichia coli* sont propagées quotidiennement dans un milieu minimum comportant une faible quantité de glucose. Un événement unique de diversification s'est produit dans une des 12 populations (Ara-2). Deux lignées de phénotypes différents sont apparues après environ 6500 générations, les S pour «Small» et les L pour «Large», chacune présentant des tailles cellulaires différentes. Les deux lignées coexistent grâce à une sélection négative dépendant de la fréquence qui favorise la lignée la plus rare et permet de supplanter sa concurrente; ainsi, aucune des deux lignées ne s'éteint. Avant l'événement de diversification, la population Ara-2 a développé un phénotype hypermutateur suite à la mutation d'un gène de réparation de l'ADN. L'objectif de cette thèse est de caractériser les mécanismes écologiques, physiologiques et moléculaires sous-tendant l'émergence et la coexistence des lignées S et L.

En premier lieu, nous avons utilisé un ensemble d'expériences d'évolution *in vivo* et *in silico* afin de déterminer les moteurs écologiques et physiologiques de l'émergence de ce polymorphisme. Plusieurs mécanismes écologiques, incluant les compromis (trade-off évolutifs), la saisonnabilité et les déplacements de caractères interviennent dans l'émergence et la persistance de la diversité au long terme. Nous avons montré que la lignée L, en produisant de l'acétate, créait une nouvelle opportunité écologique exploitée par les S. De plus, au cours du temps, les S et les L s'adaptent à leur niche écologique, respectivement l'acétate et le glucose.

En second lieu, nous avons cultivé les S et les L séparément pour éliminer la compétition entre les deux lignées. Dans ces conditions, il y a perte des interactions dépendantes de la fréquence entre les S et les L. Ceci démontre l'importance de la compétition dans le maintien du polymorphisme.

En troisième lieu, nous avons combiné des approches génétiques, physiologiques et biochimiques pour déterminer le rôle, dans l'émergence du polymorphisme, d'une mutation spécifique aux S survenant dans le gène *arcA*, codant un régulateur global. Nous avons montré que l'allèle évolué de *arcA* augmentait la transcription de gènes du métabolisme de l'acétate dans la lignée S. Au cours de cette étude, nous avons identifié une mutation supplémentaire dans le gène *acs*, impliqué dans le métabolisme de l'acétate, intervenant dans l'émergence de la lignée S. Nous avons aussi démontré que ces deux mutations étaient favorables à la lignée S au début de son émergence, puis que des mutations plus tardives agissaient de façon épistatiques avec les allèles évolués de *acs* et de *arcA*. Ainsi, ces résultats démontrent que l'établissement et le maintien du polymorphisme des S et des L est un processus en plusieurs étapes nécessitant des interactions épistatiques entre plusieurs mutations.

En quatrième lieu, nous avons identifié la dynamique au long terme des taux de mutations dans cette population. L'apparition et l'invasion rapide du phénotype hypermutateur est suivie d'une réversion complète mais indépendante dans chacune des lignées S et L indépendamment.

L'émergence d'un polymorphisme bactérien durable reflète une restructuration complexe des réseaux métaboliques et de régulation dans ces lignées qui co-existent, ce qui aboutit à l'apparition et à l'exploitation de nouvelles opportunités écologiques. La compétition et l'évolution de l'utilisation de ressources différentes sont des forces sélectives permettant le maintien du polymorphisme.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Résumé	vii
List of figures	1
PART I: INTRODUCTION	3
I. The Diversity of the Bacterial World	4
<i>i. The ecological theory of adaptive diversification</i>	5
<i>ii. Bacterial diversity in natural environments</i>	12
<i>iii. Bacterial diversity in clinical settings</i>	15
<i>iv. Bacterial diversity in experimental settings</i>	22
II. The Long-Term Evolution Experiment with <i>Escherichia coli</i>	31
<i>i. Experimental setting</i>	32
<i>ii. Parallel evolution</i>	34
<i>iii. Emergence of a stable long-term polymorphism in the LTEE</i>	38
PART II: RESULTS	50
Chapter 1: Metabolic Modelling in a Dynamic Evolutionary Framework Predicts Adaptive Diversification of Bacteria in a Long-Term Evolution Experiment	51
Chapter 2: Phenotypic Causes of Fast Loss of Long-Term <i>Escherichia coli</i> Ecotypic Coexistence After Decreased Competition	74
Chapter 3: Molecular Genetics of a New Ecological Opportunity Exploitation During Long-Term Bacterial Sympatric Adaptive Diversification	95
Chapter 4: Long-Term Dynamics of Mutation Rates in the Ara-2 Population	131

4.1. <i>Introduction</i>	134
4.2. <i>Materials and Methods</i>	139
4.3. <i>Results</i>	141
4.4. <i>Discussion</i>	147
Chapter 5: Role of Environmental Seasonality on Sympatric Bacterial Adaptive Diversification	150
PART III: GENERAL DISCUSSION	187
REFERENCES	199

List of figures

Figure 1. Cohan’s Ecological species concept	6
Figure 2. Bacterial characteristics affected by trade-offs	8
Figure 3. Sympatric divergence of ecotypes by resource competition	10
Figure 4. Estimated number of bacterial species	12
Figure 5. Representations of the bacterial diversity in freshwater	13
Figure 6. Phenotypic diversity observed in <i>E. coli</i> isolates	16
Figure 7. Isolate sampling points and patient life span	17
Figure 8. Parallel evolution of <i>P. aeruginosa</i>	18
Figure 9. Bacterial phylogeny	20
Figure 10. Phenotypic diversity and niche specificity	23
Figure 11. Growth profiles of each isolated strain	25
Figure 12. TCA, glyoxylate shunt, and acetate consumption/excretion metabolic pathways	27
Figure 13. Chemostat design and establishment of a steady state	28
Figure 14. Structure of <i>acs</i> from ancestral and acetate-scavenging strains of <i>E. coli</i>	30
Figure 15. Representation of the LTEE with <i>E. coli</i> .	33
Figure 16. Fitness trajectories of the 12 LTEE populations	35
Figure 17. Relative growth rates across a variety of growth substrates for evolved strains	36
Figure 18. Emergence of the Cit ⁺ phenotype in population Ara-3	40
Figure 19. Phylogeny of Ara-3 population	41
Figure 20. Tandem amplification in Cit ⁺ genomes	42
Figure 21. Phenotypes of the L and S lineages from population Ara-2.	44
Figure 22. Dynamics of the S and L lineages	45

Figure 23. Evolutionary dynamics of global transcription profiles for L and S lineages during 40,000 generations	46
Figure 24. Phenotypic traits conferred by the three mutations in <i>spoT</i> , <i>arcA</i> and <i>gntR</i>	48
Figure 25. Number of mutations in the Ara-2 populations over evolutionary time	142
Figure 26. Drop test of mixed Ara-2 populations sampled over evolutionary time	143
Figure 27. Drop test of isolated Ara-2 S and L clones sampled over evolutionary time	144
Figure 28. <i>In vivo</i> quantification by fluctuation tests of mutation rates in isolated clones from the S and L lineages of population Ara-2	145

PART I: INTRODUCTION

I. The Diversity of the Bacterial World

i. The ecological theory of adaptive diversification

Speciation

All living organisms have diverged from common ancestors through diversification events into divergent lineages that subsequently adapted to their environments, thereby generating the fantastic phenotypic diversity on Earth. The mechanisms leading to divergence and split into lineages often involve unique events that occurred in a distant past. The study of the origins of diversity and complexity of the living world is frequently based on direct comparisons and phylogenetic analysis of contemporary species or by taking advantage of the available fossil records that allow researchers to access to the morphological characteristics of extinct organisms. However, the fossil records are only partial with no access to the ancestral organisms.

Before the genomic era, these direct comparisons allowed researchers to “classify” the living world according to organismal similarities. With the advent (and availability) of genome sequencing technologies, this classification has been refined providing us with a clearer picture of the relationships between organisms and of their evolutionary proximity.

The first level of classification, the species, is most commonly defined by the “Biological Species Concept”. Promoted by (Mayr 1963), this genetic definition conceives a species as “*a group of interbreeding individuals that is isolated from other such groups by barriers to recombination*”. If genetic exchange, by means of sexual reproduction, is sufficiently extensive within a species, and low across species, then single species will be relatively homogeneous and ecologically distinct from other species (Prosser et al. 2007). This biological definition of species, as reproductive groups, may easily be understandable for most eukaryotic living forms on Earth, but not for asexual organisms. Prokaryotes (and some eukaryotes) are asexual, thereby violating these assumptions.

Alternatively, (Cohan 2002) defines bacterial species by their ecological properties. The “Ecological Species Concept” (Figure 1) considers a species as a set of individuals that are similar in relevant ecological properties. According to Cohan, the species concept in bacteria should rely on the notion of “*ecotype*” which itself relies on the ecological and evolutionary dynamics of subpopulations. Two bacterial subpopulations may be considered as different ecotypes if: (i) they form monophyletic clusters, (ii) they occupy different ecological niches and (iii) periodic selection purges diversity within each niche without preventing divergence between the inhabitants of different niches. Thereby, species are here defined by their level of evolutionary isolation.

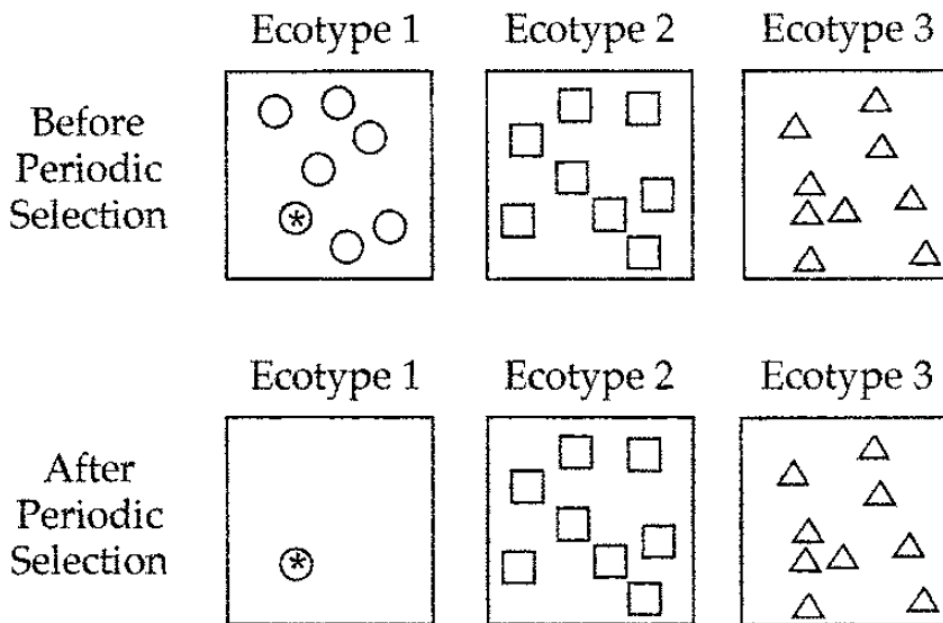


Figure 1. Cohan’s Ecological species concept, after (Cohan 2001).

Periodic selection events purge within a single ecotypic group without affecting other ecotypes, illustrating the transient nature of diversity within ecotypes and the permanence of divergence among ecotypes. Each symbol represents an individual organism and the distance between symbols the sequence divergence between organisms. The asterisk indicates an adaptive mutation in one individual of Ecotype 1. Because recombination is rare, the adaptive mutant and its clonal descendants eliminate the genetic diversity within the ecotype. Because ecotypes differ in the resources they use, the adaptive mutant from one ecotype does not outcompete the cells of other ecotypes, their genetic diversity being unaltered. Once populations of cells are divergent enough to escape one another’s periodic selection events (i.e., populations are separate ecotypes), those populations are free to diverge permanently.

Microbial evolution can occur rapidly, particularly under strong selective pressures. One intrinsic fate of evolution is speciation, the first step of which potentially being an event of adaptive radiation. The ecological theory of adaptive radiation (Schluter 2000b) describes the adaptive process in three steps:

- a. First, an environmental variation causes divergent natural selection, leading to the evolution of niche specialists with fitness tradeoff across alternative environments.
- b. Second, the rate and extent of ecological and phenotypic divergence increase, involving biotic interactions such as resource competition or predation that promote the evolution of niche specialization.
- c. Third, speciation itself. For sexual organisms, it implies evolution of reproductive isolation. This last step can be ignored in asexual organisms.

Emergence of an ecological opportunity, also referred as a new “*Adaptive Zone*”, triggers adaptive radiation. A given organism/ecotype can enter these adaptive zones by the evolution of a key innovation, dispersal into a new habitat/niche, or the extinction of antagonists/competitors (Simpson 1953). When a new ecological opportunity appears, relaxed selection on one or more ecological traits is expected, thereby enlarging the adaptive landscape and the range of viable phenotypes (Lister 1976, Yoder et al. 2010). However, ecological opportunity may sometimes not produce diversity. Indeed, some lineages may respond to ecological opportunity by becoming a single broadly adapted generalist (Yoder et al. 2010, Levins 1968).

Emergence of novel niche specialists is a complex process. The invasion of a new environment requires the survival of the population, which involves the survival and dispersal of an initially rare and not well adapted individual. The survival of the population requires the

availability of genetic variation upon which selection can act to generate adaptive divergence. A small initial population size means that standing genetic variation is likely to be low, and consequently adaptation must rely on novel genetic variants produced through mutations. Finally, the evolution of fitness tradeoffs is required for divergent selection to generate and maintain niche specialization (Kassen 2009).

Tradeoffs

Tradeoffs in fitness across environments evolve through divergent selection because specialization results in a cost of adaptation. The evolved ecotype has lower fitness in its original environment since alleles that are beneficial in one environment may be deleterious in others (Kassen 2009). The existence of such tradeoffs prevents competitive exclusion by one type, thereby potentially promoting coexistence of organisms (Rainey et al. 2000).

Each of the important properties of bacteria is constrained by tradeoffs, and thus all of them shape bacterial diversity (Figure 2). The impact of tradeoffs in bacterial properties may be classified into three processes: (i) resource allocation, (ii) design constraint, and (iii) informational processing, that are reviewed in (Ferenci 2016).

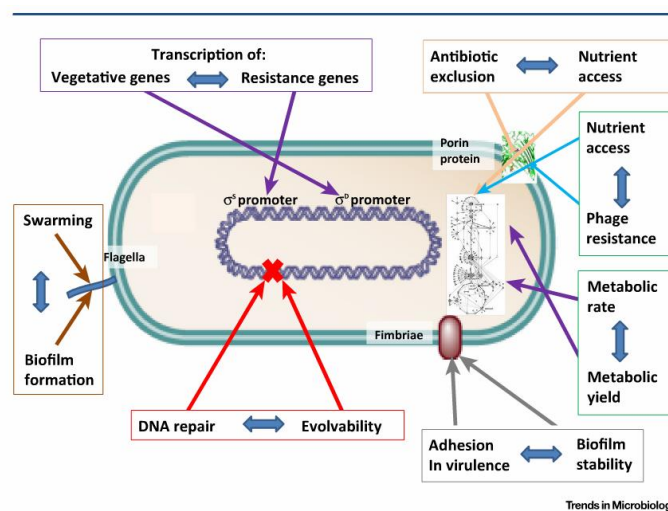


Figure 2. Bacterial characteristics affected by trade-offs, after (Ferenci 2016).

The diagram shows pairs of properties linked by double-headed arrows for which an increase in one property is accompanied by reduction in another, which is defined as tradeoffs.

Resource allocation is determined by whether bacterial cells commit RNA polymerase to make one set of proteins or another. One example involves the sigma factor RpoS that shifts the transcription machinery from vegetative genes (RpoD-dependent) to stress-response genes (King et al. 2004). Resource allocation tradeoffs often arise when global transcriptional patterns shift toward either of two alternative end-points (Ferenci 2016). Indeed, phenotypic innovations are more likely to emerge through changes in regulatory rather than structural genes. Mutations may more frequently generate variation in the quantity of enzymes, rather than in their specificity (Stern 2000). At the metabolic level, resource allocation can also affect the ability of bacteria to scavenge nutrients. Expensive transport systems to scavenge nutrients reduce growth yields of bacterial cells but provide an ecological advantage in nutrient-poor settings (Ferenci 1996).

Design constraint tradeoffs are generally associated with biological processes involving enzymes or proteins. Enzymes may exhibit a tradeoff between accuracy and speed and proteins a tradeoff between stability and activity (Tawfik 2014). An example is phosphomycine resistance of enterohemorrhagic *Escherichia coli* (EHEC) O157:H7. The CpxAR two-component system confers phosphomycine resistance through a *cpxA* mutant lacking its phosphatase activity which leads to the constitutive activation of the response regulator CpxR. This structural change in CpxA produces an impaired growth phenotype owing to reduced carbon substrate uptake (Kurabayashi et al. 2014).

Informational processing tradeoffs are related to the energetic cost of storing, transferring and processing information. Maintenance of large genomes in bacteria is energetically expensive and thus, genome reduction may be selectively advantageous. However, the loss of information can be detrimental if bacteria are exposed to rapid environmental changes (D'Souza et al. 2014). Moreover, the maintenance of genetic integrity through accuracy in processes such as DNA replication and repair imposes a tradeoff between evolvability and genome stability.

High fidelity reduces the access to new mutations and thus the diversity of the population. High mutation rates, on the other hand, reduces genome stability and may increase organismal sensitivity to stresses owing to the accumulation of deleterious mutations (Wielgoss et al. 2013).

Niche specialization and character displacement

Niche specialization involves biotic interactions such as resource competition which is a hallmark in several cases of adaptive radiation (Schluter 2000a). Resource competition increases the rate and extent of ecological and phenotypic divergence. If resources are limiting, individuals may compete strongly. If one of the competitors has a selective advantage over the other, it will propagate and drive the other to extinction (competitive exclusion). Yet, an alternative possibility is that natural selection favors, in each population, those individuals whose phenotype allows them to use resources not used by the others, resulting in divergence in phenotype and resource use that allows coexistence (Figure 3) (Losos 2000). The process of divergence in phenotype and resource use in sympatric species (without geographical isolation) is frequently referred as “*Ecological Character Displacement*”.

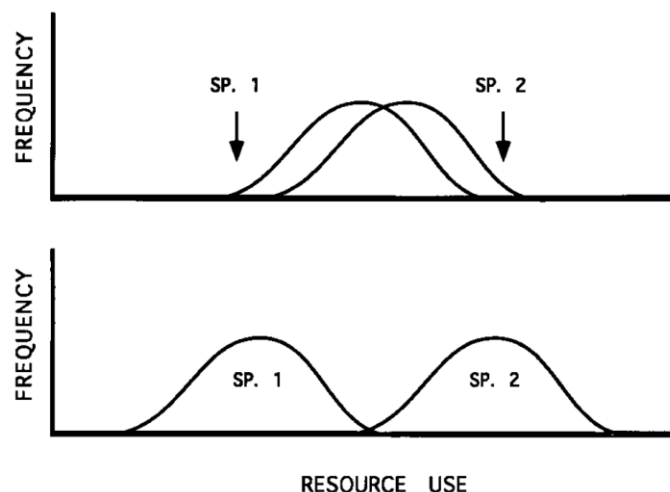


Figure 3. Sympatric divergence of ecotypes by resource competition, after (Losos 2000).

(Upper) Two sympatric species (SP. 1 and SP. 2) with broadly overlapping distributions of resource use. If resources are limiting, then natural selection may favor individuals in each species with traits that allow each of them to use that portion of the resource spectrum not used by the other species, as indicated by the arrows. (Lower) The result may be that the species diverge in trait value and resource use, thus minimizing competition for resources.

The idea of character displacement as a source of ecological diversification was first discussed by (Brown and Wilson 1956). In this seminal paper, they described it as “[...] *Two closely related species have overlapping ranges. In the parts of the ranges where one species occurs alone, the populations of that species are similar to the other species and may even be very difficult to distinguish from it. In the area of overlap, where the two species occur together, the populations are more divergent and easily distinguished, i.e., they “displace” one another in one or more characters. The characters involved can be morphological, ecological, behavioral, or physiological; they are assumed to be genetically based [...].*” In other words, the process of ecological character displacement produces exaggerated divergence in sympatry but similarities in resource use and phenotype when lineages evolved in allopatry (Schluter 2000a).

Ecological character displacement has been suggested to be a key driver of evolutionary diversification and adaptive radiation. It might be initiated when (i) interspecific competition for limited resources creates natural selection that favors those individuals most adept at partitioning resources, which (ii) drives populations to diverge adaptively, either by changing trait means or shrinking trait variance (Stuart and Losos 2013).

ii. Bacterial diversity in natural environments

Microorganisms comprise the vast majority of species on Earth. Studying their diversity in detail is nowadays possible owing to the advance and availability of molecular methods. Indeed, bacterial diversity is traditionally analyzed by sequencing the gene encoding the 16S ribosomal RNA. The 16S rRNA gene contains variable regions that allow accurate taxonomic and phylogenetic identification of organisms. However, these phylogenetic surveys are limited since they are based only on the diversity of one gene. To overcome this, during the recent years bacterial diversity is studied using metagenomics that allows the direct genetic analysis of total DNA contained within an environmental sample. Metagenomics applies genomic technologies and bioinformatics tools to directly access to the functional gene composition of entire communities of organisms (Thomas, Gilbert, and Meyer 2012).

Almost all environments contain hundreds to thousands of bacterial taxa, and bacterial diversity by far exceeds the diversity observed for plants and animals. A meta-analysis of bacterial diversity from a variety of ecosystems performed by (Fierer and Lennon 2011) determined that the richness of bacterial communities is highly variable across environments. For example, the atmosphere, glacial ice, and highly acidic stream waters harbor relatively low numbers of bacterial species, while soils, microbial mats, and marine water are likely to have thousands of bacterial species (Figure 4).

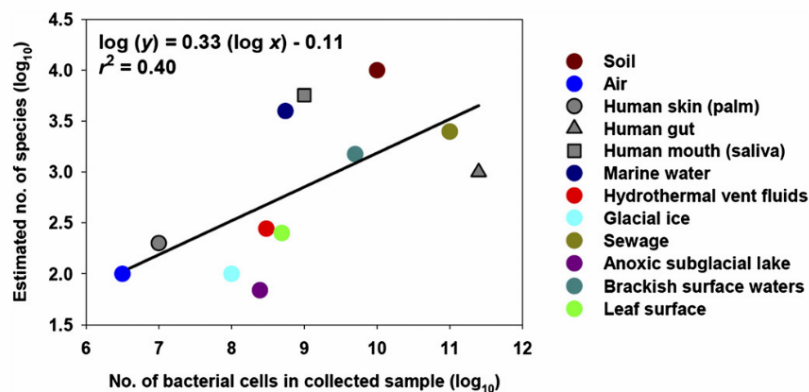


Figure 4. Estimated number of bacterial species (based on 16S rRNA similarity) versus the estimated number of cells in the collected sample, after (Fierer and Lennon 2011).

Freshwater habitats, including aquifers, groundwater, lakes, rivers, drinking water and waste-water, are the natural habitats that harbor the largest number and most diverse groups of bacterial lineages including > 1600 Operational Taxonomic Units (OTUs) (Tamames et al. 2010). The predominant bacteria belong to the phyla *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* and *Firmicutes*, irrespective of the type of water. However, different types of water present distinct patterns of bacterial diversity at lower taxonomic ranks, including genus or species (Figure 5).

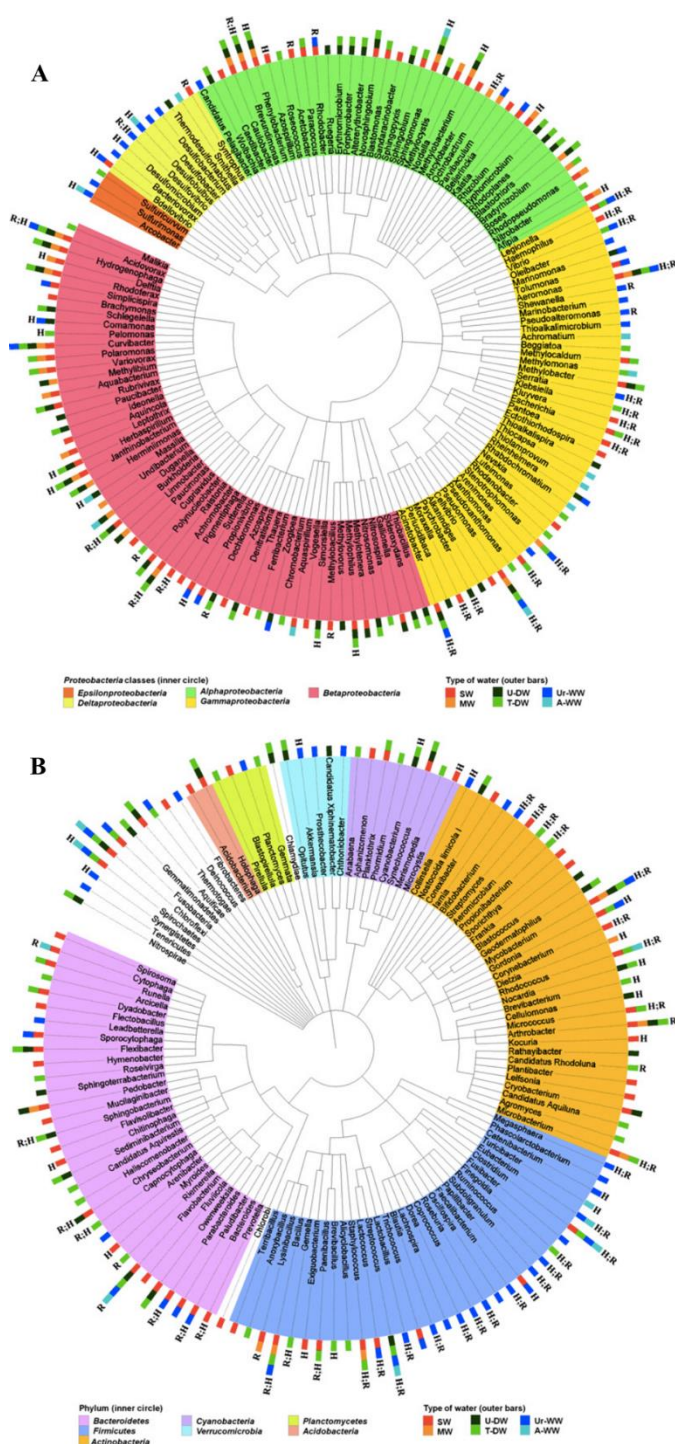


Figure 5. Representations of the bacterial diversity in freshwater, from (Vaz-Moreira, Nunes, and Manaia 2014).

(A) *Proteobacteria* classes and (B) other phyla observed in different types of water. Different phyla or *Proteobacteria* classes (inner circle) are represented by different colors. Presence in different types of water are represented by the outer bars. Types of water: SW, surface water; MW, mineral drinking water; U-DW, untreated drinking water; T-DW, treated drinking water; Ur-WW, urban domestic wastewater; A-WW, animal wastewater. Occurrence in the human associated microbiome (H) and previous description of antibiotic resistance genes (R).

Factors influencing bacterial diversity include metabolic differences, evolutionary history, microbial biogeography, among others. In a study of bacterial biogeography in soils, it was found that the structure of soil bacterial communities is not random at the continental scale and that the diversity and composition of soil bacterial communities at large spatial scales can be predicted only by the soil pH. This reflects the importance of the study of the relationship between habitat properties and bacterial diversity (Fierer and Jackson 2006).

Another example of how environmental factors shape the genetic structure and diversity of bacterial populations comes from a literature survey (Tenailon et al. 2010). Comparative analyses of *E. coli* commensal strains from the *E. coli* reference collection (ECOR) revealed an important diversity within this species, including the identification of four main phylogenetic groups (A, B1, B2 and D) and two accessory groups (C and E). Moreover, they show how the intense selective pressures that the host exerts on commensal *E. coli* shape the plasticity of their genome and thus illustrate the diversity of adaptive paths within the species.

iii. Bacterial diversity in clinical settings

a) Diversity in *Escherichia coli* infections

The evolutionary history of diversity and virulence of uropathogenic *E. coli* (UPEC) causing extraintestinal infections was assessed using 19 *E. coli* isolates from patients with urinary tract infections. The genome of the 19 isolates, among which 14 caused urinary tract infections and 5 were commensal, were sequenced using whole-genome deep-sequencing. The genome phylogeny showed that the divergence between UPEC and commensal *E. coli* strains happened over the last 32 millions of generations, and that the emergence of uropathogenicity was used opportunistically to cause extraintestinal infections (Lo et al. 2015).

In an interdisciplinary study (Levert et al. 2010), 226 *E. coli* isolates from deep and closed visceral infections were analyzed in order to determine their genomic and phenotypic diversity. A strong genomic variability was observed, either by *E. coli* polyclonal infections or by infections with a single clone that exhibit micro-heterogeneity after diversification events during the infectious process. The phenotypic diversity from the isolates with genomic micro-heterogeneity was high including variability in antibiotic resistance, outer membrane permeability, growth rate, stress resistance and virulence properties (Figure 6).

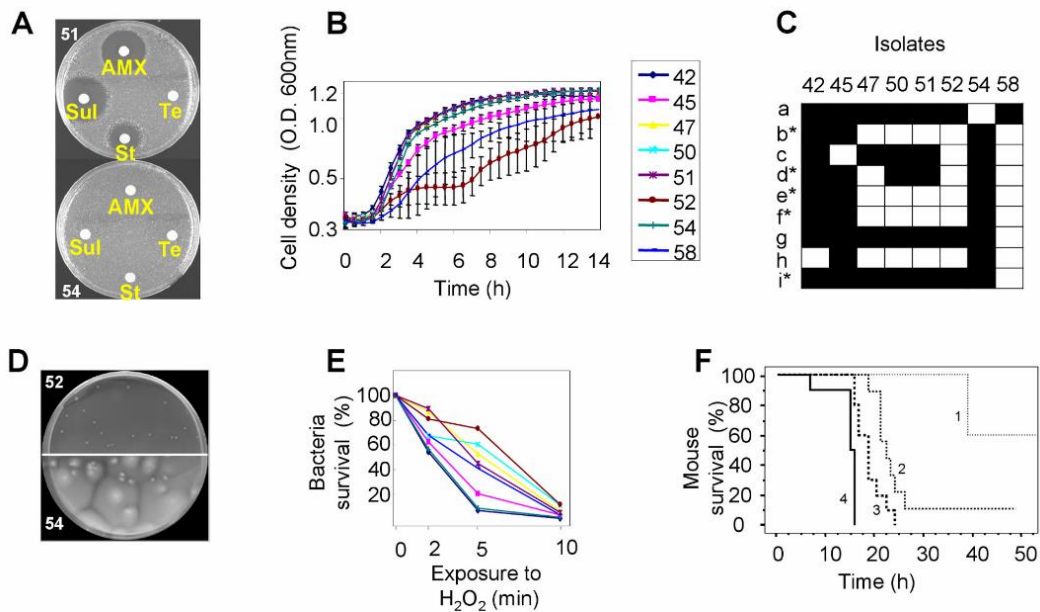


Figure 6. Phenotypic diversity observed in *E. coli* isolates of patient No. 3 infected by a unique clone isolates from blood and liver abscess, after (Levert et al. 2010).

(A) Variability in antibiotic resistance. AMX: amoxicillin, Sul: sulfonamide, St: streptomycin, Te: tetracycline. (B) Growth curves in LB medium at 37°C, showing impaired growth of isolates 45, 52 and 58. (C) Carbon source utilization (Biolog GN2 plates, 95 substrates). Black and white squares indicate the use or not of the substrate, respectively. (D) Non motile (52) and motile (54) isolates in 0.35% agar plates incubated 48 h at 37°C. (E) Sensitivity to H₂O₂. (F) Survival curves of mice subcutaneously inoculated with different isolates.

Diversity of pathogenic *E. coli* causing extraintestinal infections, as for commensal strains (Tenailon et al. 2010), varies with patient's characteristics. Indeed, Banerjee et al. found among 229 clinical isolates of *E. coli* a broad distribution of sequence types according to patient features such as age, type of infection and infection origin (Banerjee et al. 2013). This reflects how the intense selective pressures exerted by the host shapes diversity of pathogenic *E. coli*

b) Long-term study of diversity of bacterial infections

A pioneering study studied the evolutionary dynamics of *Pseudomonas aeruginosa* populations evolving over 200,000 generations in the complex natural and structured environment of cystic fibrosis (CF) airways (Yang et al. 2011). This study was based on a collection of *P. aeruginosa* isolates sampled from several hundreds of Danish CF patients between 1973 and 2008 (Figure 7).

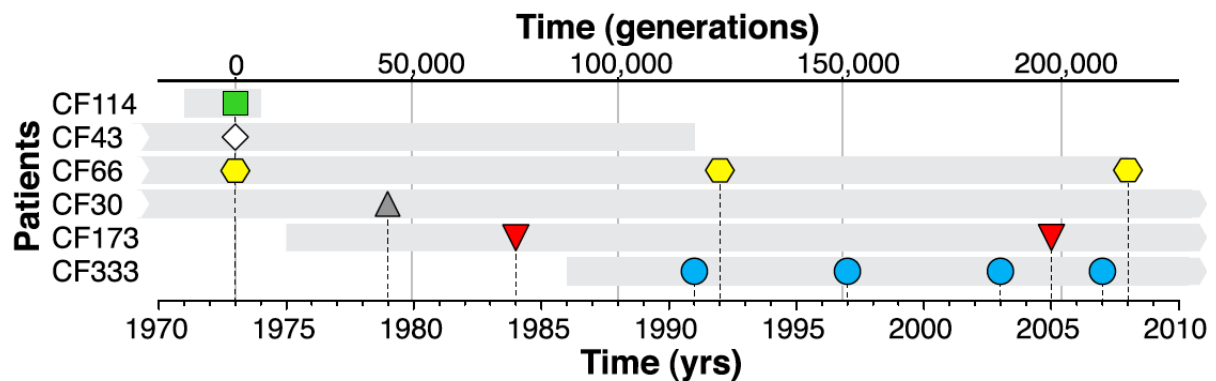


Figure 7. Isolate sampling points and patient life span, after (Yang et al. 2011).

P. aeruginosa isolates were collected from six different CF patients during a 35-y time period. Bacterial isolates are represented by the colored symbols, and gray bars represent patient life span.

Genomes from the isolated strains were sequenced. This allowed the construction of an evolutionary tree showing the relationships between strains. Within the samples, those dating from before 1979 were more heterogeneous. Important beneficial mutations accumulated in the early years (before 1979) ensuring the subsequent reproductive success of *P. aeruginosa* in the CF airways in several hosts. After 1979, the tree demonstrates a remarkably limited diversification of the successive isolates. These results are consistent with those observed during *in vitro* evolution experiments, where fitness increases rapidly at the beginning of adaptation to then increase much more slowly (Lenski and Travisano 1994, Wiser, Ribbeck, and Lenski 2013).

Among all the single nucleotide polymorphism (SNPs) affecting coding regions, 25% result in synonymous changes. This sign of negative selection is most prevalent in isolates after 1979. This observation supports a hypothesis that a selection of adaptive mutations takes place before 1979 leading to a highly adapted clone with limited diversification. Indeed, global transcription profiles and catabolic analyses revealed a strong divergence only in the isolates sampled before 1979. The phenotypic similarity of isolates sampled after 1979 indicates parallel evolution. Analyses of two isolates from two different patients (CF173 and CF333) after 1979 revealed independent evolution of parallel phenotypes (Figure 8). Indeed, both isolates lost the capacity to catabolize 4-hydroxyphenylacetic acid (4-HPA), had increased resistance to ciprofloxacin, and acquired mutations in the genes *PA5160* and *mexY* that encode components of multidrug transport systems.

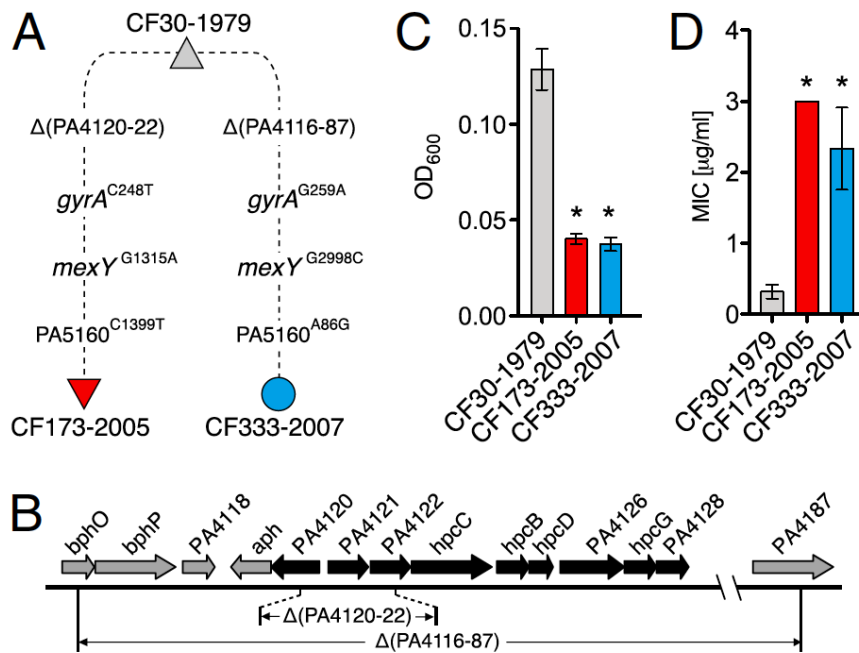


Figure 8. Parallel evolution of *P. aeruginosa* of isolates from patients CF333 and CF173, after (Yang et al. 2011).

(A) Schematic representation of the four parallel and independent genetic changes observed in the two CF333 and CF173 isolates. The genetic changes include missense/nonsense mutations in three genes (*gyrA*, *mexY*, and *PA5160*) and deletions of genes involved in 4-HPA catabolism. (B) Outline of the chromosomal region with genes required for 4-HPA catabolism (labeled in black) and positions of the deletions *Δ(PA4116-87)* (CF333) and *Δ(PA4120-22)* (CF173). (C) Growth of strains in minimal medium with 4-HPA as carbon source. (D) Ciprofloxacin minimal inhibitory concentration (MIC; µg/mL) values.

Gene expression changes can be explained by mutations occurring in three genes encoding global regulators (*mucA*, *rpoN*, *lasR*). These mutations occurred between 1973 and 1979. Indeed, during the first 42,000 estimated generations (before 1979), 11 mutations affecting these global regulators were found in contrast with 5 mutations during the following 167,000 estimated generations after 1979. These results are in accordance with the general theory of adaptation that states that phenotypic innovations are more likely to emerge through changes to regulatory rather than structural genes (Stern 2000).

This study revealed the complexity of the long-term evolution of *P. aeruginosa* in CF airways. First, bacteria experienced an initial period of rapid adaptation upon colonization that ensured survival in the new environment. Second, they experienced a long period of limited phenotypic changes suggesting that the evolving lineage has reached a major adaptive peak in the fitness landscape.

A second study of long-term evolutionary dynamics during infection was performed using isolates of a *Burkholderia dolosa* outbreak among CF patients. A total of 112 genomes were sequenced from isolates sampled from 14 patients over 16 years (Lieberman et al. 2011). The *B. dolosa* isolates were recovered from the airways and from blood of patients with bacteremia. This collection covers the epidemics with high temporal resolution and enables the study of parallel evolution of the same strain in multiple human beings. Bacterial genome sequences evidenced a steady accumulation of mutations (~2 SNPs per year) that generated high genetic diversity between the isolates. Analyzing the epidemic phylogenetic tree allowed to identify the likely network of transmission among the 14 patients (Figure 9). Bacterial isolates from the same patient tended to form genetically-related clusters, yet, the last common ancestor (LCA) of all isolates came from patient-zero, then transmission occurred among patients, either directly from one patient to another or through a healthcare worker or a medical device.

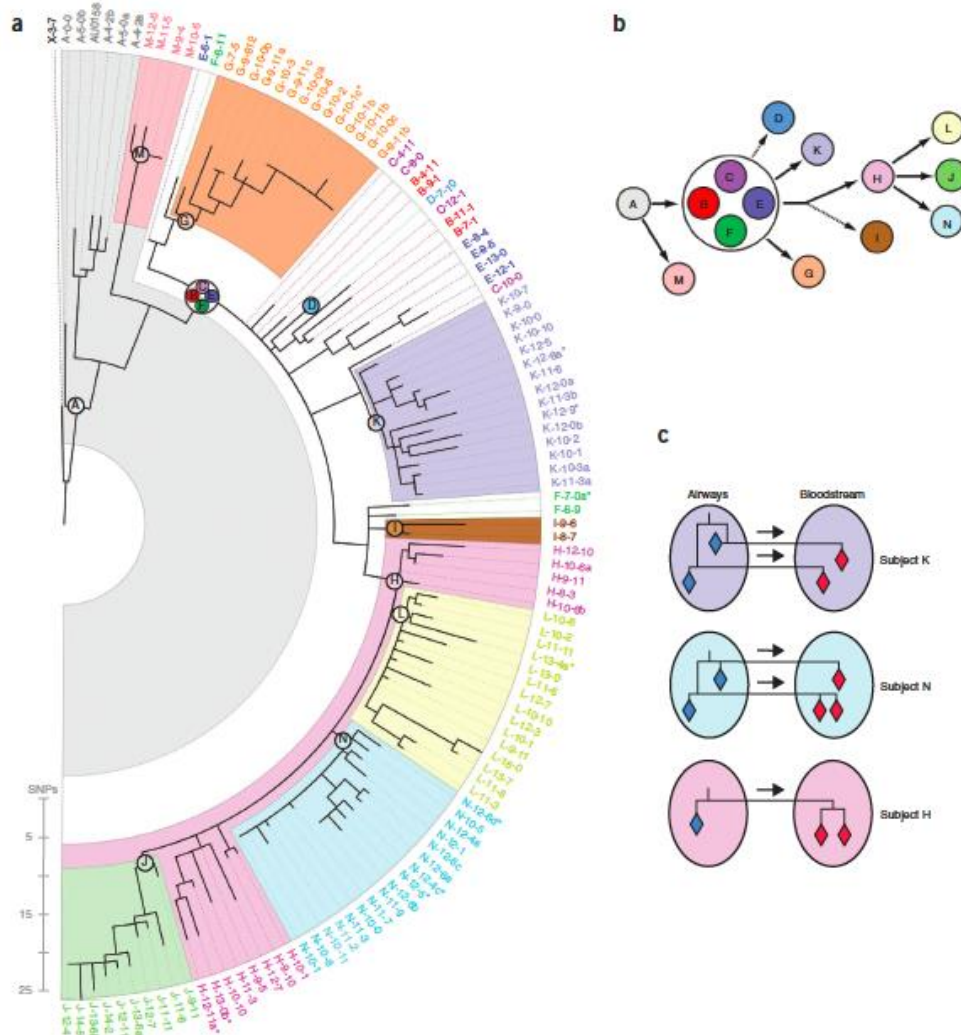


Figure 9. Bacterial phylogeny reveals a likely network of transmission between individuals and between organs, after (Lieberman et al. 2011).

(a) Maximum-likelihood phylogenetic tree (SNP scale). The 112 isolates are indicated by thin dashed lines colored according to patient and labeled according to patient and time. Blood isolates are indicated by an asterisk and isolates with the same patient ID and date are distinguished by letters. The LCA (last common ancestor) of isolates from the same patient is represented as a circle of the appropriate color and label. Colored backgrounds indicate patient-specific genetic fingerprints. Patients B, C, E and F share the same LCA (white background). (b) Phylogeny between the inferred LCAs suggests a likely network of infection between patients (arrows). Dashed arrows indicate less certainty (fewer than three isolates). (c) Phylogeny between blood and lung isolates recovered from the same patient shows the transmission of multiple clones to the bloodstream during bacteremia (multiple arrows, patients K and N).

As observed in the *P. aeruginosa* study, a strong parallelism between isolates from independent patients was detected with mutations in similar genes being independently acquired within the individual after the initial infection event. These genes confer fluoroquinolone resistance, restoration of full-length O-antigen (which was truncated in the

ancestor) and 17 of them are unknown to play a role in pathogenesis. Among these 17 genes, 11 are related to pathogenicity, including membrane synthesis (4 genes, including 2 involved in LPS biosynthesis), secretion (2 genes) and antibiotic resistance (5 genes). The other 6 are related to regulation of gene expression.

Long-term follow up of the evolutionary dynamics of microorganisms during infection allows the comprehensive understanding of genetic adaptation during pathogenesis and the identification of genes central to pathogenesis.

iv. Bacterial diversity in experimental settings

The study of adaptive diversification in clinical settings requires the combined work of clinicians and researchers in order to guaranty the long-term follow up of patients and their bacterial isolates, without talking about the ethical issues of dealing with human samples. Moreover, the intrinsic aspects of clinical practice do not allow a complete control of all evolutionary settings.

In vitro experimental settings offer an interesting alternative for the study of evolution and in particular of bacterial adaptive diversification. Indeed, easy manipulation, simple nutritional requirements, high growth rates and large population sizes of bacteria allow easy propagation of evolving populations. Moreover, bacteria may be preserved as frozen suspensions allowing the constitution of living fossil records that may be revived to study in real-time the dynamics of evolution.

The emergence and long-term establishment of adaptive diversification in experimental settings is rare. Experimental evolving populations are diverse in nature. However, this diversity is transient, since periodic selection events purge diversity. Yet, events of adaptive diversification during *in vitro* evolution experiments have been reported. They are frequently occurring in heterogeneous environments with spatial structure or several carbon sources, or in homogeneous environments where one bacterial subpopulation secretes a metabolite by-product, constructing a new niche that another bacterial type can exploit. These *in vitro* events of adaptive diversification allow the study of the main ecological rules that dictate the emergence of new ecotypes and adaptive radiation as a first step toward specialization.

a) Adaptive diversification in environments with spatial structure

Rainey et al. studied the effect of ecological opportunity on the evolution of genetic diversity using *Pseudomonas fluorescens*. They generated a series of microcosms from replicate isogenic populations of the ancestral (SM, smooth) morph in spatially heterogeneous environments provided by static broth cultures (Rainey and Travisano 1998). After only 3 days of incubation in such conditions, a morphological diversification was observed with the emergence of three dominant morphs each presenting a specific colony type and a specific ecological niche (Figure 10). The ancestral SM morph colonizes the liquid media, a wrinkly-spreader (WS) morph occupies the air-broth interface, and a fuzzy spreader (FS) morph colonizes the bottom of the media.

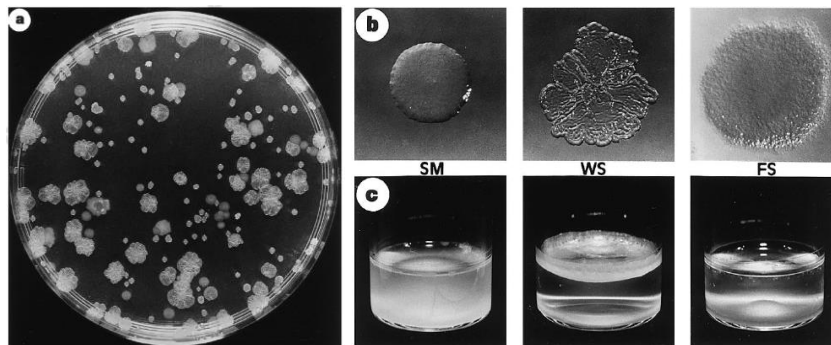


Figure 10. Phenotypic diversity and niche specificity, after (Rainey and Travisano 1998).

(A) After 7 days of incubation, populations show substantial phenotypic diversity which is observed after plating. (B) Colony types: smooth (SM), wrinkly-spreader (WS) and fuzzy spreader (FS). (C) Evolved morphs showed marked niche preferences.

The ecological mechanism maintaining diversity is strictly related to oxygen availability in the environment that stimulates competition for oxygen between morphs. For example, the WS niche (air-liquid interface) provides it with access to both oxygen and nutrients and so they are selectively favored at the expense of genotypes within the oxygen-poor broth phase. Indeed, elimination of the physical structure of the environment by shaking the cultures eliminates the multiplicity of niches and thus the morphological variation. Moreover, the maintenance of diversity is related to negative frequency-dependent interactions between the ancestor SM morph and derived WS and FS morphs. In the WS case, once it becomes common, its fitness

declines because the weight of the mat increases beyond the point at which it is self-supporting, and it sinks.

The genetic bases of the WS phenotype were identified. Indeed, WS genotypes overproduce an acetylated form of a cellulose-like polymer (CLP) that is necessary for the formation of the biofilm mat at the air-broth interface. The CLP production requires a set of 10 enzymes encoded by the operon *wss* (Spiers et al. 2002, Spiers et al. 2003). Mutations in genes from three different pathways that activate this operon through the production of cyclic-di-GMP were found in evolved WS clones (Bantinaki et al. 2007, McDonald et al. 2009). The first pathway, named Wsp, includes a four-protein receptor-signaling complex bound to the membrane (WspA WspB, WspD and WspE). This complex receives the signal from the methyltransferase WspC or the methylesterase WspF and controls the activity of WspR, which catalyzes the synthesis of cyclic-di-GMP. The second pathway, called Aws, includes two proteins with di-guanylate cyclase activity, AwsO, and AwsR together with their regulator AwsX. Finally, the Mws pathway includes the MwsR that has a DGC and negative self-regulation activity. Mutations in genes encoding WspF, WspE, AwsO, AwsR, AwsX and MwsR were found in evolved WS morphs. These mutations probably increase the production of cyclic-di-GMP, thereby activating the genes encoding the enzymes responsible for CLP production which leads to the WS phenotype.

b) Adaptive diversification in environments with nutritional structure

Adaptive diversification events in nature are frequently related to resource competition and invasion of available niches (Schluter 2000b). Several studies in the laboratory also highlighted the importance of nutritional structure on the emergence of diversity (Friesen et al. 2004, Spencer et al. 2007).

Independent experimental *E. coli* populations were propagated in serial batch cultures with minimal media supplemented with a mixture of glucose and acetate (205 mg/L of each). In the presence of two carbon sources, *E. coli* consumes first its favorite carbon source, in this case glucose. Acetate consumption is subject to catabolite repression and will be consumed only when glucose is exhausted (Vemuri et al. 2006). These metabolic patterns are known as diauxic growth with a first exponential phase, corresponding to consumption of glucose followed by a deceleration or diauxic shift, and then a second exponential phase corresponding to acetate consumption (Vemuri et al. 2006)

After 1,000 generations of evolution in these conditions, the experimental populations that were initiated from the same ancestral strain, became polymorphic for colony size after plating. Two types of colonies were identified, including small and large ones. Phenotypic analyses of the emerged polymorphic strains revealed that they differ not only in their colony size but also in their diauxic pattern. The first cluster, characterized by large colonies, grew better in glucose with a lag phase when switching to acetate similar than the ancestor. This cluster was named slow-switchers (SS). The second cluster, characterized by small colonies, had a much shorter lag when switching to acetate and was thus named fast-switcher (FS) (Figure 11A).

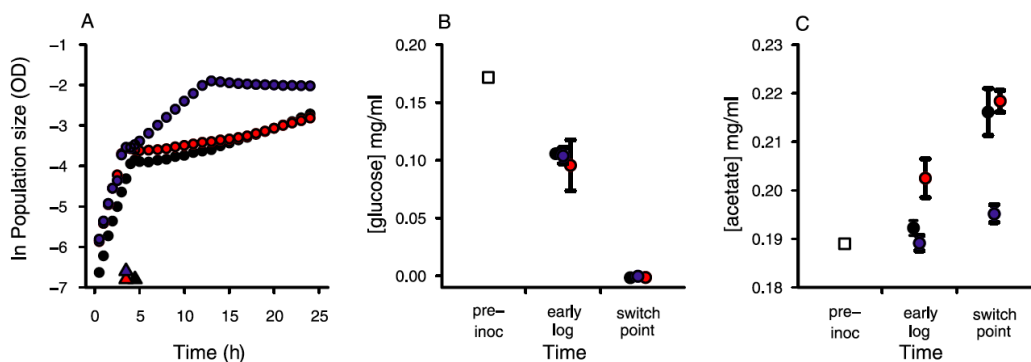


Figure 11. Growth profiles of each isolated strain on glucose/acetate mixture and the change in glucose and acetate concentrations during the first phase of diauxic growth, after (Spencer et al. 2007)

(A) Three isolated genotypes (FS, SS, and Anc) consumed first glucose (from 0 to ~4 hours), then acetate (hours ~4 until maximum OD), in a standard diauxic (two-stage) growth pattern. The SS (red) closely paralleled the ancestral (black) strain for its growth pattern on acetate, while the FS (blue) exhibited a much faster diauxic switch from growth on glucose to acetate and a higher growth rate on acetate (second stage of diauxic growth). Triangles on the horizontal axis indicate the switching points for each

strain. (B) All strains deplete glucose from the medium during the first phase of diauxic growth (from inoculation into batch culture to the time of diauxic switch). (C) Change of acetate concentration in the medium during the first phase of diauxic growth. The ancestor and the SS ecotype accumulated acetate as they depleted glucose; the FS strain did not accumulate acetate during glucose depletion, indicating that acetate is consumed.

This differential growth pattern implies differences in carbon consumption. Indeed, the authors monitored the changes of glucose and acetate concentrations during growth on glucose. Both the ancestor and SS type produced acetate as a by-product during glucose consumption, so that acetate concentration increased in the medium (Figure 11B and C). Conversely, the FS type did not accumulate acetate as it consumed glucose, indicating acetate consumption during growth on glucose (Spencer et al. 2007). The SS and FS coexistence is maintained by negative frequency-dependence generated by the daily sequential depletion of resources (Friesen et al. 2004), which is likely to be generated by a tradeoff between the metabolism of the two carbon sources, glucose and acetate, present in the medium.

Consumption of acetate is dependent on the expression of the *aceBAK* operon, which converts acetyl-CoA derivatives to malate. The operon contains genes encoding malate synthase A (*aceB*), isocitrate lyase (*aceA*), and isocitrate dehydrogenase kinase/phosphatase (*aceK*), all of which are co-expressed and subject to catabolite repression in the presence of glucose (Oh et al. 2002). In FS clones, the expression of these genes is deregulated and constitutive, even in the presence of glucose. This deregulation is due to an *IS1* insertion in the isocitrate lyase repressor gene *iclR*, encoding a negative regulator of *aceBAK*. The IS insertion inactivates *iclR*.

Global transcription profiles of SS and FS clones showed that this diversification event was associated with significant changes in expression of genes involved in central metabolism (Le Gac et al. 2008). The FS clones revealed an overexpression of genes encoding enzymes of the tricarboxylic acid (TCA) cycle (*sucA*, *sucB*, *sucC*, *sucD*), the glyoxylate shunt (*aceB*), involved in the consumption of acetate (*acs*) and anaerobic respiration (*fadD*), while the SS

clones exhibited increased transcription of genes involved in acetate secretion (Figure 12). These patterns of gene expression suggest that the emergence of this diversification event relies on competition between SS and FS for the available oxygen and carbon sources.

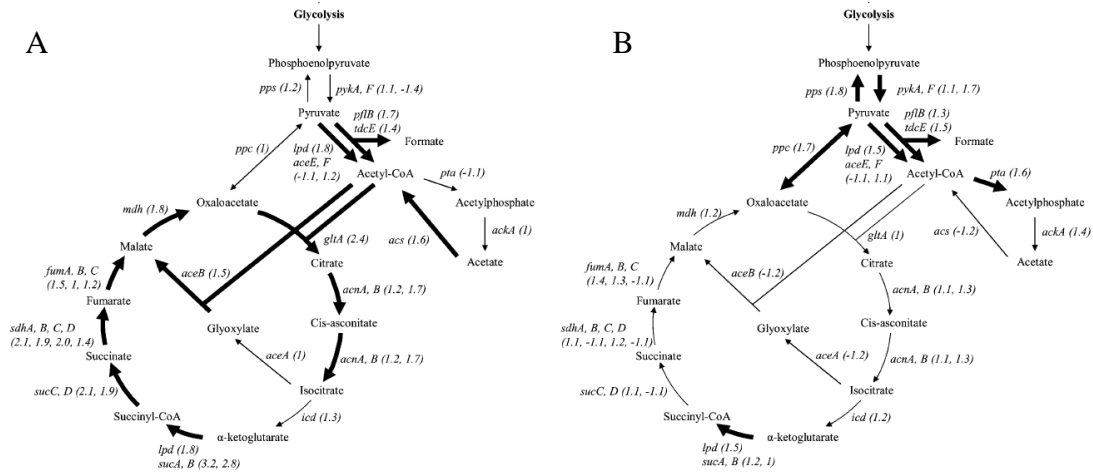


Figure 12. TCA, glyoxylate shunt, and acetate consumption/excretion metabolic pathways showing different trends in FS and SS compared to the ancestral strain, after (Le Gac et al. 2008).

Enzymes are indicated by italics and fold changes compared to the ancestor are shown in parentheses. A thick arrow indicates a fold change ≥ 1.5 . (A) Comparison between FS and ancestor. (B) Comparison between SS and ancestor.

To identify genes with parallel changes in gene expression, which is known to be a hallmark of adaptive changes (Croizat et al. 2005), the authors performed a cluster analysis of differentially expressed genes. Three main gene clusters were identified. The first includes 218 genes that had higher expression in both FS and SS. These genes are likely to be linked to the adaptation of the derived types to log growth in the evolution environment and these changes probably appeared before the diversification event. The second cluster includes 144 genes that were more highly expressed in both FS and SS in stationary phase. These genes may thus be involved in adaptation to survival in starvation. The third cluster includes 88 genes that have higher expression in FS at the switching timepoint and are therefore most likely involved in allowing FS to grow while SS is in stationary phase.

Among the genes differentially expressed between SS and FS, many are controlled by ArcA, a global regulator of genes related to central metabolism in *E. coli*. Mutations were found

in the *arcA* gene only in FS clones, thereby suggesting its involvement in the differential gene expression patterns observed between SS and FS. However, the role of ArcA in FS emergence and establishment is not clear. Indeed, the *arcA* mutation was identified at 1000 generations, while FS emerged at generation 200 (Le Gac and Doebeli 2010).

c) Adaptive diversification in homogeneous environments

These environments can be achieved by the use of batch cultures or chemostats with a single carbon source. In batch cultures, after a lag phase, population expansion proceeds at a maximal rate (exponential phase) due to nutrient abundance. Once a population reaches high density, the exhaustion of nutrients leads to the cessation of growth (stationary phase).

In a chemostat, the continuous addition of medium containing a single growth-limiting nutrient is performed simultaneously to the removal of culture, to achieve a stable equilibrium. In this steady state, the rate at which the population grows is equal to the rate of dilution which allows the experimental control of growth rate by modulation of the rate of culture dilution (Figure 13) (Gresham and Hong 2015).

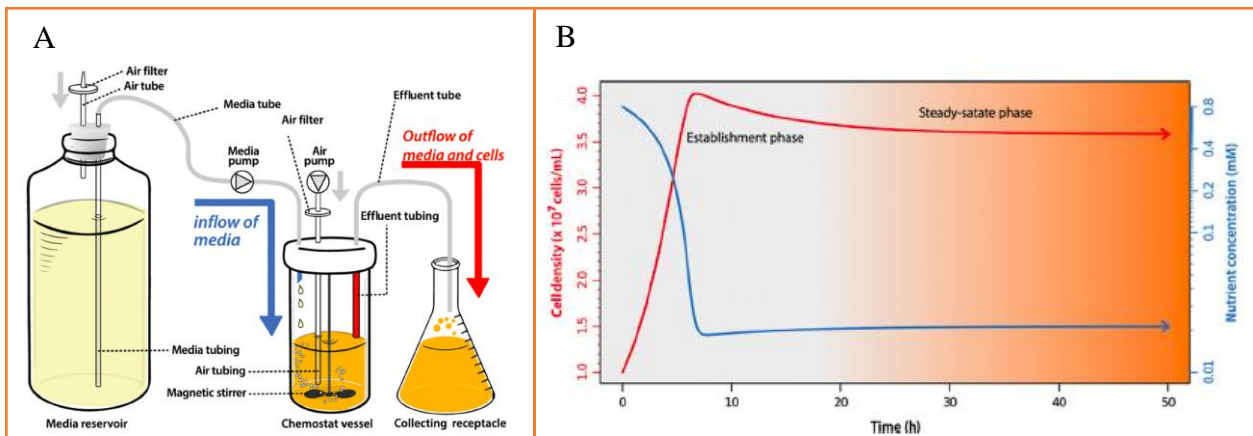


Figure 13. Chemostat design and establishment of a steady state, after (Gresham and Hong 2015).

(A) A chemostat includes a culture vessel in which the population grows under continuous agitation and aeration. New media flows into the vessel at a defined rate. At the same rate, culture containing cells and medium is removed. The flow of media and culture is maintained using a pumping apparatus and holding the chemostat vessel under positive pressure by means of a constant air flow. (B) Following inoculation and initiation of culture dilution, the chemostat is characterized by a period during which the population increases and nutrient abundance declines. Eventually, a steady state is established in which the cell population remains high and the concentration of the limiting nutrient low.

One of the first experiments of *in vitro* evolution describing adaptive diversification in homogeneous environments was published by Rosenzweig et al. in 1994. In this experiment, evolving populations were initiated from a single *E. coli* clone and maintained in chemostat cultures of minimal media supplemented with only glucose at a concentration fluctuating between 0.00625% (w/v) and 0.025% (w/v). After 765 generations of evolution, the populations evolved extensive polymorphisms that remained stable for at least ~1200 generations. Three morphotypes, defined by colony size after plating on agar plates, were identified, and named CV101, CV103 and CV116 (Rosenzweig et al. 1994).

Estimation of maximum growth rates under non-limiting conditions and analyses of spent media of the three evolved types revealed significant differences among them. The CV103 strain revealed improved glucose uptake kinetics with associated secretion of both acetate and glycerol. These new constructed ecological niches allowed the emergence of the other two types CV101 and CV116. The CV101 strain consumes constitutively acetate without being subject to catabolic repression in the presence of glucose. The CV116 strain presents an improved consumption of glycerol.

The glucose-limited chemostat evolution experiment was reproduced by starting 12 independent populations from the same ancestor (Treves et al. 1998). Six of the twelve populations evolved polymorphisms maintained by acetate cross-feeding, all of them related to semi-constitutive overexpression of *acs*, the gene encoding acetyl CoA synthase. Changes in *acs* expression were associated with modifications of its promoter region (Figure 14). Indeed, mobile genetic element insertions or single nucleotide substitutions in the *acs* promoter region were found in the acetate scavengers (Treves, Manning, and Adams 1998).

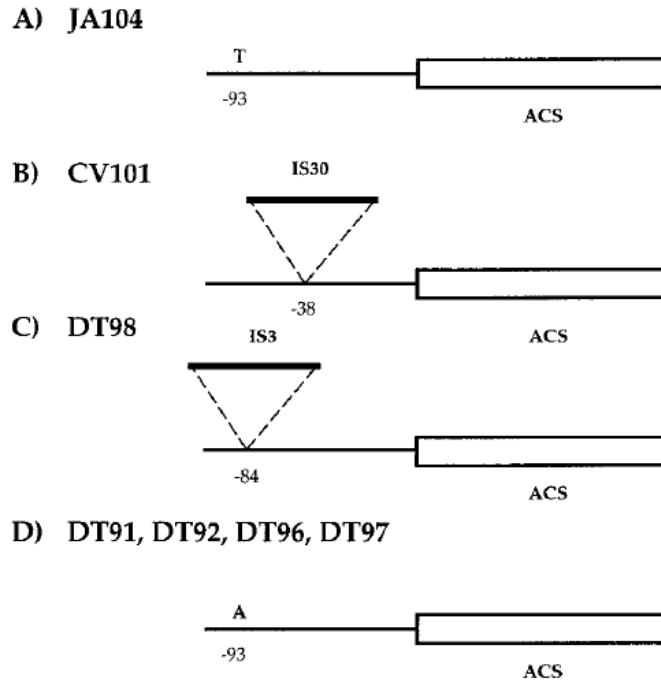


Figure 14. Structure of *acs* from ancestral and acetate-scavenging strains of *E. coli*, after (Treves, Manning, and Adams 1998).

(A) The ancestral progenitor strain JA104. (B) IS30 insertion, (C) IS3 insertion, (D) T>A transversion in the acetate-scavenging clones from three independent populations.

This study revealed how *E. coli* populations growing in a simple unstructured environment may evolve an interactive, polymorphic state owing to the emergence of ecotypes capable of modifying their environment, thereby generating new ecological niches that are exploitable for new emergent lineages.

II. The Long-Term Evolution Experiment with *Escherichia coli*

i. Experimental setting

The longest still running evolution experiment is the Long-Term Evolution Experiment (LTEE) with *E. coli*. This experiment was started in February 1988 by Richard E. Lenski and aims to study the principles of evolutionary processes. Today, the LTEE has enlarged this initial aim and is not only used as a model to study evolution but also fundamental issues of bacteria biology such as genome plasticity, regulatory networks and even antibiotic resistance.

The LTEE was started from a clonal population of *Escherichia coli* B REL606 (Jeong et al. 2009), and its Ara⁺ revertant named REL607. Indeed, REL606 is unable to use arabinose as sole carbon source, and spontaneous revertants can be selected on arabinose plates owing to a point mutation in the gene *araA* which is able to restore the Ara⁺ phenotype. The capacity to consume arabinose is used as a phenotypic marker in competition experiments aiming to determine the fitness of two strains in direct competition and it was shown to be neutral in the conditions of the LTEE (Lenski et al. 1991).

From each of the two REL606 and REL607 strains used as ancestors, 6 independent populations were started, and named Ara⁻1 to Ara⁻6 and Ara⁺1 to Ara⁺6, respectively. Each population is propagated since 1988 in a 50-mL Erlenmeyer flask containing 10 mL of Davis minimal medium (Davis and Mingioli 1950) supplemented with 2 µg/L of thiamine and 25 g/L of glucose (DM25). Cultures are propagated with constant agitation at 120 rpm at 37°C for 24h and are transferred daily by diluting 100 µL into 9.9 mL of fresh media. During this 24h cycle, bacterial cells undergo a cycle of feast and famine reaching stationary phase densities of $\sim 5 \times 10^7$ cells/mL. The 100-fold daily growth of each population allows for ~ 6.7 generations for each population. Therefore, today the LTEE accounts for more than 65,000 generations of evolution, twelve times independently from a common ancestor. The ancestor and all evolved intermediates are stored every 500 to 1000 generations as glycerol suspensions at -80°C,

thereby providing a complete fossil record that can be revived at any time to perform direct comparisons (Figure 15).

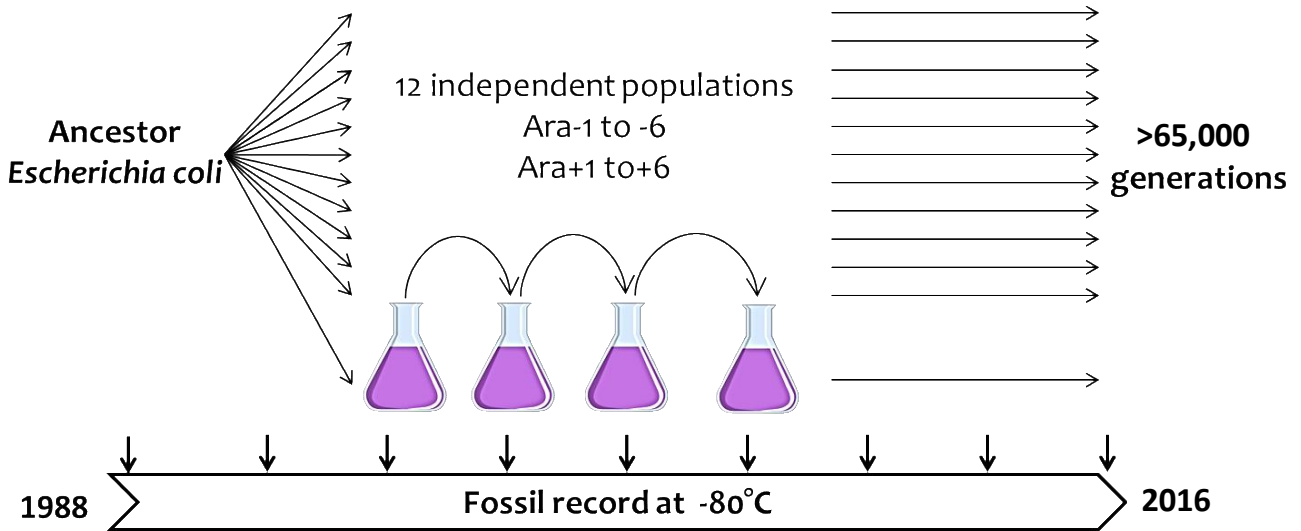


Figure 15. Representation of the LTEE with *E. coli*.

Twelve populations, called Ara-1 to Ara-6 and Ara+1 to Ara+6, are propagated since 1988 from a common ancestor of *E. coli* B.

The REL606 strain is strictly asexual since it carries no plasmid and no functional bacteriophage. Its asexual nature prevents the generation of genetic variation by recombination with DNA from external sources (*e.g.*, horizontal transfer). Therefore, genetic variation results from mutations that may fix owing to selective sweeps.

ii. Parallel evolution

One important question in evolutionary biology is whether adaptive changes through natural selection are “reproducible”. In other words, if life would be set up several times independently in the same conditions (in parallel), what is the probability to obtain similar evolutionary outcomes in each replicate, at each possible level (phenotype, genotype). The LTEE may answer this question, and it is one of the reasons the experiment was started with 12 independent replicate populations which would allow to identify, if they exist, parallel changes.

Many studies have been performed in the evolving populations of the LTEE to determine whether phenotypic and genotypic traits evolved in parallel. The most important trait that shows parallel evolution in all 12 LTEE populations is fitness, the ultimate measure of ecological success. The 12 replicate populations increased their relative fitness compared to their common ancestor by about 80% in the evolutionary medium after 20,000 generations (Philippe et al. 2007). The rate of fitness increase was very fast during the first 2,000 generations, and became progressively slower over evolutionary time (Lenski and Travisano 1994). Even if the fitness trajectories of the LTEE populations declined over time, it has no upper limit, implying that adaptation can continue indefinitely, or at least for a long time, even in a constant environment (Figure 16) (Lenski et al. 2015).

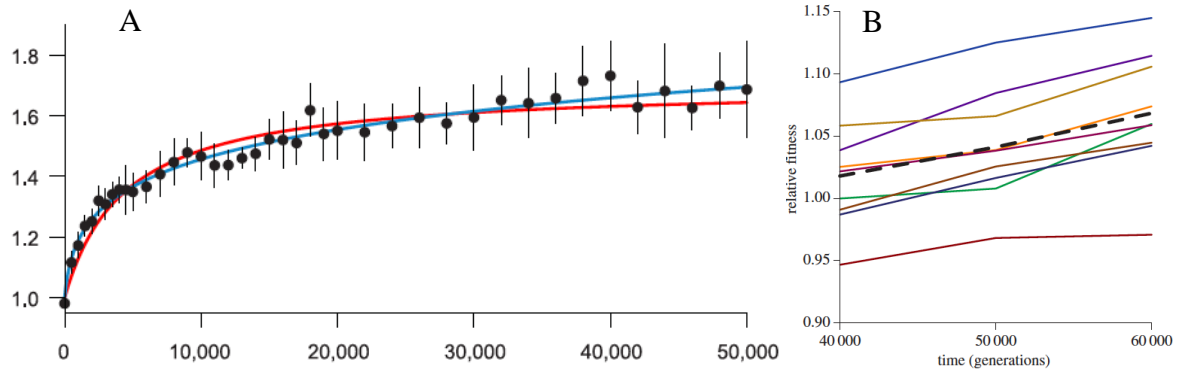


Figure 16. Fitness trajectories of the 12 LTEE populations, after (Wiser, Ribeck, and Lenski 2013, Lenski et al. 2015).

(A) Fitness trajectories during 50,000 generations. Hyperbolic (red) and power-law (blue) models fit to the set of mean fitness values (black symbols) from all 12 populations. (B) Trajectories of mean fitness for nine of the 12 LTEE populations (thin lines) and grand-mean fitness (thick dashed line).

These gains on fitness proved the adaptation of bacterial cells to their environment and have been related to the improvement of glucose consumption. During evolution, bacteria reduced their lag-phase and grew faster during exponential phase in the evolutionary medium (Vasi, Travisano, and Lenski 1994). These adaptations are accompanied by decay of unused catabolic functions in all populations, thereby revealing glucose specialization (Cooper and Lenski 2000). These catabolic declines occurred early in the experiment during the fast adaptation period when beneficial mutations of high fitness effects were rapidly fixed. This process is correlated with antagonistic pleiotropy meaning that the same beneficial mutations confer high fitness in the glucose evolutionary environment and at the same time were responsible for the other catabolic decays. The most striking example is the loss of ability to use D-ribose as a carbon source during the first 2,000 generations of evolution in all 12 evolving populations. The Rib⁻ phenotype is associated with a 1 to 2% gain of fitness in the evolutionary medium when compared to the ancestral Rib⁺ strain. The loss of ribose catabolic function involved the deletion of part or all of the *rbs* operon containing the genes necessary for ribose consumption. This deletion was associated with an *IS150* transposition event (Cooper et al. 2001). On the other hand, not only declines in catabolic functions were detected. Indeed,

increased performance on a variety of substrates that the ancestral strain utilized poorly was also observed (Figure 17) (Leiby and Marx 2014).

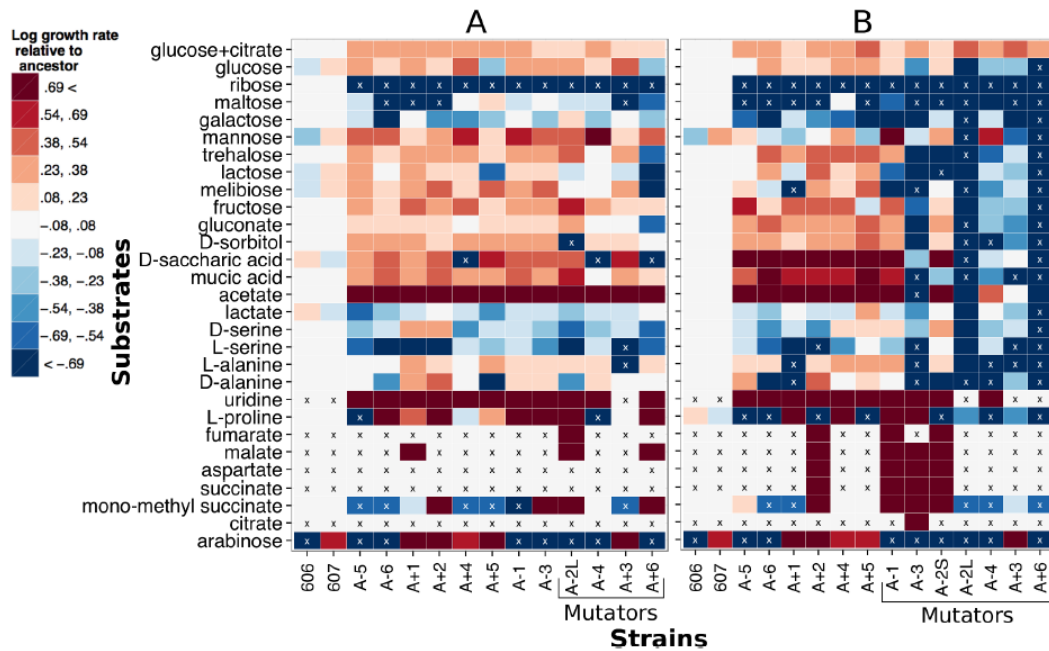


Figure 17. Relative growth rates across a variety of growth substrates for evolved strains, after (Leiby and Marx 2014).

(A) Evolved strains from each of the 12 populations, sampled at 20,000 generations. (B) Evolved strains from each of the 12 populations, sampled at 50,000 generations. Heatmaps indicate the log ratio of growth rates relative to the average of the two ancestors REL606 and REL607 on that carbon source. White indicates a growth rate equal to that of the ancestor average, red faster, and blue slower. The growth rates are plotted on a log scale with the limits of the color range set for twice as fast and half as fast as the ancestor average. An “x” in a box indicates that no growth was observed for that combination of strain and substrate over 48 h. Strains that were mutators by that timepoint are indicated.

Besides the parallel phenotypic changes, the LTEE populations have also experienced parallel changes in genome structure, expression and regulation. Global transcription profiles of one clone isolated at 20,000 generations from each of two populations showed that the expression of many genes was changed in parallel in these two evolved clones compared to their common ancestor (Cooper, Rozen, and Lenski 2003a). These parallel changes affected genes that belong to the cAMP-receptor protein (CRP) and guanosine tetraphosphate (ppGpp) regulons, both of which are known to be regulated by *spoT*. Mutations were subsequently found in the *spoT* gene in 8 of the 12 populations. Moreover, changes in protein expression profiles, analyzed by two-dimensional protein electrophoresis (Pelosi et al. 2006), also revealed parallel

changes in several populations. These analyses allowed the identification of mutations in *malT*, encoding the activator of the maltose operons, in most of the 12 populations. These mutations explained the decay of the ability to use maltose as a carbon source for the evolved populations (Pelosi et al. 2006, Cooper and Lenski 2000).

Parallel changes in DNA structure also occurred in most populations. First, the DNA supercoiling level increased over time in the LTEE populations. Changes in DNA supercoiling were associated to beneficial mutations in the *topA*, *fis* and *dusB* genes (Croizat et al. 2005, Croizat et al. 2010). Second, large DNA rearrangements (deletions, inversions, duplications) have been detected in all 12 populations (Raeside et al. 2014). Moreover, half of the 12 populations have evolved 50 to 100-fold increases in mutation rates, owing to mutations that affect genes involved in DNA repair or removal of oxidized nucleotides (Sniegowski, Gerrish, and Lenski 1997, Wielgoss et al. 2013, Viraphong 2015, Cooper and Lenski 2000). The time of emergence of these hypermutator phenotypes varies among the mutator populations: four evolved it early during the experiment before 10,000 generations, and two much later after 20,000 generations.

Recently, analyses of the genomes of 264 evolved clones sampled over time from all 12 populations of the LTEE allowed to characterize the mutational dynamics over 50,000 generations (Tenaillon et al. 2016). A strong gene-level parallel evolution was detected between the 12 populations. Moreover, the populations that retained the ancestral mutation rate fixed mostly beneficial mutations, the fraction of beneficial mutations declines as fitness increases and neutral mutations accumulate at a constant rate (Tenaillon et al. 2016). These results support the idea that selection favors the fixation of beneficial mutations and places adaptation by natural selection as the keystone of phenotypic evolution.

iii. Emergence of a stable long-term polymorphism in the LTEE

Despite the strong parallelism discussed above, two populations of the LTEE present a unique evolutionary outcome, the emergence of a stable polymorphism, which was unexpected. Indeed, the medium in which the 12 populations are propagated contains only one carbon source usable by *E. coli* during aerobic growth, *i.e.* glucose. Furthermore, the limiting low concentration of glucose (25 mg/L) was expected to support only biomass formation without any carbon left for metabolic by-products. However, against these expectations, two independent diversification events occurred in two populations of the LTEE. The first one, in population Ara-3, is linked to the emergence of a phenotypic innovation, the capacity to metabolize citrate in aerobic conditions. The inability to use citrate in aerobic conditions is even described as a trait of the *E. coli* species. The second, in population Ara-2, is related to the process of niche construction where one type of cells produces a metabolic by-product that can be exploited by an emergent cell type that occurred after mutations. These two events of adaptive diversification provide an excellent opportunity to study the ecological, physiological and genetic bases of the emergence of diversity in unstructured conditions.

a) Emergence of a phenotypic innovation: *E. coli* uses citrate in aerobic conditions

During the LTEE, bacterial cells are evolving in Davis minimal media (DM) supplemented with 25 mg/L of glucose (DM25). The DM composition includes dibasic and monobasic potassium phosphate for their buffering properties and a large amount of sodium citrate (1.7 mM) which serves as a chelating agent. These high concentrations of citrate represent a second potential carbon source. It is however inaccessible for consumption to *E. coli* in these aerobic conditions. Indeed, even if the TCA cycle allows internal citrate metabolism during aerobic growth on other substrates, *E. coli* is unable to metabolize external citrate as a sole carbon source owing to the lack of transporter in oxygenated conditions (Lara

and Stokes 1952). This feature is extremely stable and the resulting Cit⁻ phenotype has long been used as one of the key traits to define *E. coli* as a species (Koser 1923, Scheutz and Strockbine).

A single case of a spontaneous Cit⁺ mutant has been described in *E. coli* (Hall 1982). This mutant derived from an *E. coli* D21 strain and was accidentally isolated after growth on a medium containing phenyl-arabinoiside as a carbon source and citrate as a chelating agent. Citrate utilization arose as the consequence of two mutations in *citA* and *citB*, whose products form the DpiAB two-component system. However they could not explain completely the Cit⁺ phenotype (Hall 1982). These data suggest that even if *E. coli* has the potential to develop the ability to use citrate as a carbon source under aerobic conditions, a complex genetic path is needed to achieve this phenotypic innovation.

The DM25 medium used during the LTEE contains very low glucose concentration and the cultures reach a density of only 5×10^{-7} cells/mL before each daily transfer. Therefore, the culture flasks are only very slightly turbid owing to the low cell concentrations. After ~33,000 generations, the population Ara-3 displayed very high turbidity. Contamination was ruled out by sequencing genomic markers that carry mutations diagnostic from Ara-3. It was however shown that high turbidity was associated with a Cit⁺ phenotype (Blount, Borland, and Lenski 2008). Therefore, Cit⁺ clones emerged in that population and, owing to the high citrate concentration in the DM25 medium, attained high density after growth. The Cit⁺ clones emerged by 31,500 generations, causing an increase in population size and diversity (Figure 18). However, they did not drive Cit⁻ clones to extinction since both types of clones were found to co-exist in the Ara-3 population.

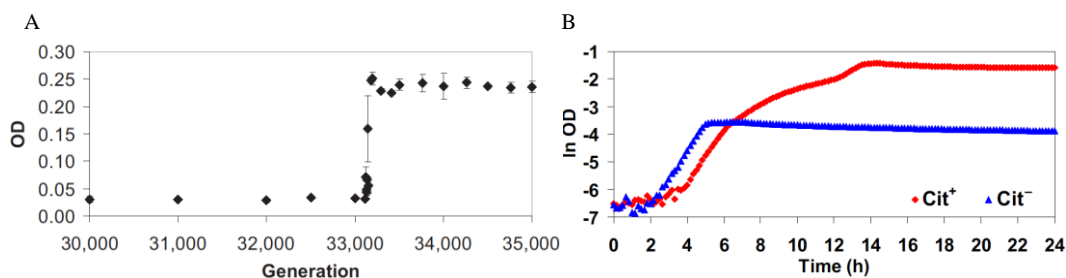


Figure 18. Emergence of the Cit⁺ phenotype in population Ara-3, after (Blount, Borland, and Lenski 2008).

(A) Population expansion during evolution of the Cit⁺ phenotype. (B) Growth of Cit⁻ (blue triangles) and Cit⁺ (red diamonds) cells in DM25 medium.

From the fossil records preserved at -80°C of Ara-3 evolutionary history, “replay” evolution experiments were performed. When evolved clones up to 15,000 generations were used as ancestors during these replay experiments, no emergence of Cit⁺ mutants were observed. However, with clones isolated at 20,000 generations as ancestors, the propagated populations showed a high tendency to evolve the Cit⁺ phenotype, indicating that some potentiating mutation arose between 15,000 and 20,000 generations, and that the Cit⁺ emergence is contingent on the evolutionary history of the population (Blount, Borland, and Lenski 2008).

The co-existence of the two Cit⁻ and Cit⁺ lineages of population Ara-3 relies on negative frequency-dependent interactions whereby one lineage is able to invade the other when rare. This process therefore maintains the polymorphism. This stable co-existence involves differences in metabolite usage, the Cit⁻ lineage being superior to the Cit⁺ in glucose consumption (Figure 18B), allowing it to be maintained as a glucose specialist.

The history and genetic basis of the Cit⁺ trait was analyzed by sequencing and analyzing the genome of 29 evolved clones sampled over time from population Ara-3 (Blount et al. 2012). The phylogenetic history of the population shows that it is polymorphic with several clades (Figure 19). Clade 1 diverged from the ancestor of clades 2 and 3 before 15,000 generations.

Clades 2 and 3 had diverged by generation 20,000. The Cit⁺ lineage arise from clade 3 at about 31,000 generations. Moreover, a mutator phenotype emerged in clade 3.

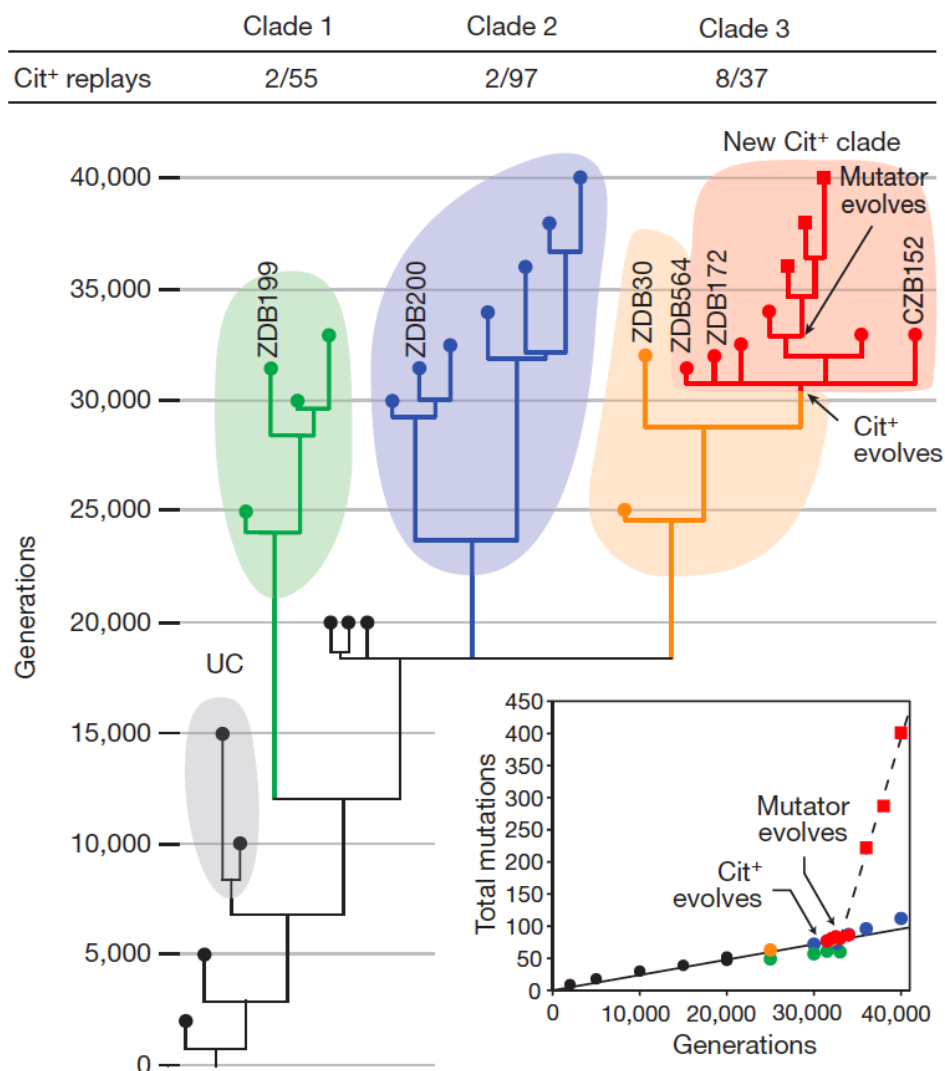


Figure 19. Phylogeny of Ara-3 population, after (Blount et al. 2012).

Symbols at branch tips indicate the 29 clones for which the genome has been sequenced. Shaded areas and colored symbols identify major clades. Fractions above the tree show the number of clones belonging to the clade that yielded Cit⁺ mutants during replay experiments (numerator) and the corresponding total of clones used in those experiments (denominator). In the clade 3, a mutator phenotype emerged. The inset shows the number of mutations relative to the ancestor. The solid line is the least-square linear regression of mutations in non-mutator genomes; the dashed line is the corresponding regression for mutator genomes.

The genetic bases of the Cit⁺ trait evolution involved three successive processes: potentiation, actualization and refinement (Blount et al. 2012). Potentiation refers to the evolution of a genotype in which the emergence of the phenotypic innovation (here citrate consumption) is not yet realized but is possible by further mutation(s). Actualization consists

in the emergence of a weak Cit⁺ phenotype (slow growth on citrate) by 31,500 generations. Refinement of the new function involves additional genetic changes that allow the efficient exploitation of citrate and the rise of the Cit⁺ lineage to numerical dominance.

The genetic bases of potentiation are difficult to study since it may involve several mutations that emerged at different timepoints during evolution. Actualization of the Cit⁺ trait arose from an amplification-mediated promoter capture (Figure 20), whereby the initially-silent *citT* gene, encoding a C4-di- and tri-carboxylic acid transporter, is placed under the control of the *rnk* promoter which is active in aerobic conditions (Blount et al. 2012). The amplification mutational event generated a new junction fragment by joining upstream *rnk* and downstream *citG* fragments, producing an *rnk-citG* hybrid gene under the control of the *rnk* promoter. Because *citT* and *citG* are monocistronic, the downstream copy of *citT* is co-transcribed with the hybrid gene during aerobic metabolism and confers the Cit⁺ phenotype. Further refinement of this phenotype is explained by an increased number of *rnk-citT* modules. Indeed, later Cit⁺ genomes have four-copy tandem arrays.

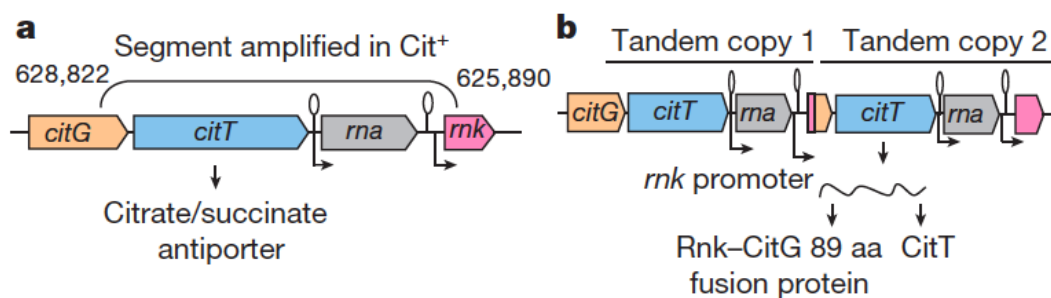


Figure 20. Tandem amplification in Cit⁺ genomes, after (Blount et al. 2012).

(A) Ancestral arrangement of *citG*, *citT*, *rna* and *rnk* genes. (B) Altered spatial and regulatory relationships generated by the amplification.

b) Emergence of a stable polymorphism by niche construction: the case of population Ara-2

The emergence of long-term stable diversity by niche construction in the LTEE was unexpected. Indeed, the medium contains only glucose as a single limiting carbon source and at very low concentrations (25 mg/L). The low concentration of glucose was expected to avoid the emergence of stable diversity for two reasons: first, low carbon availability may only support biomass formation without any carbon left for metabolic by-products. Second, the low carbon concentration may support low population sizes which reduces the possibility of emergence of differentiated clones constituting different lineages or of evolved clones with different beneficial mutations of similar fitness effects that would compete with each other (clonal interference). However, both niche construction and clonal interference have been reported as drivers of adaptation and diversity in the LTEE (Rozen and Lenski 2000, Maddamsetti, Lenski, and Barrick 2015).

The emergence of a stable polymorphism in the LTEE was first reported in one of the 12 LTEE lines, population Ara-2 (Rozen and Lenski 2000). At 18,000 generations, after plating the whole Ara-2 population, two colony types were identified. They were distinguished by their colony size and time of appearance on the plates (Figure 21A). The L lineage is characterized by clones forming large colonies after plating and incubation for 24h at 37°C, whereas clones from the S lineage form small colonies that appear on the surface of the plates only after ~48h of incubation at 37°C. These phenotypic traits are conserved after growing the different clones, indicating their heritable nature and genetic bases (Rozen and Lenski 2000).

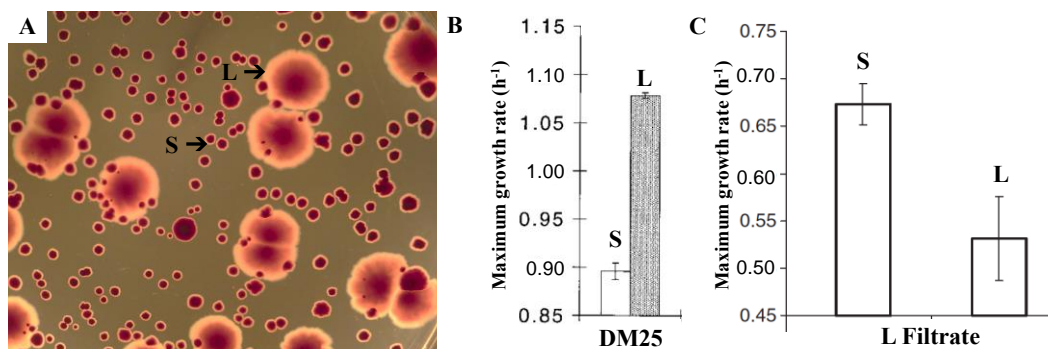


Figure 21. Phenotypes of the L and S lineages from population Ara-2.

(A) A 48-h culture of the Ara-2 population after plating. Notice the differences in colony size between the S and L lineages indicated by arrows. (B) Maximum growth rates of S and L in DM25, modified from (Rozen and Lenski 2000). (C) Maximum growth rate of S and L in filtered media, modified from (Rozen et al. 2009).

The L lineage is favored on the supplied glucose (Rozen and Lenski 2000), with a higher growth rate compared to the S lineage (Figure 21B). The role of niche construction as a driver of the emergence of the S and L lineages was confirmed by growth experiments in spent media (Figure 21C). Filtrates of a 24-h culture of one clone from each of the L and S lineages were prepared and tested to determine whether they could sustain growth of either lineage. Both S and especially L secrete metabolites that promote the growth of S, but L cannot efficiently use these metabolites, indicating that cross-feeding occurs specifically from L to S (Rozen and Lenski 2000, Rozen et al. 2009).

Based on phenotypic differences between L and S clones sampled over evolutionary time from population Ara-2 (colony size after plating), the time of emergence of the S lineage was established to be ~6500 generations (Figure 22A) and the S and L lineages co-exist until at least 50,000 generations (Le Gac et al. 2012, Tenailon et al. 2016).

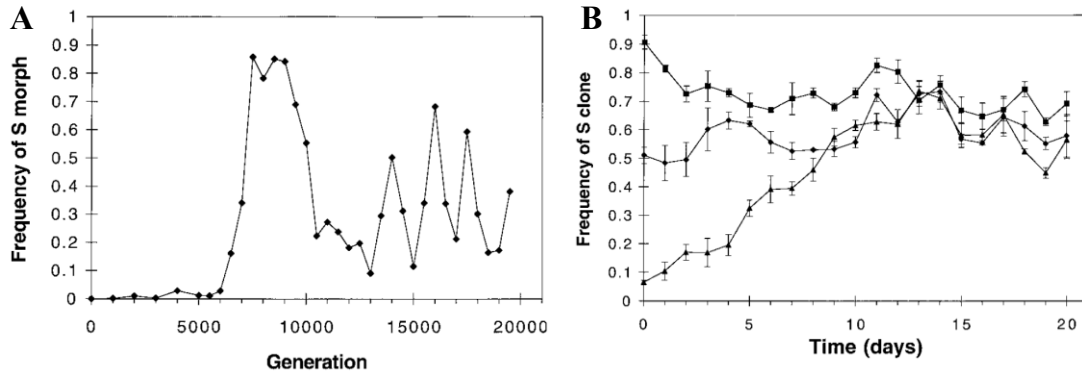


Figure 22. Dynamics of the S and L lineages, after (Rozen and Lenski 2000).

(A) Long-term dynamics of the polymorphism. Each point shows the frequency of the S morph, obtained by scoring several hundreds of evolved clones as S or L on the basis of their colony morphology after plating. (B) Convergence on a stable equilibrium over 20 serial transfer cycles (~130 generations). Genotype S is able to invade when rare, but it declines in frequency when it is initially common, leading to a balanced polymorphism.

Long-term maintenance of diversity is frequently attributed to negative frequency-dependent interactions (Friesen et al. 2004, Rainey et al. 2000, Blount, Borland, and Lenski 2008). When mixtures of S and L clones at different initial frequencies are propagated for ~130 generations, they are able to invade each other when initially rare, thereby converging to a stable equilibrium (Figure 22B). Moreover, direct competition experiments between S and L clones show that the initially rare ecotype is able to invade the other. Therefore, frequency-dependent interactions between S and L are strongly involved in the long-term maintenance of the polymorphism (Rozen and Lenski 2000). Indeed, the L and S lineages exploit different ecological niches in the sympatric culture. The L lineage has a large advantage over S during exponential growth on the supplied glucose. By contrast, the S lineage is favored in stationary phase owing to its preferential growth on the niche created by L that secrete by-products (Rozen and Lenski 2000). Furthermore, in mixed cultures of S and L, the L lineage has a high mortality in stationary phase that is even increased in the presence of S (Rozen et al. 2009).

These complex interactions are maintained over evolutionary time. The long-term maintenance of the S and L lineages was investigated by reciprocal invasion experiments at various initial frequencies using S and L clones sampled over time (Le Gac et al. 2012). When

rare, contemporary clones from each lineage had a significant advantage over clones from the other lineage. When the competitors came from different generations, both L and S clones from later generations could invade, when rare, earlier clones from the alternative lineage. However, only L clones from earlier generations are able to invade S clones from later generations. These results demonstrate the complex and dynamic co-existence of S and L: L encroach the ecological niches of S which is continuously escaping extinction, by unknown mechanisms. Therefore, the dynamics of this polymorphism involves co-evolution (Le Gac et al. 2012).

The ecological differences between S and L are associated with substantial reorganization of gene regulatory networks (Le Gac et al. 2012). Global expression profiles of S and L clones changed profoundly over evolution time. Comparison of the transcriptional profiles of L and S clones showed significant differences for 73 genes after 6500 generations, 263 after 17,000 generations and 618 after 40,000 generations. Interestingly, the expression of 176 genes significantly differed between the L and S lineages between 6500 and 40,000 generations. These genes are likely to be important for the ecological specialization of each lineage (Figure 23).

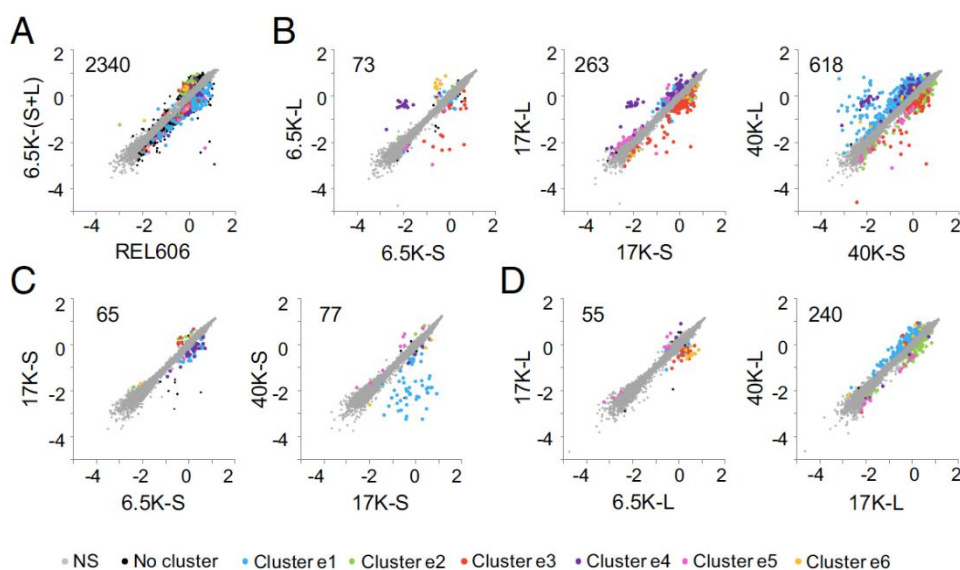


Figure 23. Evolutionary dynamics of global transcription profiles for L and S lineages during 40,000 generations, after (Le Gac et al. 2012).

(A) Comparison of the ancestral strain (REL606) with the combined S and L samples at 6500 generations [6.5K – (S + L)]. (B) Comparisons of S and L clones sampled at each of three timepoints: 6500, 17,000, and 40,000 generations. (C) Comparisons of S clones sampled over time. (D)

Comparisons of L clones sampled over time. Each dot corresponds to a gene, and the values are \log_{10} -transformed expression levels. The number of genes with significantly different expression is shown for each comparison at the upper left corner of the graphs. Gray symbols indicate genes without significant differences in expression in any of the eight comparisons. Those genes with significant differences in expression in at least one of the seven comparisons between two evolved samples were clustered according to their expression patterns, and the colored symbols indicate genes in expression clusters e1–e6. Black symbols indicate genes that either did not cluster or showed significant differences in expression only between the ancestor and evolved samples.

Genes involved in the Entner-Doudoroff (ED) metabolic pathway (Peekhaus and Conway 1998) and in the glyoxylate bypass (Chung, Klumpp, and Laporte 1988) show higher expression in the S lineage (cluster e3 in Figure 23, Le Gac et al. 2012). These differences may contribute to the lower growth rate of the S cells on glucose and to the ability of the S cells to exploit metabolic by-products secreted by L cells. The genes with higher expression in the L lineage (cluster e4 in Figure 23) are involved in glycerol metabolism. Higher expression is also observed in L clones for *manXYZ*, encoding a secondary glucose transporter. Its increased transcription may contribute to the faster growth of L on glucose (Le Gac et al. 2012).

The genome sequences of two clones sampled at 6,500 generations from each of the two S and L lineages (6.5KS1, 6.5KS2, 6.5KL4, and 6.5KL9) allowed to identify the genetic bases of the emergence of the S lineage (Plucain et al. 2014). These clones share 68 mutations that probably occurred before the divergence of the S and L lineages. Fifty-five mutations are specific to the S clones and 36 to the L clones. Two of the 55 S-specific mutations affected genes that can explain the changes in the global transcription profiles that were observed in the S clones over evolution. Indeed, mutations were identified in the *arcA* and *gntR* genes, which encode regulators of the TCA cycle and the ED pathway, respectively (Plucain et al. 2014). The *arcA* mutation arose at ~6000 generations and the *gntR* mutation at 6500 generations. Both have been fixed in the S lineage and are still present at 50,000 generations. The replacement of either *arcA* or *gntR* evolved alleles in the 6.5KS evolved background by its ancestral counterpart abolishes the negative frequency-dependent interactions with L indicating that both alleles contributed to the emergence of the S lineage. It was further shown that the S-specific traits

(growth on L filtrate and negative frequency-dependent interactions with L) needed a third mutation, in addition to the *arcA* and *gntR* mutations (Plucain et al. 2014). This third mutation occurred in *spoT* which is involved in the metabolism of ppGpp the effector of the stringent response to starvation (Hernandez and Bremer 1991). The *spoT* mutation appeared in the Ara-2 population before the diversification event and if thus shared by S and L. *spoT* mutations are known to be beneficial in the LTEE conditions (Cooper, Rozen, and Lenski 2003b, Pelosi et al. 2006). Combining the three mutations in *spoT*, *arcA* and *gntR* in the ancestral background mimicked the S-specific phenotypes (Figure 24). The emergence of the S lineage corresponds to a multistep process, where the successive mutations in global regulatory genes allow the establishment of the polymorphism, most likely by restructuring the regulatory networks. Moreover, even if other replicate populations fixed mutations in the same genes, none has evolved such long-term polymorphisms, implying that specific alleles produced qualitatively different evolutionary dynamics (Plucain et al. 2014).

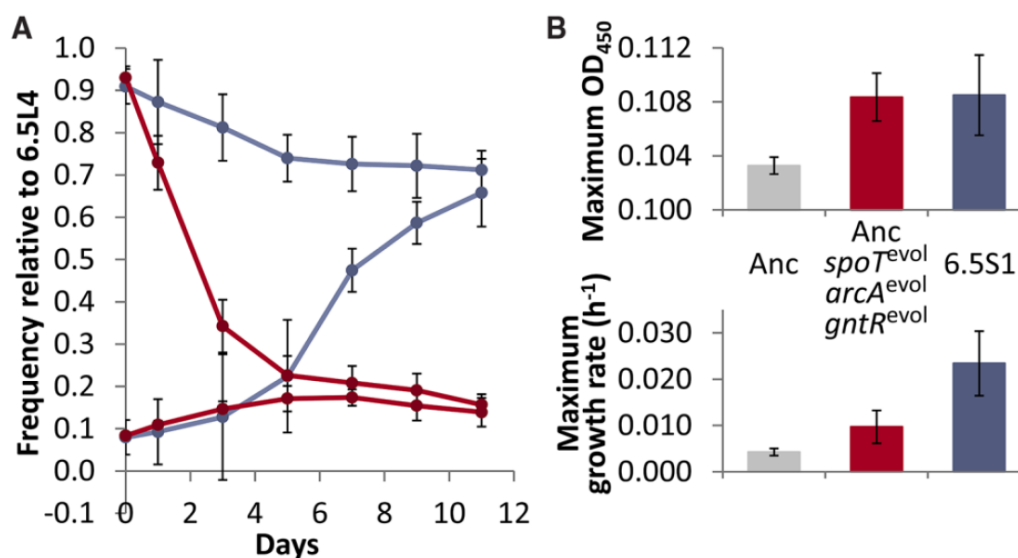


Figure 24. Phenotypic traits conferred by the three mutations in *spoT*, *arcA* and *gntR*, after (Plucain et al. 2014).

The three mutations in *spoT*, *arcA* and *gntR* were combined in the ancestral background, resulting in a constructed S ecotype. (A) Invasion and coexistence of the constructed S ecotype (red) and clone 6.5KS1 (blue) with clone 6.5KL4, starting from different initial frequencies. (B) Maximum optical density (OD₄₅₀) and growth rate of the ancestral strain (gray), constructed S ecotype (red) and 6.5KS1 (blue) in the supernatant obtained from a 6.5KL4 culture.

Despite all the work performed to understand the mechanisms underlying the Ara-2 polymorphism since its first description in 2000 by Rozen et al., several aspects related to the deep ecological and molecular interactions that allowed its emergence and long-term co-existence remain unclear. For instance, the metabolic and molecular bases of emergence, the identification of the L metabolic by-product, and the effect of important ecological parameters such as tradeoff, seasonality and ecological character displacement are still mostly unknown. The objective of the present doctoral work is the complete characterization of the ecological, physiological and molecular parameters that allowed the emergence and stable co-existence of the S and L lineages in population Ara-2. We used a multidisciplinary approach that includes:

- i. *in vivo* experimental analyses of S and L lineages (Results, Chapters 1 and 2).
At the molecular level, we studied the particular case of the global regulator ArcA by an experimental approach (Results, Chapter 3). To characterize all aspects involved in adaptation, I investigated the long-term dynamics of mutation rates in Ara-2 (Results, Chapter 4).
- ii. Implementation of a dynamic evolutionary framework that allowed the *in silico* study of sympatric adaptive diversification in bacteria, based on metabolic trade-offs using Flux Balance Analyses (Results, Chapter 1).
- iii. An *in silico* model intending to simulate the ecological bases of sympatric adaptive diversification (Chapter 5).

PART II: RESULTS

**Chapter 1: Metabolic Modelling in a Dynamic
Evolutionary Framework Predicts Adaptive
Diversification of Bacteria in a Long-Term
Evolution Experiment**

The first part of my doctoral research focuses in the understanding of the ecological traits of S and L emergence and co-existence. The data presented in this first chapter of the Results section resulted from a close collaboration we established with Professor Orkun S. Soyer and Dr. Tobias Großkopf from the Warwick Centre for Integrative Synthetic Biology of the University of Warwick, United Kingdom. The main objective of this collaboration was the analysis of the ecological dynamics of the emergence and long-term co-existence of the S and L lineages of population Ara-2. The experimental results I obtained during my PhD allowed to identify the main parameters needed for the implementation of a dynamic evolutionary framework that allowed the *in silico* study of sympatric adaptive diversification in bacteria. The implemented model is based on metabolic tradeoffs using Flux Balance Analyses (evoFBA). This chapter is presented as a paper entitled, “*Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a Long-Term Evolution Experiment*” (Großkopf T. *, Consuegra J. *, Gaffé J., Willison J.C., Lenski R.E., Soyer O.S, Schneider D.) which was published in *BMC Evolutionary Biology*, 2016, 16:163.

Initial studies on S and L lineages of population Ara-2 established that the maintenance of the polymorphism involved cross-feeding interactions. Indeed, using 24-h culture filtrates of L clones, it was shown that L secreted by-products that S cells could better exploit in terms of both growth rate and yield (Rozen and Lenski 2000, Rozen et al. 2009). However, the by-product secreted by L and the ecological and metabolic mechanisms of S and L emergence and co-existence were still unknown. Therefore, the mechanism of emergence of adaptive diversification in Ara-2 in the LTEE was incompletely understood.

The limiting low concentration of glucose in the evolution medium (25 mg/L, DM25 medium) was expected to support only biomass formation without any carbon left for metabolic by-products. With the evidence of cross-feeding, owing to the filtrate experiments, we

hypothesized an overflow of the metabolism with subsequent production of 3-carbon by-products by the L lineage despite the low glucose concentration.

We produced large quantities of filtrates from cultures of an L clone in DM25 and analyzed them by high-performance liquid chromatography (HPLC), in collaboration with John C. Willison (Laboratoire Chimie et Biologie des Métaux CEA – Grenoble). Unfortunately, we were unable to detect any metabolite in these filtrates. We then grew an L clone in DM media containing 10 times more glucose (DM250), produced filtrates, and analyzed them by HPLC. We detected 480 μM of acetate in these filtrates. However, we still had to confirm that acetate was produced during the LTEE in DM25. To confirm this, we used Gas Chromatography Mass Spectrometry (GC-MS) of volatile compounds. This method is more sensitive than HPLC which may allow us to detect lower concentrations of acetate. The presence of acetate in the L filtrate after growth in both DM25 and DM250 was confirmed by GC-MS. The acetate concentration from the DM25 filtrate was estimated at 51 μM , which was approximately 10 times lower than the acetate concentration detected by HPLC in DM250 filtrates. Knowing that the amount of acetate produced by L was proportional to the concentration of glucose in the DM media, we used DM250 for further experiments in order to improve the accuracy of measurements. To confirm that acetate was indeed the metabolite secreted by L that allowed S to grow on L filtrates, we analyzed L filtrates before and after S grow. Acetate was not detected anymore after growing S in that supernatant (Supp Fig. 1 of the following paper), confirming that acetate was indeed the L-secreted metabolite and was involved in the emergence of the S/L polymorphism.

Since the 12 populations of the LTEE experience growth cycles of 24 hours before being transferred in fresh medium, we next analyzed the 24-h kinetics of glucose and acetate production and consumption in allopatric cultures of S and L (Fig. 3B of the following paper). Both L and S clones consumed glucose faster than the ancestor. We showed that acetate was

secreted by all strains (the ancestor and both S and L clones), while only S clones consumed it efficiently and exhibited a faster metabolic switch from glucose to acetate. These results support previous results on L filtrates and the hypothesis that the stable coexistence of S and L depends on acetate cross-feeding.

The secretion of acetate by L and its improved consumption by S showed that presence of acetate was an important selective pressure during the evolution of the S lineage and its coexistence with L. If the previous hypothesis is true, we expected a progressive improvement of acetate consumption by S through evolutionary time. We therefore measured the growth rates of S and L in DM media with either acetate or glucose at seven different generational time points over evolution (from the time of S emergence at 6500 generations to 50,000 generations). Not only S progressively improved on acetate but also revealed declined abilities to grow on glucose. Opposite trends were observed in the L lineage, with improvement on glucose and decline on acetate (Fig. 5 of the following paper). These evolutionary growth trajectories on acetate and glucose indicated character displacement and suggested tradeoffs that prevented the simultaneous optimization of growth on both carbon sources by either of the two lineages.

One important objective of this work was to determine the metabolic mechanism of S and L emergence and co-existence. In order to reach this objective, we developed a metabolic model based on a dynamic evolutionary framework (evoFBA) which allowed us the study of sympatric adaptive diversification. Model simulations predicted the adaptive diversification that occurred in Ara-2 and generated hypotheses about the mechanisms that promoted the co-existence of the diverged lineages. These predictions allowed us to target the experiments that were performed and described above.

First, evoFBA predicted the niche construction promoted by L cells owing to fast consumption and partial oxidation of glucose, resulting in acetate production as a by-product.

This partial oxidation of glucose was predicted to be due to a strong tradeoff between uptake reactions. Hence, it predicted the emergence of both glucose and acetate specialists (L- and S-like, respectively). Second, it predicted metabolic fluxes for the two model organisms. On glucose, both the glucose and acetate specialists displayed similar behaviors, using the TCA cycle only partially and the glyoxylate shunt not at all. After switching to acetate consumption (which the L-like glucose specialist was unable to do), the acetate specialists showed very different fluxes, with reverse glycolysis and full use of the TCA cycle including the glyoxylate shunt. We tested these predictions by measuring, the promoter activities of genes encoding *aceB*, *acnB*, *ackA* and *pgi* using transcriptional fusions with the *gfp* reporter gene. Both S and L clones showed moderately increased promoter activity for *pgi* relative to the ancestor, and large increases in the promoter activities of *acnB* and *aceB* relative to the ancestor, with the S clones showing much greater increases than the L clones. This was consistent with the possibility of greater flux through the TCA cycle and glyoxylate shunt as predicted in the S-like model organism (Fig. 6 of the following paper).

The combination of evoFBA modelling and experimental evolution provided us with a powerful approach that gave us insights into the ecological and metabolic mechanisms involved in the emergence and long-term co-existence of the S and L lineages.

RESEARCH ARTICLE

Open Access



Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment

Tobias Großkopf¹, Jessika Consuegra^{2,3}, Joël Gaffé^{2,3}, John C. Willison^{4,5,6}, Richard E. Lenski^{7,8}, Orkun S. Soyer^{1*} and Dominique Schneider^{2,3*}

Abstract

Background: Predicting adaptive trajectories is a major goal of evolutionary biology and useful for practical applications. Systems biology has enabled the development of genome-scale metabolic models. However, analysing these models via flux balance analysis (FBA) cannot predict many evolutionary outcomes including adaptive diversification, whereby an ancestral lineage diverges to fill multiple niches. Here we combine *in silico* evolution with FBA and apply this modelling framework, evoFBA, to a long-term evolution experiment with *Escherichia coli*.

Results: Simulations predicted the adaptive diversification that occurred in one experimental population and generated hypotheses about the mechanisms that promoted coexistence of the diverged lineages. We experimentally tested and, on balance, verified these mechanisms, showing that diversification involved niche construction and character displacement through differential nutrient uptake and altered metabolic regulation.

Conclusion: The evoFBA framework represents a promising new way to model biochemical evolution, one that can generate testable predictions about evolutionary and ecosystem-level outcomes.

Keywords: Adaptive diversification, Experimental evolution, FBA, *In silico* evolution, Tradeoffs

Background

The ability to predict evolution would be valuable not only for understanding such processes as adaptation and speciation [1–3], but also for engineering robust industrial strains, anticipating ecosystem responses to climate change, and combatting antibiotic resistance [4–7]. Models that capture the relationship between genotypes and environments, the structure and state of regulatory and metabolic networks, and the resulting phenotypes are likely to be important for developing these predictive abilities [1, 3, 8].

Ultimately, models of the relationship between genotype and phenotype will need to be combined with models of evolutionary and ecological dynamics in integrated frameworks that can predict the trajectory of evolution [5, 9].

The dynamics of evolutionary change reflect multiple processes and varying selective pressures that are influenced by many ecological, physical, and cellular constraints that may conflict with one another. Understanding whether and how these dynamics lead to the splitting and divergence of lineages is of central interest, as these processes represent the initial steps towards speciation. To this end, several theoretical studies have shown that cellular tradeoffs can promote lineage divergence [10–16]. The importance of such tradeoffs can be readily understood in the context of metabolism and growth. For example, if there were no tradeoffs, then one would predict that cells should maximize their expression of transporters and their surface area to achieve the highest

* Correspondence:

o.soyer@warwick.ac.uk; dominique.schneider@ujf-grenoble.fr

Tobias Großkopf and Jessika Consuegra are shared first authors.

¹School of Life Sciences, University of Warwick, Coventry, UK

²University of Grenoble Alpes, Laboratoire Techniques de l'Ingénierie

Médicale et de la Complexité - Informatique, Mathématiques et Applications, Grenoble (TIMC-IMAG), F-38000 Grenoble, France

Full list of author information is available at the end of the article



possible rate of substrate uptake [17]. However, such cellular investments would impinge on other cellular processes owing to competing requirements for membrane and cytosol space [18, 19], ribosomes [15, 16], and redox carriers [20, 21]. Thus, cells may appear suboptimal for individual physiological parameters, but this might be merely a consequence of being optimal for the combined set of parameters and associated cellular tradeoffs.

Historically, the interplay between cellular tradeoffs and evolutionary and ecological dynamics has been analyzed using game theory and differential equation-based models that consider small or idealized metabolic systems [10, 11, 14, 22]. These studies have highlighted that tradeoffs in cellular metabolism can lead to incomplete degradation of a resource, resulting in the evolution of cross-feeding interactions [10, 11]. This phenomenon has been seen in several evolution experiments under both batch and chemostat conditions [23–27]. To increase predictive power in microbial ecology and evolution, it is now desirable to develop models that can take into account cellular metabolism at a larger scale and across different organisms. Stoichiometric models offer a promising approach because, in principle, they can capture all enzyme-mediated metabolic reactions of an organism in an unbiased and non-supervised way using genomic information [8].

Flux Balance Analysis (FBA) has been developed to determine the optimal metabolic state of an organism, given knowledge of its biochemical network, biomass composition, and uptake flux rates [28]. This approach is based on the assumptions that evolution has optimized metabolism and that metabolic fluxes can be predicted by setting the growth rate for a given rate of substrate uptake (such that the ratio of the two rates represents a yield) as an optimization criterion that can be solved by linear programming [28–30]. Early applications of FBA ignored the essential role of tradeoffs in the computation of metabolic fluxes [28, 31, 32], but more recent applications have incorporated tradeoffs as constraints on total fluxes [18, 19, 33, 34] and thereby achieved better prediction of experimentally observed metabolic states, such as preferential substrate utilization [19] and acetate overflow [18]. Experimentally measured reaction thermodynamics and gene expression levels have also been used to constrain optimal metabolic states that reflect tradeoffs [35–37], and there have been efforts to combine FBA with ecological interactions between multiple species in microbial communities [38–45]. These approaches use species-specific models in a shared environment to maximize a predefined, community-level objective [39, 41, 43, 44] or apply FBA within a dynamic framework [46]. The latter approach enables prediction of ecological interactions such as competition and cross-feeding between different species making up the model

community, given defined substrate uptake constraints for each model species [40, 42, 45]. However, none of these approaches can currently be used to predict the interplay between ecological and evolutionary dynamics.

Here, we begin to overcome these limitations by integrating a FBA model of multi-phenotype systems with both cellular constraints and evolutionary dynamics. We define an overall constraint on uptake rates to enforce tradeoffs while simulating multiple model organisms living in the same environment without the need to specify each organism's uptake preferences a priori (for details on how evolution and mutations are simulated see Methods section). By limiting total uptake in the model, and including O₂ “uptake” in that total, we seek to represent cellular limitations that can arise from many diverse processes, including redox cycling [20, 47], respiratory chain [18], enzyme expression [16, 48], and substrate uptake [17]. Although O₂ uptake *per se* might not be limiting, limitations in the electron transport chain can effectively limit O₂ respiration. Accounting for all the different possible limitations arising from cellular processes in a mechanistic manner is beyond the scope of stoichiometric models; however, limiting total uptake provides a general constraint that allows us to implement the tradeoffs observed in different studies in a simple, consistent, albeit approximate manner [16–18, 20, 47].

This approach allows integration of evolutionary dynamics by mutations that change substrate uptake rates along with the optimization of each model organism in the context of other model organisms that are present and coevolving in the same environment. The combined framework, which we call evoFBA, thus aims to provide a more realistic way to model the interplay between ecological and evolutionary dynamics with global constraints arising from cellular tradeoffs. To the best of our knowledge, this is the first FBA modeling approach that captures the continuous adaptation of organisms to the interplay between ecological and evolutionary dynamics in systems with multiple strains or species.

To examine the ability of evoFBA to capture ecological and evolutionary dynamics, we used it to simulate the evolution of *Escherichia coli* populations in a defined glucose-limited environment with daily transfers. We then experimentally analyzed the predictions of evoFBA in the context of the long-term evolution experiment (LTEE) with *E. coli*, in which 12 populations started from a common ancestor have been propagated in a glucose-limited medium for more than 60,000 generations [2, 49]. We found that the evoFBA simulations predicted the emergence of cross-feeding model organisms as a stable end-point, which in fact has occurred in at least one of the LTEE populations [26, 50]. Moreover, we saw that key metabolic features of the model organisms were in qualitative agreement with the physiological properties we

measured for the two biological lineages that emerged and subsequently coexisted for more than 50,000 generations.

Results

Microbial communities and their underlying metabolic interactions reflect the ecological and evolutionary histories of the component species [51]. To capture these interactions, we combine stoichiometric metabolic models with ecological and evolutionary dynamics in the multi-layered evoFBA framework (see Methods). To test the utility of this framework, we apply it to the LTEE in which *E. coli* populations evolve in a defined glucose-limited environment [2, 52].

To model the LTEE, we ran evoFBA simulations starting with a metabolic model of *E. coli* that accounts for 14 carbon sources including glucose and byproducts that can be scavenged from the environment to produce biomass and fuel associated core metabolic reactions. In each evoFBA simulation, we allowed the metabolic model to

change by random mutations under global constraints that must be obeyed. Thus, each simulation produced mutant model organisms exhibiting different uptake rates, metabolic flux patterns, and resulting growth rates.

evoFBA predicts evolution of cross-feeding between lineages with different metabolic flux distributions

Starting from a population of identical model organisms under conditions similar to the LTEE, a typical evoFBA simulation produced through random mutations more than 90,000 genetically distinct model organisms over 550 simulated daily transfer cycles (Fig. 1). The evolutionary dynamics across replicate simulations were highly reproducible in their key features, in particular the diversification of the population into two coexisting lineages (Fig. 2). Thus, throughout the paper, we will focus on results from a typical representative simulation that resulted in 97,912 different model genotypes, of which 3943 survived at least one transfer event (Fig. 1a) and 12 reached a population

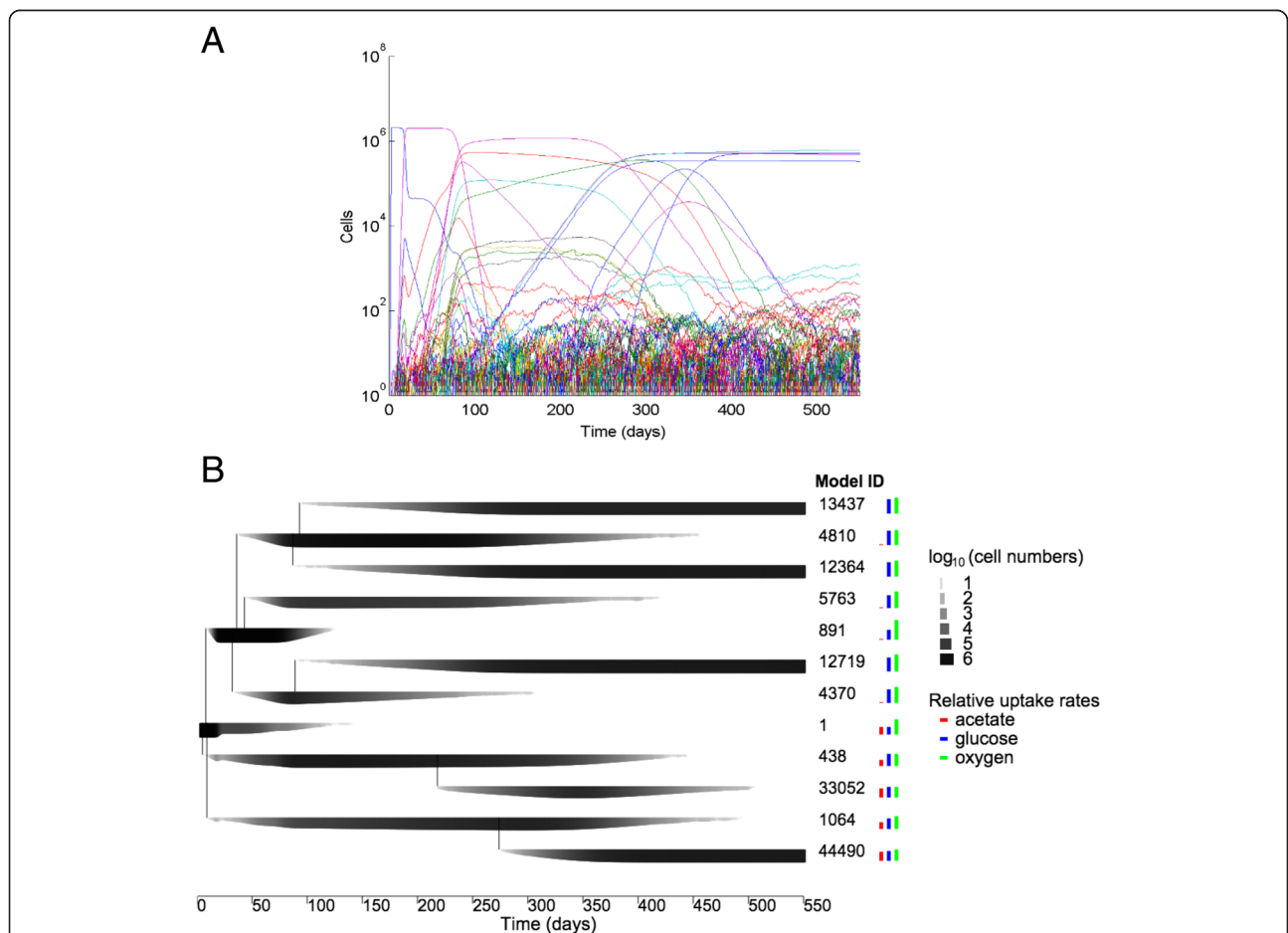
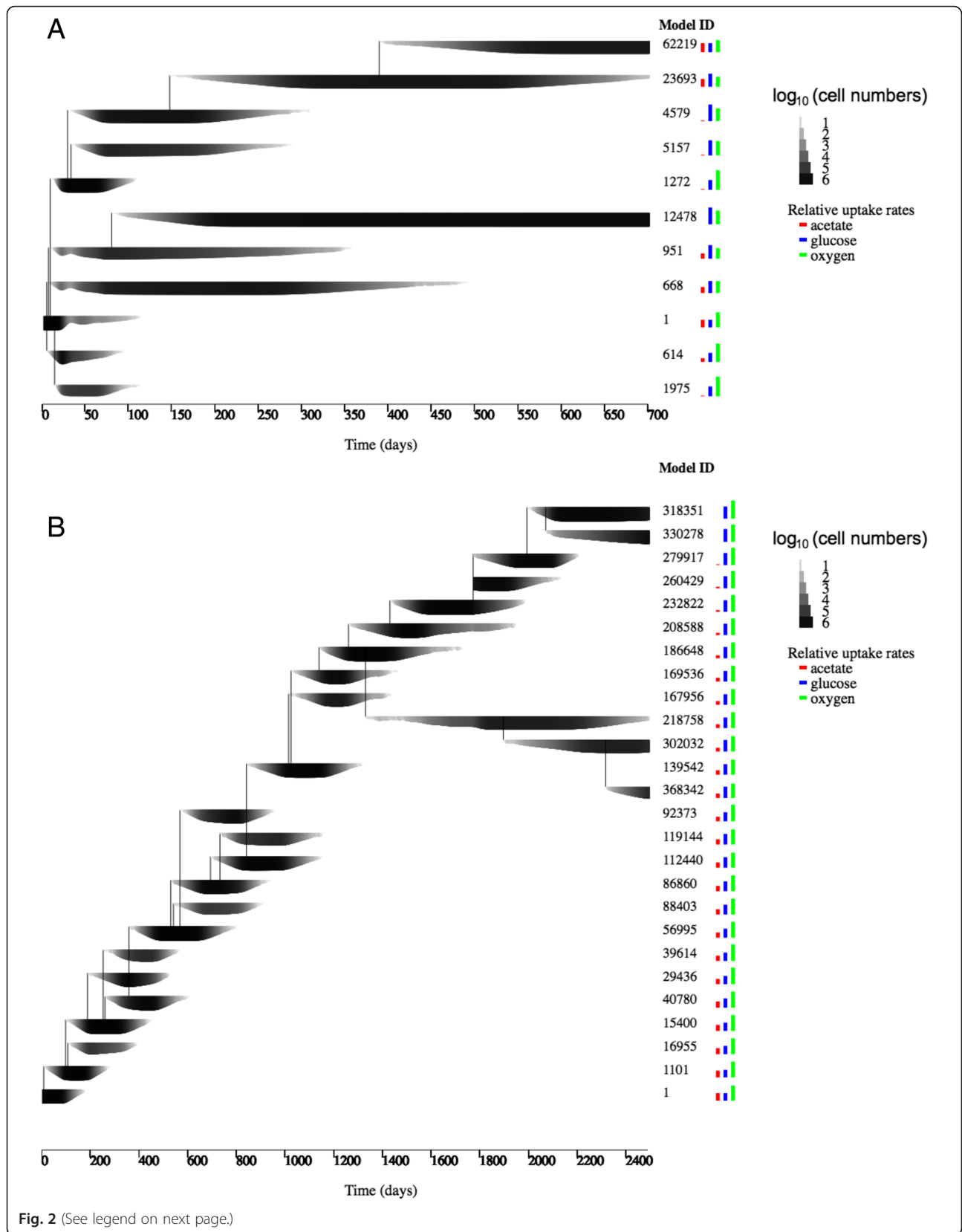


Fig. 1 Evolutionary dynamics *in silico*. **a** Numbers of surviving cells (i.e., post dilution) after each simulated cycle on a logarithmic scale. Each curve shows one of the 3943 model organism genotypes that survived at least one cycle (see text). **b** Relationships among ancestral and mutant model genotypes for those that reached a population of at least 10⁵ cells at any point during the simulation (see Methods). Model ID indicates the identifier assigned to each model genotype, with 1 being the ancestor. Line thickness is proportional to the log₁₀-transformed number per 10-ml volume at the start of each cycle. Coloured bars show relative uptake rates for glucose (blue), acetate (red), and oxygen (green)



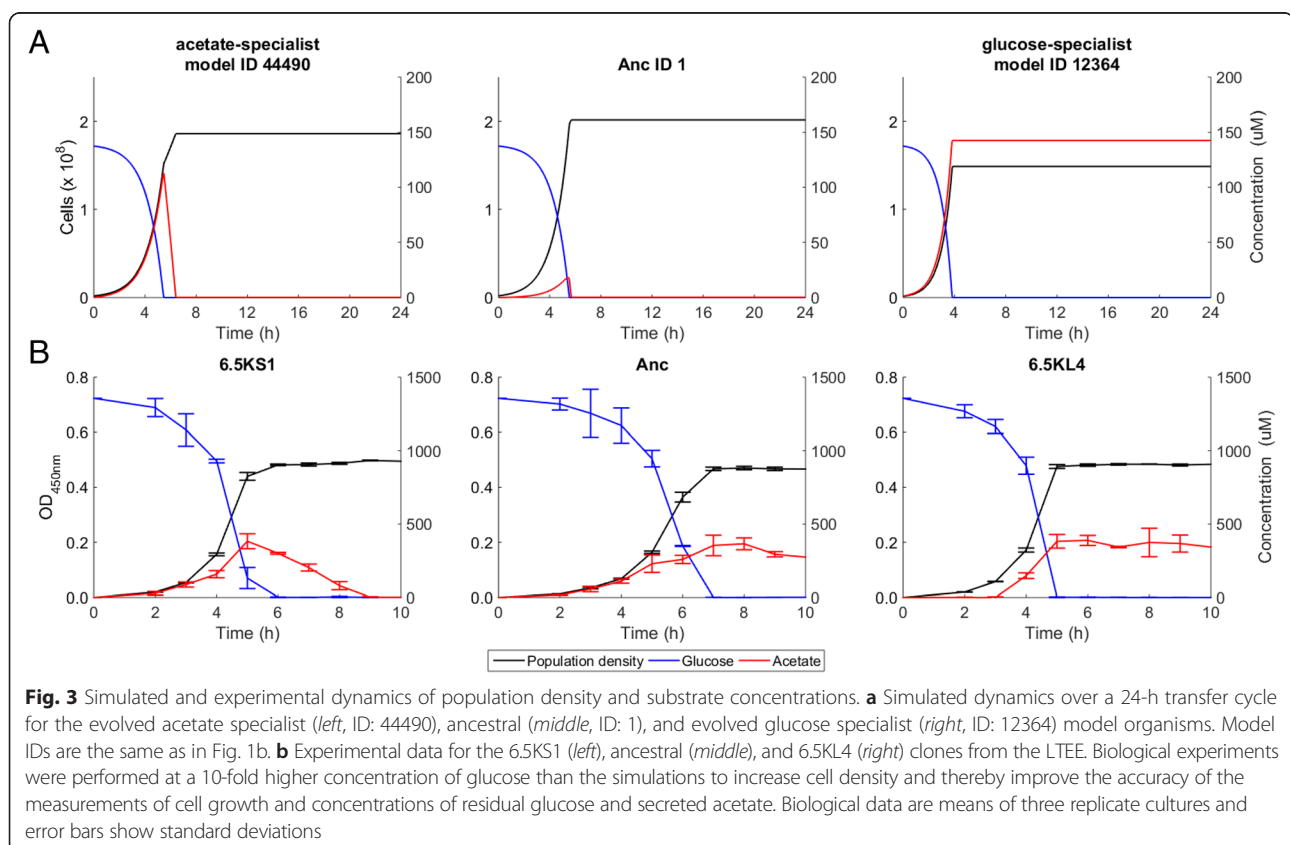
(See figure on previous page.)

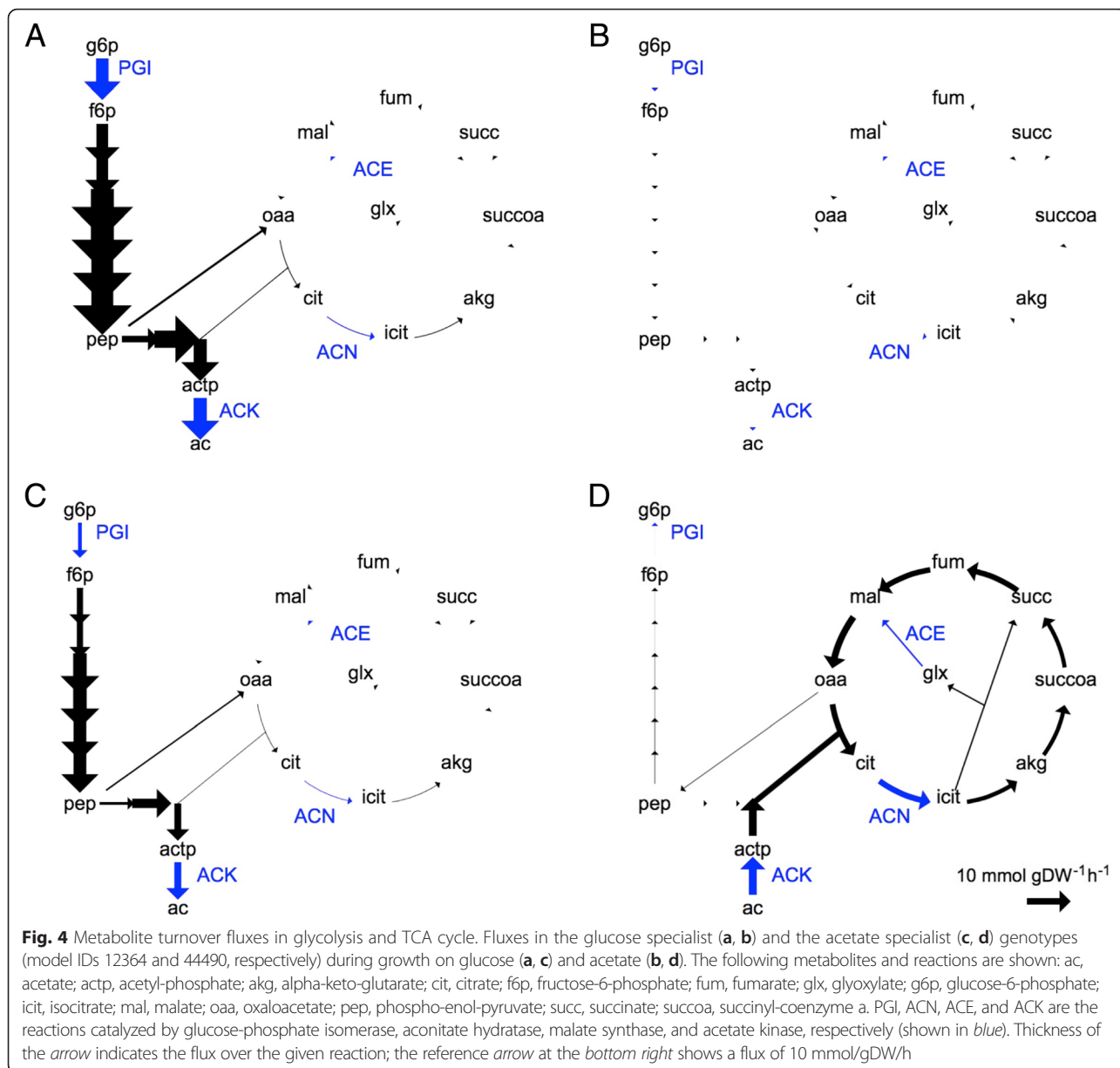
Fig. 2 Replicate runs of evoFBA. **a** One of five replicate simulations using the same parameter set as described in the main text and shown in Fig. 1. All simulations led to qualitatively similar outcomes. **b** Running evoFBA simulations with a smaller maximum mutation step size (± 1 mmol/gDW/h), see Methods eq. 5), led to the same diversification into glucose specialist and glucose-acetate co-utilizing model organisms, although the time required to achieve the diversification was substantially longer. Model ID, line thickness and coloured bars are the same as in Fig. 1

size of at least 10^5 cells at some point (Fig. 1b). These simulations revealed specific changes in oxygen, glucose, and acetate uptake by the model organisms (Fig. 1b). Glucose uptake and incomplete oxidation resulted in acetate secretion by the ancestral model organism, which would then switch to acetate uptake and oxidation after the glucose was exhausted. Thus, the ancestral model displayed a diauxic shift (Fig. 3a), as observed in *E. coli* [21]. As the *in silico* evolution proceeded, new model organisms arose that had increased glucose uptake and acetate production. The resulting increase in acetate concentration generated an ecological niche that was colonized by other model organisms with increased acetate uptake but reduced glucose uptake. After ~300 simulated daily transfer cycles (~2000 generations), the simulated evolution came to a halt, with no mutant model organisms able to replace the dominant ones. Thus, the *in silico* dynamics produced two distinct lineages that specialized on glucose and acetate,

respectively. The glucose-specialist model organisms lost the ability to consume acetate, whereas the acetate-specialist model organisms retained the ability to consume glucose but at a lower rate, and the timing of their diauxic shift was changed (Fig. 3a). As a consequence, the simulation led to a stable cross-feeding relationship between two lineages of model organisms.

We then examined the metabolic fluxes for the two model organisms when growing on glucose and acetate (Fig. 4). On glucose, both the glucose and acetate specialists displayed similar behaviours, using the TCA cycle only partially and the glyoxylate shunt not at all (Fig. 4a and c). After switching to acetate consumption (which the glucose specialists could not do), the acetate specialists showed very different fluxes, with reverse glycolysis and full use of the TCA cycle including the glyoxylate shunt (Fig. 4b and d). We emphasize that the emergence of cross-feeding model organisms and their





associated fluxes in the evoFBA simulation represents an idealized evolutionary stable state given the assumptions of the evoFBA framework.

Adaptive diversification in one LTEE population, matching evoFBA predictions

Two distinct lineages had emerged in one of the LTEE populations, called Ara-2, by 6500 generations, and they have coexisted ever since [26, 50]. The lineages are called S (small) and L (large) after their colony sizes on agar plates. The maintenance of this polymorphism depends on a cross-feeding interaction in which the L type is a better competitor for the exogenously supplied glucose and the S type is better at using one or more

secreted byproducts [26], although the precise ecological and metabolic mechanisms are still unknown. Therefore, we used predictions from the evoFBA simulations to generate hypotheses about these mechanisms.

We hypothesized that, first, L specializes on glucose and secretes acetate and, second, S specializes by improved acetate consumption. We tested this hypothesis by analyzing two evolved clones sampled at generation 6500 from the S and L lineages, named 6.5KS1 and 6.5KL4, respectively. HPLC analyses confirmed the presence of acetate in a 24-h supernatant of 6.5KL4 that was grown in the same medium as the LTEE (see Methods). Acetate was not detected after growing 6.5KS1 in that supernatant (Additional file 1: Figure S1). We then

measured the acetate and glucose concentrations over time in cultures of the ancestor, 6.5KS1, and 6.5KL4 clones in DM250-glucose medium (Fig. 3b). Both the L and S clones consumed glucose faster than the ancestor, consistent with previous assays [53]. Moreover, in agreement with the evoFBA results, 6.5KL4 secreted acetate, with its concentration remaining high for many hours in the monoculture, and 6.5KS1 drew down its own acetate secretion much faster than both 6.5KL4 and the ancestor. After exhausting the glucose by 6 h, 6.5KS1 showed diauxic growth and consumed acetate until it was depleted after 9 h, whereas 6.5KL4 had barely, if at all, begun to consume acetate at that time even as it had exhausted the glucose by 5 h (Fig. 3b). These results support the hypothesis that the stable coexistence of S and L depends on acetate cross-feeding, with acetate production by both the L and S lineages and more efficient acetate scavenging by the S lineage, which exhibits a faster metabolic switch from glucose to acetate (Additional file 2: Figure S2).

Physiology and fluxes in S and L clones agree qualitatively with evoFBA

The evoFBA simulation reaches an evolutionary equilibrium, whereas the interaction between the S and L lineages remained highly dynamic over thousands of generations [26]. Therefore, we examined the metabolic divergence of the S and L lineages over the course of the LTEE. We first measured the ability of clones from earlier and later generations to grow in minimal media containing glucose or acetate. S clones from later generations typically grew faster and with a shorter lag phase on acetate and more slowly on glucose than S clones from earlier generations, while the opposite trends were observed in the L lineage (Additional file 3: Figure S3) (in line with previous observations [53]). Compared to the ancestor, S clones improved their growth on acetate over evolutionary time, while L clones initially improved somewhat but were variable, with the 50,000-generation L clone showing weak growth similar to the ancestor (Fig. 5). On glucose, the opposite trend was observed with L clones consistently improving compared to the ancestor, while S clones improved initially but declined in later generations (Fig. 5). These patterns of growth relative to the ancestor are consistent with previous assays using the LTEE clones [53, 54]. These evolutionary trajectories of growth on acetate and glucose indicate character displacement and suggest tradeoffs that prevent the simultaneous optimization of growth on both carbon sources. The trajectories are qualitatively consistent with the evoFBA simulations, although the evoFBA predicts complete specialization on glucose without any acetate consumption. This evoFBA prediction represents a potential evolutionarily stable end point, which might

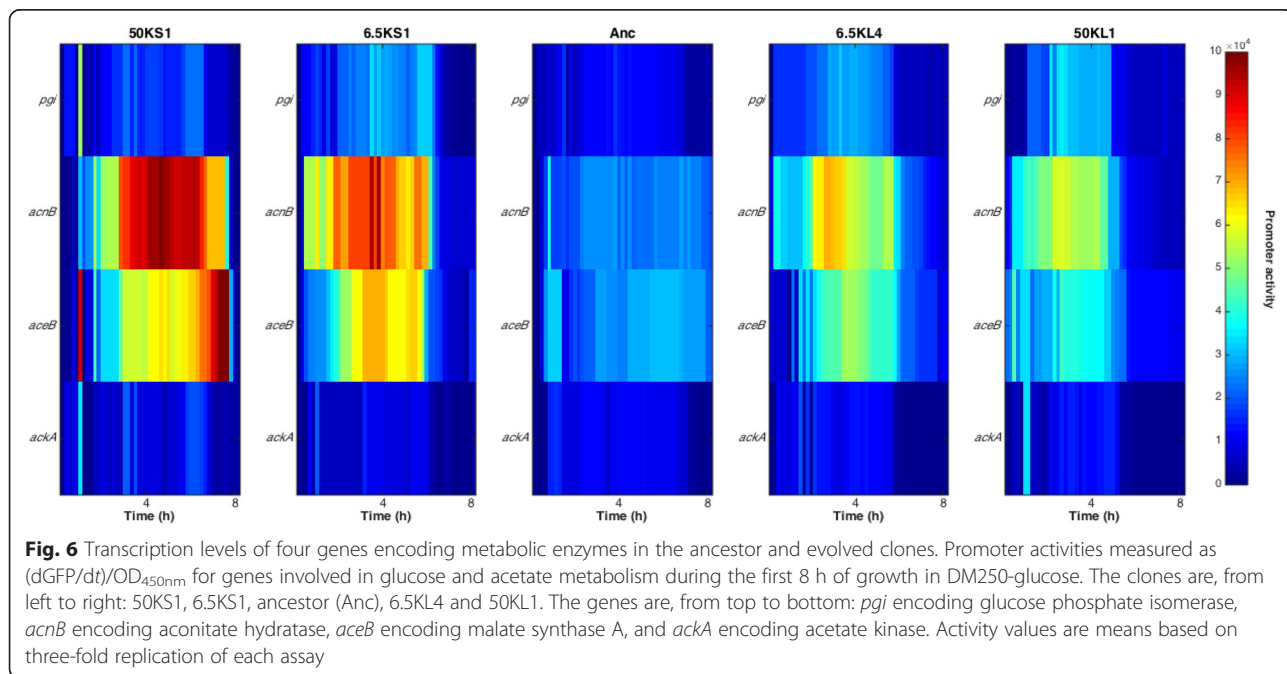
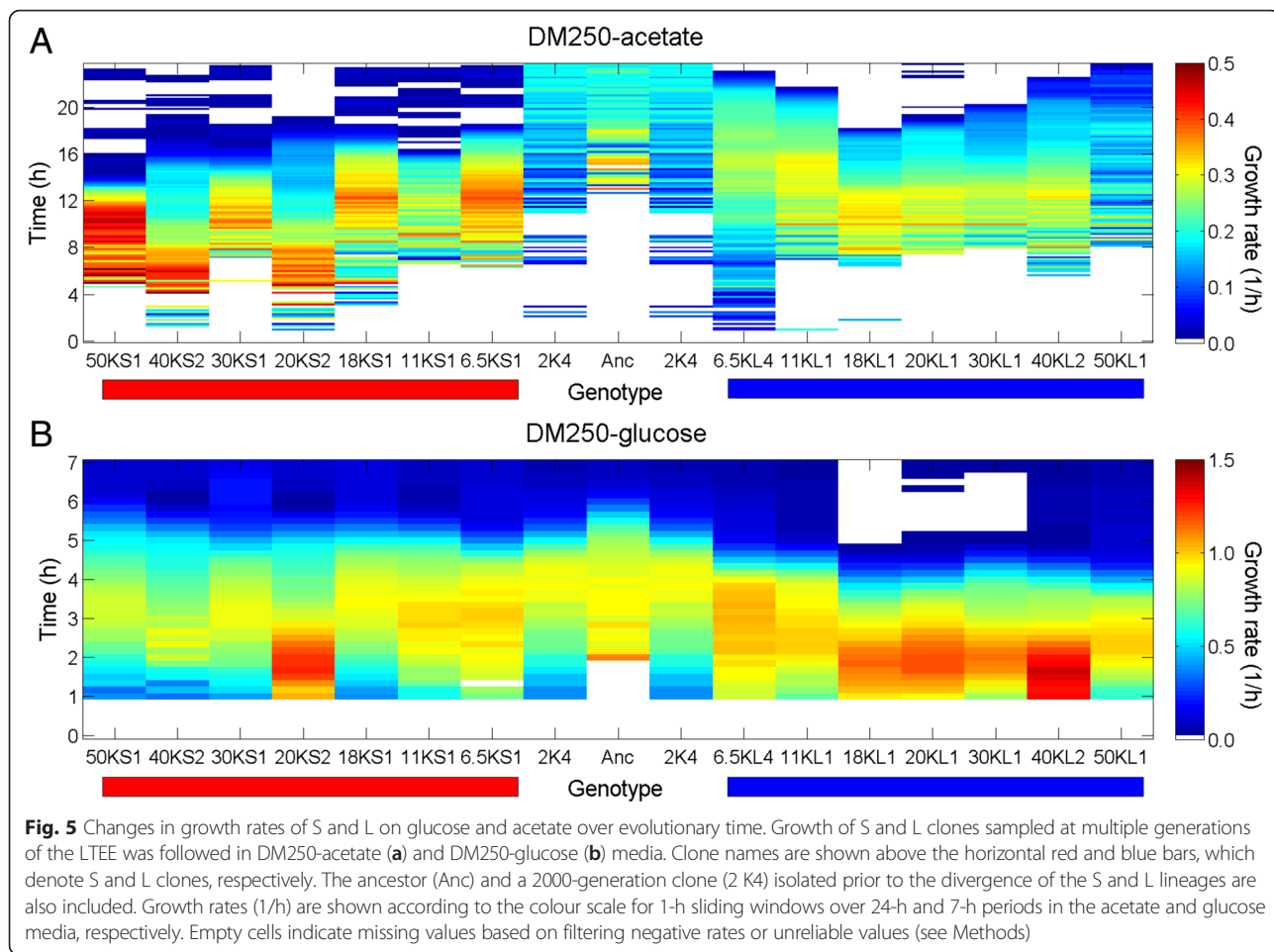
eventually occur in the S and L lineages after more generations.

We then tested the flux patterns predicted by evoFBA (Fig. 4) by measuring, in several LTEE clones, the promoter activities of genes encoding four key metabolic enzymes, using transcriptional fusions with the *gfp* reporter gene (see Methods). Both S and L clones showed moderately increased promoter activity for *pgi* relative to the ancestor (Fig. 6). Both S and L clones exhibited larger increases in the promoter activities of *acnB* and *aceB* relative to the ancestor, with the S clones showing much greater increases than the L clones, consistent with the possibility of greater flux through the TCA cycle and glyoxylate shunt in the S acetate specialists. There were no obvious changes in the promoter activities of *ackA* in either the S or L lineages. Of course, there may be discrepancies between promoter activities and actual enzyme activities [55, 56]. Nonetheless, these patterns agree reasonably well with the flux predictions from the evoFBA simulations, especially as they relate to the higher activities in the S lineage of the genes that specifically promote growth on acetate. As noted above, we reiterate that the evoFBA simulations predict an eventual complete loss of the acetate-specific activities in the L lineage, whereas thus far they are merely expressed at a lower level in the L lineage than in the S lineage.

Discussion

We developed a modeling framework, called evoFBA, which combines metabolic models that are amenable to FBA with an evolutionary algorithm to simulate the interplay of evolutionary and ecological dynamics in systems with multiple strains or species. We applied evoFBA to the LTEE with *E. coli* and predicted the emergence of two stably coexisting lineages with distinct metabolic flux distributions that promote a cross-feeding interaction. These predictions fit with the polymorphism seen in the Ara-2 population, where two lineages emerged early in the LTEE that have now coexisted for tens of thousands of generations [26, 50]. The evoFBA simulations enabled us to hypothesize specific ecological and physiological mechanisms that generate and sustain this polymorphism, and we then tested these hypotheses using the bacteria from that LTEE population. On balance, we found that the ecological, physiological, and metabolic properties of these coexisting lineages agree reasonably well with the predictions of the evoFBA.

Only one of the 12 LTEE populations evolved a persistent polymorphism that has been studied in such detail [26, 50]. However, other LTEE populations show evidence of negative frequency-dependent fitness, deep phylogenetic divergence, or both, which are consistent with adaptive diversification [57–60]. These results



suggest that other populations may have evolved cross-feeding interactions similar to the one studied here, even if they were not always so persistent [59]. One possible explanation for why persistent polymorphisms did not evolve in the other populations is that the establishment of the S lineage appears to have involved epistatic interactions between multiple mutations [50], which may have limited its evolutionary accessibility [58]. This possibility reflects one of the limitations of evoFBA, which cannot capture all of the intricacies of biological evolution but instead predicts optimal end states that emerge given the simplifying assumptions of this approach. For example, mutations in evoFBA affect the rates of resource uptake, but not the rates of internal reactions in the model. This limitation reflects the computational burden of simulating a multitude of mutant genotypes, the number of which would increase greatly if all reaction rates were subject to mutation. This limitation could be relieved by the development of more efficient algorithms (allowing mutations to affect all reactions in the model), but the final evolved model organisms might not differ functionally from those based on the current approach because changes in uptake rates can already affect downstream flux distributions. Another limitation of the evoFBA approach at this time is the assumption that constraints on the uptake fluxes can be changed only by mutation, while the optimization of fluxes within those constraints is immediate through FBA [32]. In other words, metabolic fluxes change within physiological limits without delay, whereas changing the limits themselves requires mutations. These assumptions are reasonable starting points for incorporating evolutionary dynamics into an FBA framework, but physiological delays in metabolic adjustments are also sometimes important [61, 62]. Expanding the evoFBA framework to include the dynamics of physiological transitions could start by integrating previous work on incorporating gene regulation into FBA [63, 64].

Adaptive diversification is expected, and has occurred, in other evolution experiments besides the LTEE, such as when two exogenous carbon sources are provided [27, 65] and in high-glucose chemostats, where substantial acetate is produced [21, 23, 25]. However, the adaptive diversification observed in the LTEE was unexpected owing to the presence of a single carbon source, glucose, which was supplied at a low concentration [52]. Using the evoFBA framework, we predicted that acetate secretion was the primary metabolic driver leading to the emergence of the polymorphism, and that prediction was supported by our experiments. The long duration of the LTEE—including several thousand generations to establish the S and L polymorphism [26] and its persistence for tens of thousands of generations [60]—may have facilitated adaptive divergence under these more restrictive

ecological conditions, in comparison with other studies of much shorter duration where glucose, acetate, or both were supplied exogenously at high concentrations [23, 25, 45]. In fact, low levels of acetate, as observed in our study, have previously been reported to favor generalists as opposed to divergence into coexisting specialists [10]. Nonetheless, the L lineage evolved higher glucose uptake rates, which led to acetate overflow and the construction of a new niche that benefited the S lineage, as occurred in the evoFBA simulations. Thus, niche construction by the bacteria led to the emergence of this polymorphism, in contrast to experiments where both carbon sources were added to the medium [45]. Despite the differences between the LTEE and previous evolution experiments [45, 65], similar metabolic processes emerged.

Conclusion

The combination of evoFBA and experimental evolution provides a useful approach that can give insights into general mechanisms involved in the emergence of bacterial diversity and community construction. This approach may stimulate the development of even more detailed and integrated studies aimed at predicting the outcomes of evolution experiments and dynamics in multi-species systems including synthetic microbial communities [51].

Methods

Evolutionary flux balance analysis (evoFBA)

In each evoFBA simulation, stoichiometric metabolic models were used to simulate clonal populations with distinct genotypes. Each genotype was represented by a metabolic model, which was simulated in a dynamical FBA formalism [46] to evaluate its growth and metabolic flux rates over time. At each time step of the dynamic FBA, the metabolic model was optimized using linear optimization and a pseudo-reaction representing biomass as the objective function [66]. This optimization thus maximized growth rate given the constraints on uptake rates, i.e. it optimized biomass yield per substrate [67]. Instead of defining specific uptake rates for a particular set of media components (as in standard FBA studies), we assumed a global constraint for all carbon and oxygen uptake reactions in each model organism. By limiting total uptake in the model (including O₂ “uptake”), we represent cellular limitations that can arise from many different factors, including redox cycling [20, 47], respiratory chain [18], enzyme expression [16, 48], and substrate uptake [17]. Similar implementations of global constraints in FBA models have been employed previously to study diauxic shift and substrate preference in *E. coli* [18, 19, 37]. The global uptake constraint implemented here favored a minimization of fluxes given the

maximization of the FBA objective in order to achieve the most efficient use of cellular resources for growth [34].

New model organisms were generated by random mutation from existing ones. Mutations altered specific bounds on individual uptake rates, while maintaining an overall total flux constraint of carbon and oxygen into the model organism. Thus, mutations change how the overall uptake flux is distributed across different substrates, and they allow a second level of optimization to occur over evolutionary time in addition to the optimization that occurs by FBA over the physiological time scale. Focusing evolution on a subset of reactions made computation of the ecological and evolutionary dynamics feasible; even so, the simulations presented here took over 20 days on a dedicated high-performance computer to simulate over 90,000 different model organism genotypes and their associated population and metabolite dynamics. A more complete simulation might encompass genome-scale models with evolution of all reactions in the model and with global constraints on total protein biomass [19] and membrane space [18]. The current implementation of evoFBA was unable to perform such simulations in a reasonable timeframe and with appropriate numbers of replicate simulations; efforts to run evoFBA with mutations allowed for all reactions caused a slowdown of over 10-fold relative to the current implementation.

For the evoFBA simulations, we implemented the *in silico* equivalent of the LTEE with *E. coli*. The simulations started with a population comprised of a single model genotype that represented the central metabolism of *E. coli* [68]. This model included 95 reactions, 75 metabolites, and 20 exchange reactions. The uptake of nutrients from the medium (i.e., the flux over the exchange reactions) was simulated by a Michaelis-Menten function for each substrate, v_j , as follows:

$$v_j = \frac{v_{max,j} \cdot [S_j]}{K_m + [S_j]} \quad (1)$$

where $v_{max,j}$ is the maximum uptake rate of the j th substrate in millimoles per g dry weight per h (mmol/gDW/h), $[S_j]$ is the concentration of the j th substrate in mmol/l, and K_m is the half-saturation constant of the transporter in mmol/l. For simplicity, we arbitrarily set the initial K_m values for all uptake reactions to 0.01 mmol/l. The $v_{max,j}$ values were allowed to evolve by mutation (see next section below). The value of v_j was then used as the uptake bound for the exchange reaction of each substrate when running FBA. For each simulated day, we evaluated each model using dynamic FBA [46] over the course of 24 h with 1-min steps; the simulation used a 10-ml batch reactor, as in the LTEE. At each time step, we set the v_j values for each

model using Eq. 1, used FBA to determine growth rate, and updated the biomass as follows:

$$BM_{t+1,i} = BM_{t,i} \left(1 + \frac{\mu_i}{\ln(2) \cdot 60} \right) \quad (2)$$

where $BM_{t,i}$ is the biomass in gDW of the clone represented by the i th model genotype at time t , and μ_i is the growth rate of that model computed by FBA [29]. After updating the biomass of all model genotypes, the resulting concentration of each substrate was reset as follows:

$$[S]_{j,t+1} = [S]_{j,t} - \sum_i \frac{BM_{i,t} \cdot v_{j,i}}{60 \cdot V} \quad (3)$$

where $[S]_{j,t}$ is the substrate concentration of the j th substrate at time t (mmol/l), BM_i is the biomass of the i th model clone at time t (gDW/l), $v_{j,i}$ is the uptake rate of the j th substrate by the i th model genotype as computed by FBA (mmol/gDW/h), and V is the culture volume (l). To mimic the LTEE, we started each day's culture with glucose at 0.1389 mmol/l. The culture was started with one model genotype (i.e., the core *E. coli* model [68]) having a v_{max} for glucose and acetate of 10 mmol/gDW/h each, and for oxygen of 20 mmol/gDW/h, i.e. the total uptake constraint was set at 40 mmol/gDW/h based on previous values for the combined uptake of carbon and oxygen [67]. The individual uptake rates for 14 carbon sources represented in the *E. coli* core model (acetate, acetaldehyde, α -ketoglutarate, ethanol, formate, fructose, fumarate, glucose, glutamine, glutamate, lactate, malate, pyruvate, succinate) and oxygen were subject to mutation in evoFBA. The exchange rates for phosphate, ammonia, water, protons, and carbon dioxide had no limits, reflecting the fact that carbon is the growth-limiting factor in the LTEE.

Representing mutations in evoFBA

The point mutation rate of *E. coli* in the LTEE (excluding populations that evolved mutator phenotypes [60]) has been estimated to $\sim 10^{-10}$ per base pair per generation, which equals $\sim 4 \times 10^{-4}$ per genome per generation [69]. Directly mapping mutations from bacteria to evoFBA model organisms is not possible. The mutable "genome" in the model organism has only 15 targets, as opposed to thousands of genes and millions of base pairs in an *E. coli* genome. Owing to these differences and computational limitations, we introduced mutations at the rate of 10^{-6} per model cell per generation in evoFBA. During the simulations, mutations were introduced into the population of each model genotype at each time step (i.e., simulated minute) according to the number of cells

produced in that step and those expected to contain a mutation ($N_{m(i)}$) as follows:

$$N_{m(i)} = \frac{N_i \mu_i}{\ln(2) \cdot 60 \cdot 10^6} \quad (4)$$

where N_i is the population size of model organism i . N_i was calculated from the biomass of model clone i divided by the mass of one cell in gDW; we used a mass of 600 fg/cell, which was reported previously for exponentially growing *E. coli* cells [70]. When $N_{m(i)}$ was between 0 and 1, $N_{m(i)}$ was used as a probability to determine whether or not a mutant was introduced; when $N_{m(i)}$ was ≥ 1 , a single mutant was always introduced.

As noted, each genotype in evoFBA corresponds to a different stoichiometric model with associated v_{max} values. When a mutation occurred, one of the uptake reactions was chosen at random and the maximum rate for that reaction was changed as follows:

$$v_{max,m,new} = v_{max,m,old} + a \quad (5)$$

where $v_{max,m,new}$ and $v_{max,m,old}$ are the new and old rates for the mutated reaction m , and a is a random number from the uniform distribution over the interval $(-10,10)$. Each individual uptake rate was further constrained to lie between 0 and 40, such that the total uptake rate of 40 mmol/gDW/h was not violated. After any mutation, all other uptake reactions were updated as follows:

$$v_{max,j,new} = \frac{v_{max,j,old}}{\sum_k v_{max,k,old}} (40 - v_{max,m,new}) \quad (6)$$

where k includes all uptake reactions except the mutated one. This adjustment ensures a constant total uptake flux across the membrane of 40 mmol/gDW/h. This mutation scheme generates strong tradeoffs between uptake reactions. The large effects of the mutations on reaction rates were chosen for computational speed of the evolutionary simulations; additional simulations with smaller maximum mutation steps produced qualitatively similar results (Fig. 2).

Simulating serial transfer and selection

To simulate the LTEE's daily transfer cycles [2], we used dynamical FBA to compute growth over 24 h; selection is a direct consequence of the differential growth of the model genotypes. After 24 h, a dilution was performed by randomly drawing 1 % of the model organisms, which constituted the initial population for the next simulated day. The next day's medium included 99 % of the initial medium and 1 % of the spent medium from the end of the previous day. The simulated growth and dilution ran for a total of 550 cycles. Results from replicate simulations (Fig. 2) are qualitatively similar to those in Fig. 1.

The population dynamics arising from these simulations are expected to give rise eventually to one dominant clone in each stably coexisting lineage. However, similar model organisms may occur within a simulation as a result of independent mutations before any one of them has reached its population maximum (e.g., model genotypes 13437, 12364 and 12719 in Fig. 1). However, if the model organisms differ even slightly in their uptake rates, then one genotype should eventually prevail through competitive exclusion, unless the model organisms occupy distinct ecological niches (Fig. 1).

Computation

Simulations were performed using MATLAB (Math Works, Natick, Massachusetts) and dynamic FBA calculations using the COBRA toolbox [71]. The MATLAB scripts used to run evoFBA and analyze the data are freely available at [72].

LTEE and bacterial strains

The LTEE consists of 12 populations founded from the same ancestral strain of *E. coli*, REL606 [73], that have been propagated since 1988 by daily 1:100 dilutions in Davis minimal medium [52] supplemented with glucose at 25 mg/l (DM25). Here, we focused on one population, called Ara-2, in which two lineages, S and L, diverged before 6500 generations and have co-existed ever since [26, 50, 60]. We studied the ancestor and one clone sampled from each lineage at 6500, 11,000, 18,000, 20,000, 30,000, 40,000 and 50,000 generations. Each evolved clone is named by its generation followed by S or L according to its lineage and an arbitrary numeral for a given clone. For example, 6.5KS1 is a clone from the S lineage that was sampled at 6500 generations.

Media and culture conditions

Bacteria were grown in the same medium as used in the LTEE [52], except that the carbon source was glucose at 250 mg/l (DM250-glucose), glucose at 1000 mg/l (DM1000-glucose), or acetate at 250 mg/l (DM250-acetate). These higher concentrations were used to increase cell density and thereby improve the accuracy of measurements of cell growth (e.g., Fig. 5) and concentrations of residual resources and secreted metabolites (e.g., Additional file 1: Figure S1). After overnight growth in DM1000-glucose, strains were inoculated by a 10,000-fold dilution into DM250-glucose, where they grew for 24 h at 37 °C with shaking at 120 rpm as an acclimation step. For each strain, three replicate acclimation cultures were then inoculated as duplicates, each at a 1:100 dilution, into DM250-glucose or DM250-acetate and incubated in 96-well microtiter plates at 37 °C for 24 h. Growth was monitored using an Infinite M200 microplate reader (Tecan, Lyon, France) by measuring the OD_{450nm} every

10 min. Growth rates were computed from filtered OD data as $\ln(\text{OD}_{450})/dt$ over a sliding window of 1 h, using MATLAB. We report the mean of the three replicates. Filtering was performed by removing negative growth rates and mean growth rates that were more than 0.2 units above or below the immediately adjacent data points (outliers).

Measuring glucose and acetate concentrations

The ancestor, 6.5KS1, and 6.5KL4 clones were grown in DM250-glucose as described. Samples were taken at time 0 and every h for 9 h. After centrifugation to remove cells, we measured glucose and acetate concentrations in the supernatant using the Glucose Assay Kit (Merck Millipore, Lyon, France) and Acetic Acid Assay Kit (Megazyme, Pontcharra-sur-Turdine, France), respectively, following the manufacturers' recommendations.

Analysis of flux patterns in individual model organisms

We simulated the growth of the evoFBA model organisms with IDs 44490 and 12364 (Fig. 1) to obtain the flux values for their biochemical reactions. Each model organism was simulated using dynamical FBA in medium containing 0.1389 mM glucose, the same concentration as in the LTEE. Each simulation ran for ten 24-h periods with daily 1:100 dilutions; the last day was used to record the flux values, in order to remove any effect of the initial conditions. The flux patterns for growth on glucose were taken 10 min after the onset of growth (Fig. 4a and c), and for growth on acetate at 388 min because glucose was exhausted while acetate was still present at a substantial level (Fig. 4b and d). From the flux patterns, we identified several reactions of interest that showed differences between the two evolved model organisms (highlighted in blue in Fig. 4).

Identification of metabolites in filtrates of spent cultures of 6.5KL4

We analyzed by HPLC and GC-MS the metabolic by-products secreted by clone 6.5KL4 using filtrates from 24-h spent cultures of that clone in DM25- and DM250-glucose, both before and after growth of clone 6.5KS1. For HPLC, 1 ml of filtrate was acidified with 5 μ l 1 M H_2SO_4 , incubated at room temperature for 5 min, and passed through a 0.45- μ m regenerated cellulose syringe filter (PHENEX RC Membrane, Phenomenex, Le Pecq, France). Samples were then analyzed on an Agilent 1260 Infinity HPLC system equipped with a RezEX ROA-Organic Acid (8 %) 300 \times 7.8-mm column (Phenomenex) and a diode array detector. The analytical conditions were as follows: mobile phase, 5 mM H_2SO_4 ; flow rate, 0.6 ml/min; column temperature, 35 $^\circ\text{C}$; injection volume, 50 μ l; wavelength scan range, 190–400 nm; detection wavelength, 210 nm; and run time, 35 min. Concentrations

of acetate and fumarate in the L-clone filtrates were determined from linear standard curves over the ranges of 0–10 mM and 0–100 μ M, respectively, and with lower detection limits of 0.1 mM and 0.3 μ M, respectively. Succinate, lactate, formate, propionate, and butyrate can also be separated under these analytical conditions with detection limits similar to acetate, but they were not detected in any samples.

GC-MS analysis of volatile compounds was performed using an Agilent GC HP6890 gas chromatograph equipped with a Varian CP-WAX 58 column (length, 25 m; internal diameter, 0.25 mm; film thickness, 0.20 μ m), and coupled to an MSD5973 mass sensitive detector. The sample (600 μ l) was cooled on ice, acidified with 50 μ l 4 M HCl, and extracted with 0.375 g NaCl and 650 μ l ice-cold ether. After vortexing three times for 10 s each, with 30 s cooling intervals, the sample was centrifuged for 5 min at 10,000 rpm and placed on ice for 5 min. The upper organic layer (2.5 μ l) was then injected manually into the GC using an ice-cold syringe (injection in split mode, split ratio = 10). The column was held at 40 $^\circ\text{C}$ for 1 min, ramped to 200 $^\circ\text{C}$ at a rate of 5 $^\circ\text{C}/\text{min}$, and held for a further 3 min, giving a total run time of 36 min. The solvent delay for the MSD was 1.4 min and the mass scan range was set to 35–300 atomic mass units. The presence of acetate (retention time 12.2 min) in the L filtrate after growth in both DM25- and DM250-glucose was confirmed by this method. The estimated concentration from the DM250-glucose filtrate was 510 μ M, which is close to the 480 μ M detected by HPLC (Additional file 1: Figure S1). The concentration of acetate in the L filtrate from DM25-glucose was about one-tenth that detected in DM250-glucose. Ethanol (retention time 2.1 min) was also detected in the filtrates of all three strains tested (ancestor, 6.5KL4, and 6.5KS1). Other metabolites including isopropanol, butanol, acetoin, acetone, formic acid, propionic acid, butyric acid, isobutyric acid, and valeric acid were not detected (with lower detection limits around 50–100 μ M in scan mode).

Analysis of promoter activities in LTEE clones

We measured the activities of the promoters of four genes—*pgi*, *acnB*, *aceB*, and *ackA*—that encode enzymes associated with reactions of interest (Table 1) given the results of evoFBA (Fig. 4). We used the corresponding

Table 1 Genes used in the analysis of promoter activities

Name	Gene ID	Gene	Protein	FBA model term
Glucose-phosphate isomerase	948535	<i>pgi</i>	PGI	PGI
Aconitate hydratase	944864	<i>acnB</i>	ACN	ACONTb
Malate synthase A	948512	<i>aceB</i>	ACE	MALS
Acetate kinase	946775	<i>ackA</i>	ACK	ACKr

reporter plasmids from the *E. coli* library of *gfp* transcriptional fusions [74]. Each of the four plasmids, as well as the empty pUA66 reference plasmid, was introduced into the ancestor, 6.5KS1, 6.5KL4, 50KS1, and 50KL1 clones. Each plasmid-bearing clone was grown in DM250-glucose supplemented with 25 µg/ml kanamycin. Both OD_{450nm} and GFP fluorescence were measured every 10 min for 24 h in the microplate reader. Promoter activities were estimated as the rate of GFP production from the promoter region [74]. They were computed using MATLAB as $(dGFP/dt)/OD_{450nm}$, where GFP is the fluorescence signal after subtracting the value for the empty plasmid and division by OD_{450nm} standardizes the data with respect to cell biomass density. We show the mean values from three replicate experiments (Fig. 6).

Additional files

Additional file 1: Figure S1. HPLC profiles of filtrate from spent cultures of 6.5KL4 before and after growth of 6.5KS1. Partial HPLC chromatograms, scaled in milli Absorbance Units (mAU) at 210 nm, showing elution time (min) of key metabolites for the filtrate of a 24-h spent culture of clone 6.5KL4 in DM250-glucose (A), and for the same filtrate after 24 h of growth of clone 6.5KS1 at 37 °C (B). The L filtrate contained 2-hydroxyglutarate, acetate, and fumarate. The S clone consumed the acetate and fumarate, but not the 2-hydroxyglutarate. The acetate peak indicates a concentration of 480 µM, whereas the fumarate peak indicates a concentration of only 0.67 µM; the molar absorption coefficient of fumarate at 210 nm is more than 300 times greater than that of acetate. (PDF 64 kb)

Additional file 2: Figure S2. Diauxic growth in DM250 medium containing glucose and acetate at 10:90 ratio. Clone 6.5KS1 (red) exhibits a diauxic shift from glucose to acetate consumption much earlier than either the ancestor (green) or clone 6.5KL4 (blue). Curves show the average of three biological replicates. (PDF 113 kb)

Additional file 3: Figure S3. Growth curves of the ancestor and evolved clones in DM250-glucose and DM250-acetate media. **A** Growth curves of the Anc and pre-divergence clone 2 K4 (both shown in green) on glucose (left) and acetate (right). **B** Growth curves of S (red) and L (blue) clones sampled at seven generations (6.5, 11, 18, 20, 30, 40, and 50 K arranged chronologically from top to bottom) on glucose (left) and acetate (right). In each panel, curves show the average (heavy line) of 3–6 replicate assays (lighter lines) for each clone; curves for individual replicates are not always visible when they are close to the mean or other replicates. (PDF 186 kb)

Acknowledgements

JC acknowledges the French ministry of research for a research fellowship. This work was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC, BB/K003240/1 to OSS), Agence Nationale de la Recherche Programme Blanc (ANR-08-BLAN-0283-01 to DS), European Union programme FP7-ICT-2013-10 project EvoEvo (610427 to DS), Université Grenoble Alpes (to DS), Centre National de la Recherche Scientifique (CNRS, to DS), the U.S. National Science Foundation (DEB-1451740 to REL), and the BEACON Center for the Study of Evolution in Action (NSF Cooperative Agreement DBI-0939454 to REL).

Authors' contributions

JC, TG, JG performed the experiments, analysed and interpreted the data, and drafted the article; JW performed the HPLC experiments and drafted the article; JC, TG, DS, OSS designed the study, analysed and interpreted the data, drafted and revised the article; REL directs the LTEE and helped draft and revise the article. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Life Sciences, University of Warwick, Coventry, UK. ²University of Grenoble Alpes, Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications, Grenoble (TIMC-IMAG), F-38000 Grenoble, France. ³Centre National de la Recherche Scientifique (CNRS), TIMC-IMAG, F-38000 Grenoble, France. ⁴University of Grenoble Alpes, Institut de recherches en technologies et sciences pour le vivant – Laboratoire de chimie et biologie des métaux (iRTSV-LCBM), Grenoble F-38000, France. ⁵CNRS, iRTSV-LCBM, F-38000 Grenoble, France. ⁶Commissariat à l'énergie atomique (CEA), iRTSV-LCBM, F-38000 Grenoble, France. ⁷Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA. ⁸BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824, USA.

Received: 21 April 2016 Accepted: 4 August 2016

Published online: 20 August 2016

References

- Stern DL, Orgogozo V, Pierre U, Bâtiment A, Saint Q. Is genetic evolution predictable? Nonrandom distribution of evolutionarily relevant mutations. *Science*. 2009;323:746–51.
- Wiser MJ, Ribeck N, Lenski RE. Long-term dynamics of adaptation in asexual populations. *Science*. 2013;342:1364–7.
- de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet*. 2014;15:480–90.
- Nesse RM, Stearns SC. The great opportunity: Evolutionary applications to medicine and public health. *Evol Appl*. 2008;1:28–48.
- Soyer OS, O'Malley MA. Evolutionary systems biology: What it is and why it matters. *BioEssays*. 2013;35:696–705.
- Lohbeck KT, Riebesell U, Reusch TBH. Adaptive evolution of a key phytoplankton species to ocean acidification. *Nat Geosci*. 2012;5:346–51.
- Novais Â, Comas I, Baquero F, Cantón R, Coque TM, Moya A, et al. Evolutionary trajectories of beta-lactamase CTX-M-1 cluster enzymes: Predicting antibiotic resistance. *PLoS Pathog*. 2010;6:e1000735.
- Papp B, Notebaart RA, Pál C. Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet*. 2011;12:591–602.
- Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, et al. The role of ecological theory in microbial ecology. *Nat Rev Microbiol*. 2007;5:384–92.
- Doebeli M. A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms. *Popul Ecol*. 2002;44:59–70.
- Pfeiffer T, Bonhoeffer S. Evolution of cross-feeding in microbial populations. *Am Nat*. 2004;163:E126–35.
- Gudelj I, Beardmore RE, Arkin SS, MacLean RC. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. *J Evol Biol*. 2007;20:1882–9.
- Schuster S, Pfeiffer T, Fell DA. Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol*. 2008;252:497–504.
- Pfeiffer T, Schuster S, Bonhoeffer S. Cooperation and competition in the evolution of ATP-producing pathways. *Science*. 2001;292:504–7.
- Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol*. 2009;5:323.
- Weiße AY, Oyarzún DA, Danos V, Swain PS. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci U S A*. 2015;112:E1038–47.
- Button D. Nutrient uptake by microorganisms according to kinetic parameters from theory as related to cytoarchitecture. *Microbiol Mol Biol Rev*. 1998;62:636–45.
- Zhuang K, Vemuri GN, Mahadevan R. Economics of membrane occupancy and respiro-fermentation. *Mol Syst Biol*. 2011;7:500.
- Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási A-L, et al. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci U S A*. 2007;104:12663–8.
- van Hoek MJ, Merks RMH. Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Syst Biol*. 2012;6:22.
- Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA. Overflow metabolism in *Escherichia coli* during steady-state growth: Transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol*. 2006;72:3653–61.

22. Schuster S, de Figueiredo LF, Schroeter A, Kaleta C. Combining metabolic pathway analysis with evolutionary game theory. Explaining the occurrence of low-yield pathways by an analytic optimization approach. *Biosystems*. 2011;105:147–53.
23. Rosenzweig RF, Sharp RR, Treves DS, Adams J. Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics*. 1994;137:903–17.
24. Turner PE, Souza V, Lenski RE. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology*. 1996;77:2119–29.
25. Treves DS, Manning S, Adams J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol Biol Evol*. 1998;15:789–97.
26. Rozen DE, Lenski RE. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am Nat*. 2000;155:24–35.
27. Friesen ML, Saxer G, Travisano M, Doebeli M. Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution*. 2004;58:245–60.
28. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol*. 1994;60:3724–31.
29. Orth J, Thiele I, Palsson B. What is flux balance analysis? *Nat Biotechnol*. 2010;28:245–8.
30. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Curr Opin Biotechnol*. 2003;14:491–6.
31. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*. 2007;3:121.
32. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*. 2002;420:20–3.
33. Heinrich R, Schuster S, Holzhütter H-GG. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur J Biochem*. 1991;201:1–21.
34. Holzhütter HG. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem*. 2004;271:2905–22.
35. Hoppe A, Hoffmann S, Holzhütter H-G. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst Biol*. 2007;1:23.
36. Thiele I, Jamshidi N, Fleming RMT, Palsson BØ. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol*. 2009;5:e1000312.
37. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol*. 2013;9:693.
38. Klitgord N, Segrè D. Ecosystems biology of microbial metabolism. *Curr Opin Biotechnol*. 2011;22:541–6.
39. Zomorodi AR, Maranas CD. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol*. 2012;8:e1002363.
40. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep*. 2014;7:1104–15.
41. Klitgord N, Segrè D. Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol*. 2010;6:e1001002.
42. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, et al. Competitive and cooperative metabolic interactions in bacterial communities. *Nat Commun*. 2011;2:589.
43. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol*. 2007;3:92.
44. Khandelwal RA, Olivier BG, Röling WFM, Teusink B, Bruggeman FJ. Community flux balance analysis for microbial consortia at balanced growth. *PLoS One*. 2013;8:e64567.
45. Louca S, Doebeli M. Calibration and analysis of genome-based models for microbial ecology. *Elife*. 2015;4:1–17.
46. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*. 2002;83:1331–40.
47. el-Mansi EM, Holms WH. Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. *Microbiology*. 1989;135:2875–83.
48. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*. 2015;528:99–104.
49. Lenski RE, Wisner MJ, Ribbeck N, Blount ZD, Maddamsetti R, Burnmeister AR, et al. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proc R Soc B*. 2015;282:20152292.
50. Plucain J, Hindré T, Le Gac M, Tenaillon O, Cruveiller S, Médigue C, et al. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science*. 2014;343:1366–9.
51. Großkopf T, Soyer OS. Synthetic microbial communities. *Curr Opin Microbiol*. 2014;18:72–7.
52. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat*. 1991;138:1315–41.
53. Leiby N, Marx CJ. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS Biol*. 2014;12:e1001789.
54. Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput Biol*. 2013;9:e1003091. Stelling J, editor.
55. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol*. 2010;6:390.
56. Rossell S, van der Weijden CC, Lindenberg A, van Tuijl A, Francke C, Bakker BM, et al. Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 2006;103:2166–71.
57. Elena SF, Lenski RE. Long-term experimental evolution in *Escherichia coli*: VII. Mechanisms maintaining genetic variability within populations. *Evolution*. 1997;51:1059–67.
58. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2008;105:7899–906.
59. Maddamsetti R, Lenski RE, Barrick JE. Adaptation, clonal interference, and frequency dependent interactions in a long term evolution experiment with *Escherichia coli*. *Genetics*. 2015;200:619–31.
60. Tenaillon O, Barrick JE, Ribbeck N, Deatherage DE, Blanchard JL, Dasgupta A, et al. Tempo and mode of genome evolution in a 50,000 - generation experiment. *Nature*. 2016; in press.
61. Lenski RE, Souza V, Duong LP, Phan QC, Nguyen TNM, Bertrand KP. Epistatic effects of promoter and repressor functions of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli*. *Mol Ecol*. 1994;3:127–35.
62. Le Gac M, Brazas MD, Bertrand M, Tyerman JG, Spencer CC, Hancock REW, et al. Metabolic changes associated with adaptive diversification in *Escherichia coli*. *Genetics*. 2008;178:1049–60.
63. Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*. 2007;3:101.
64. Covert MW, Palsson B. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem*. 2002;277:28058–64.
65. Spencer CC, Bertrand M, Travisano M, Doebeli M. Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *PLoS Genet*. 2007;3:e0083–8.
66. Feist A, Palsson B. The biomass objective function. *Curr Opin Microbiol*. 2010;13:344–9.
67. Edwards JS, Ibarra RU, Palsson BO. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol*. 2001;19:125–30.
68. Orth JD, Fleming RM, Palsson BØ (2010a) 10.2.1 – Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. In *EcoSal – Escherichia coli and Salmonella Cellular and Molecular Biology*, Karp PD (ed), 10.2.1. Washington DC: ASM Press.
69. Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Médigue C, et al. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3*. 2011;1:183–6.
70. Loferer-Krößbacher M, Klima J, Psenner R. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Appl Environ Microbiol*. 1998;64:688–94.

71. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc.* 2007;2:727–38.
72. MATLAB scripts used to run evoFBA and analyze data. <http://osslab.lifesci.warwick.ac.uk/?pid=resources>.
73. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, et al. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol.* 2009;394:644–52.
74. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods.* 2006;3:623–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Figure S1

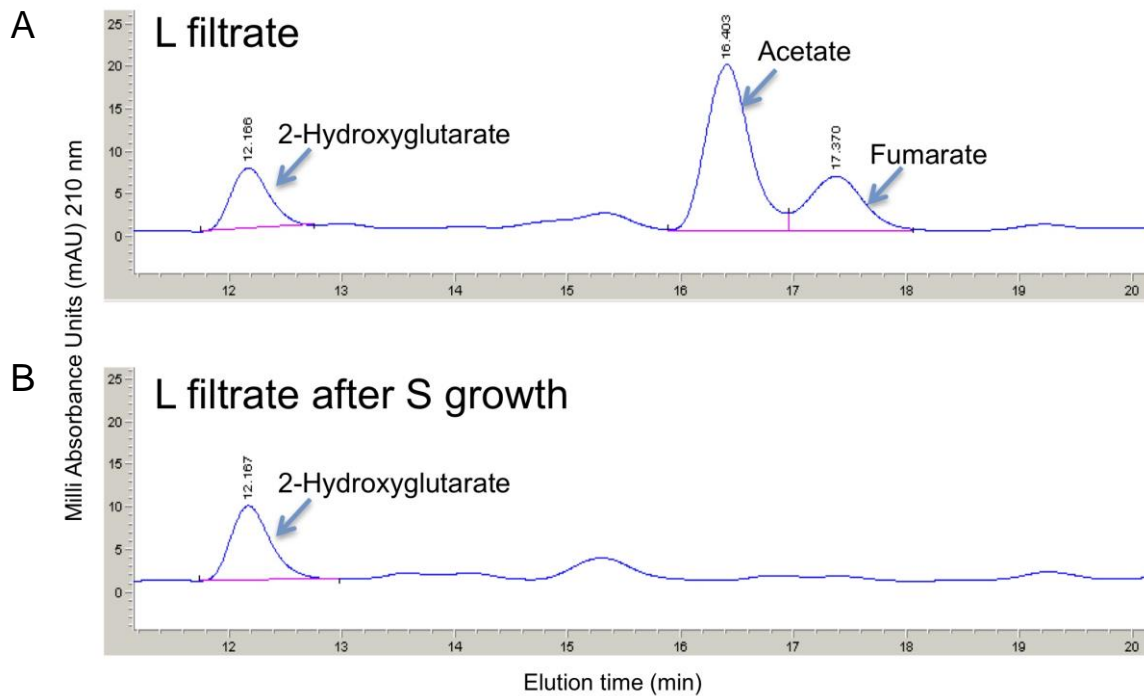
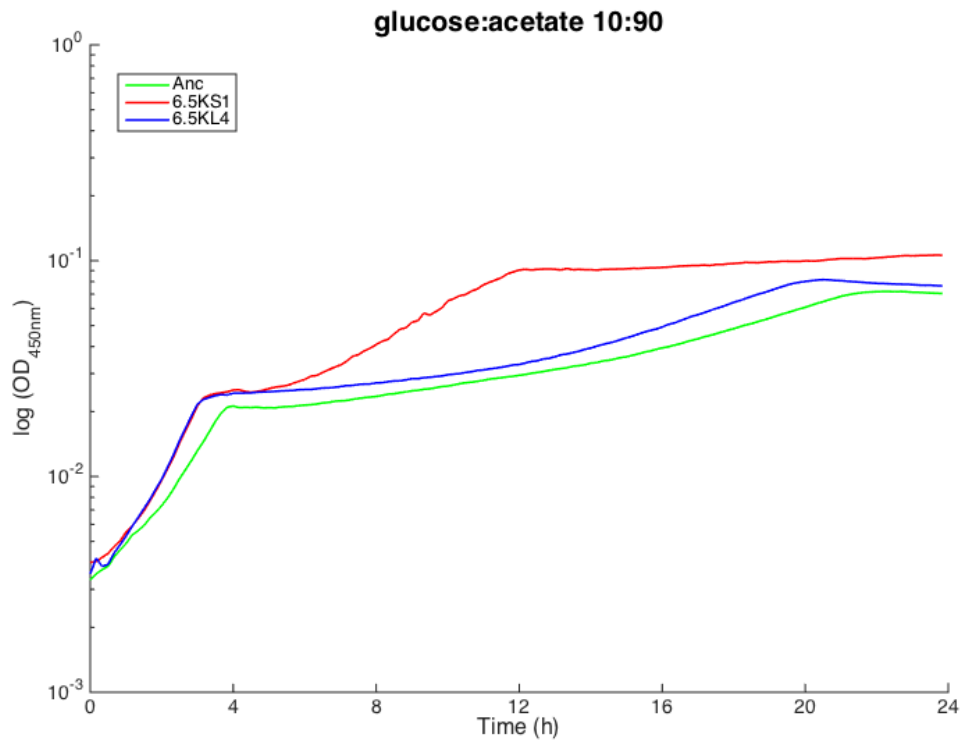


Figure S2



**Chapter 2: Phenotypic Causes of Fast Loss of
Long-Term *Escherichia coli* Ecotypic
Coexistence After Decreased Competition**

We showed in the previous chapter that ecological character displacement was a key driver of adaptive diversification resulting in the S and L polymorphism. This subject will be developed in this chapter, which is presented as a manuscript entitled, “*Phenotypic causes of fast loss of long-term Escherichia coli ecotypic co-existence after decreased competition*” (Consuegra J, Rozen D, Hindré T, Lenski RE, Gaffé J, Schneider D.) that will be submitted to *BMC Evolutionary Biology*. In addition to the results presented below, we obtained genome sequences from the evolved clones that are analyzed here. These genome sequences were available a few days before sending the PhD manuscript and their analyses will later be included in the manuscript that will be submitted.

Previous studies (Tyerman et al. 2008), using experimental bacterial populations evolving in the presence of two carbon sources, have shown that competition was a cause of character displacement resulting in adaptive diversification. Therefore, we studied here the role of competition in the long-term co-existence of the S and L lineages in the population Ara-2 of the LTEE. The two S and L lineages emerged in sympatry owing to niche construction and their long-term co-existence is due to negative frequency-dependent selection and specialization of S and L for consumption of acetate and glucose, respectively (See Results, Chapter 1). We ran a short 300-generation evolution experiment in allopatry, *i.e.* S and L clones were propagated separately, thereby abolishing competition between S and L. We showed that these conditions resulted in the fast loss of the negative frequency-dependent interactions between S and L (Fig. 1) due to phenotypic and resource use convergence (Figs. 2 & 3).

In sum, abolition of competition allows the simultaneous invasion of both ecological niches (acetate and glucose) by each of the two S and L lineages, and thus the loss of ecological character displacement.

Phenotypic causes of fast loss of long-term *Escherichia coli* ecotypic co-existence after decreased competition.

Jessika Consuegra^{1,2}, Daniel Rozen³, Thomas Hindré^{1,2}, Richard E. Lenski^{4,5}, Joël Gaffé^{1,2*},
Dominique Schneider^{1,2}

1 Université Grenoble Alpes, Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications, Grenoble (TIMC-IMAG), F-38000 Grenoble, France

2 Centre National de la Recherche Scientifique (CNRS), TIMC-IMAG, F-38000 Grenoble, France

3 Institute of Biology, Leiden University, 2333 BE Leiden, The Netherlands

4 Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

5 BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824, USA

jessika.consuegra-bonilla@univ-grenoble-alpes.fr

joel.gaffe@univ-grenoble-alpes.fr

d.e.rozen@biology.leidenuniv.nl

lenski@msu.edu

thomas.hindre@univ-grenoble-alpes.fr

dominique.schneider@univ-grenoble-alpes.fr

*Corresponding author

Abstract (350 words)

Background: Ecological character displacement is an important mechanism that can drive adaptive diversification in sympatry when resources are limiting and thus impose strong competition. In the absence of geographic isolation, natural selection favors individuals from the nascent community that are able to use alternative resources. In the resulting divergent populations competition between individuals is reduced and sympatric coexistence may ultimately emerge through the establishment of frequency dependent interactions between ecotypes. One example of character displacement as driver of adaptive diversification in sympatry comes from one population of the Long-Term Evolution Experiment (LTEE) with *Escherichia coli*, called Ara-2, where two lineages, named S (Small) and L (Large) according to their cell size, emerged from a common ancestor and co-exist ever since through negative frequency-dependent selection. L evolves higher and earlier glucose uptake rates together with acetate overflow that drives the emergence of the S ecotype that colonized this new ecological niche. Further specialization of the two ecotypes on those alternative niches leads to long term co-existence of the S and L lineages. Here, we test the role of competition in the maintenance of S and L co-existence through negative frequency-dependent selection. To that aim, one S and one L co-existing clones were isolated from the Ara-2 population and further propagated for 300 generations but in allopatric conditions.

Results: Abolishing competition by means of a short allopatry propagation, rapidly induced the loss of frequency dependent interactions. with most L^{Allo} lineages now out-competing the S lineages whatever their initial frequency. These L lineages however remain susceptible to invasion by S^{Sym} clone whereas the allopatric propagation allowed the S lineage to evolve the capacity to out compete the initial L^{Sym} clone. Further characterization of allopatrically evolved populations revealed that S improved on both available carbon sources whereas L maintained its optimal glucose usage but improved only on acetate.

Conclusions: This work evidences competition between diversifying lineages as a major driver of ecological character displacements sustaining long-term co-existence. It also reveals an

asymmetry in the selective pressure that co-existing ecotypes impose on each other, with the L forcing the S to further specialization on acetate at the cost of trade-offs on the alternative niche whereas the already optimized glucose usage by the L preserve the later from niche colonization by the S.

Keywords (3 to 10): Allopatric evolution, LTEE, character displacement, negative-frequency dependence, adaptive diversification.

Background

Evolutionary diversification and adaptive radiation arise when sympatric species compete for the same set of limited resources, with natural selection favoring divergence in resource use and phenotype [1]. This idea of ecological diversification, was first evoked by Brown in 1956 and named “Character displacement” [2]. In this seminal paper, Brown considered that the process of ecological character displacement produces exaggerated divergence among lineages evolving in sympatry whereas similarities in resource use and phenotype are generally observed for lineages evolving in allopatry [1].

Ecological character displacement have been linked with changes in the relative abundances of two resources which produces selective pressures for divergence and with it shifts in characters that have genetic variability in their relative intake rates of those resources [3]. When two similar species evolve in sympatry, but with limited resources, the species are likely to compete strongly. One possibility is that one competitor drives the other to extinction. However, natural selection may also favor, in each population, individuals using a resource not used by others, thus driving populations to diverge in phenotype and resource use. This divergence reduces resource competition and allows coexistence of more specialized individuals [4] that is generally maintained by frequency dependent interactions [1]. Such divergent selection is supposed to favor trade-offs in fitness across ecological niches, because specialization may be

accompanied by a cost of adaptation, from antagonistic pleiotropy (alleles beneficial in one niche are deleterious in others) or mutation accumulation (mutations neutral in the niche of selection may disrupt function in novel niches) [5].

Ecological character displacement due to resource competition has been widely studied in mathematical models [6-8]. Evidence of character displacement in nature have been shown in lizards [9] and Galapagos finches [10] among others (review in [1]) all of them having in common resource competition as the trigger of adaptive diversification. Likewise, several *in-vitro* evolution experiments attribute resource competition as the origin of diversification [11, 12].

Friesen et al. showed adaptive diversification in a lineage of *Escherichia coli* evolved for 1,000 generations in minimal medium containing glucose (205 mg/L) and acetate (205 mg/L) as carbon sources [11]. Further studies demonstrate a causal role for competition in ecological character displacement in this population, and how the abolishment of competition induces phenotypic convergence [13]. However, the long-term fate and evolution of this population are unknown.

The second example of character displacement triggered by resource competition comes from the Long-Term Evolution Experiment (LTEE) with *E. coli* [14]. One of the 12 long-term evolved populations, called Ara-2, presents an adaptive diversification event. In this population, two different lineages, named S (Small) and L (Large) according to their cell size, emerged from a common ancestor after 6,500 generations of evolution in Davis minimal medium supplemented only with 25 mg/L of glucose as limiting carbon source. The two lineages co-exist ever since, owing to negative frequency-dependent selection whereby each lineage is favored and invades the other when rare, assuring that no lineage gets extinct [15-18]. L evolves higher glucose uptake rates accompanied with the secretion of metabolites that sustains S growth [16, 19]. Further studies demonstrated that acetate is the secreted L by-product and that this newly available niche is the main metabolic driver leading to the emergence of the S lineage

[12]. These environmental variations lead to the evolution of niche specialists with fitness trade-off across the alternative niches, with L having high growth rates on glucose but poor growth on acetate whereas S exhibits the opposite pattern. Indeed, the fast scavenging of acetate by S in sympatry prevents L improvement on this carbon source since it doesn't represent a selective pressure for selection [12]. S and L emergence implicates ecological character displacement where neither lineage is able to evolve simultaneously on both carbon sources and their long-term co-existence is maintained by negative frequency-dependent selection [12, 16-20].

Here, we explore the ecological dynamics of character displacement in a sympatric diverged population. We test the role of competition in the maintenance of S and L co-existence through negative frequency-dependent selection. We showed that once competition is abolished, by means of allopatric evolution, frequency dependence is rapidly lost. We then addressed the causes of this loss that we showed to rely on phenotypic convergence disrupting the existent specialization of the two lineages.

Methods

Media and culture conditions

The LTEE has been carried out since 1988. In this experiment, twelve independent populations were derived from the same ancestor clone of *E. coli* B REL 606 and have been propagated for over 60,000 generations in Davis minimal medium supplemented with 25 mg/L of glucose (DM25-glucose). For all the 12 populations, 10-mL cultures are maintained in 50-mL Erlenmeyer flasks placed in a rotatory shaker at 37°C and 120 rpm [14]. Here, bacteria were grown in Davis medium as in the LTEE and the carbon source was changed according with the experiment, glucose at 1000 mg/L (DM1000-glucose), glucose at 250 mg/L (DM250-glucose), or acetate at 250 mg/L (DM250-acetate). These higher concentrations were used to increase cell density and improve the accuracy of measurements of cell growth and metabolites

concentration. In general, after overnight growth in DM1000-glucose from glycerol stocks, strains were acclimated in triplicate into DM250-glucose, for 24 h at 37°C and shaking at 120 rpm after a 10,000-fold dilution. Each acclimated culture was then inoculated as duplicate, at a 1:100 dilution in the test media.

Allopatric evolution

We isolated one S and L clone from the sympatric culture of population Ara-2 at 18,000 generations, 18KS^{Sym}, 18KL^{Sym}, respectively. A spontaneous Ara⁺ revertant was isolated from 18KS^{Sym} and used to start the allopatric evolution in order to facilitate future competition experiments. After acclimation in DM25-glucose we started six allopatric replicated populations of each, 18KS^{Allo(1 to 6)} and 18KL^{Allo(1 to 6)}. These populations were evolved by serial daily transfers in the same conditions as the LTEE during 300 generations. Isolated clones of these populations are designed by the name of the population followed by a capital letter, for example a clone isolated from 18KS^{Allo1} population is labelled 18KS^{Allo1A}.

Competition experiments.

We performed pairwise competition experiments to determine whether the sympatric and/or allopatric L and S lineages could invade one another when rare. The pairs included, 18KL^{Sym} vs. 18KS^{Sym}, 18KL^{Sym} vs. 18KS^{Allo} (Clones Allo1A to Allo6A), 18KS^{Sym} vs. 18KL^{Allo} (Clones Allo1A to Allo6A), 18KS^{Allo} (Clones Allo1A to Allo6A) vs. 18KL^{Allo} (Clones Allo1A to Allo6A). Each competition experiment was run for 24 h in the same medium and conditions as in the LTEE. The competitors were mixed at two initial volumetric ratios (1:9 and 9:1). The L and S competitors were distinguished on the basis of an arabinose marker that has no significant effect on fitness under these conditions. The fitness of one competitor relative to the other was calculated as the ratio of their net growth rates during the competition experiment [14, 16].

Growth rates assay

After acclimation cultures, the ancestor (Anc), 18KS^{Sym}, 18KL^{Sym}, 50KS^{Sym}, 50KL^{Sym} (one S and L clone from the sympatric culture of population Ara-2 isolated at 50,000 generations, respectively) and 18KS^{Allo(1 to 6)}, 18KL^{Allo(1 to 6)} populations were inoculated as duplicates in DM250-glucose or DM250-acetate and incubated in 96-well microtiter plates at 37°C for 24 h. Growth was monitored measuring the OD_{450nm} every 10 min with an Infinite M200 microplate reader (Tecan®). Growth rates were determined by fitting an exponential growth model using custom analysis software, Curve Fitter (N. F. Delaney, CJM, unpublished) (<http://www.evolvedmicrobe.com/Software.html>).

Metabolites kinetics

The 18KS^{Sym}, 18KL^{Sym}, 18KS^{Allo1A}, 18KS^{Allo4A}, 18KL^{Allo1A}, 18KL^{Allo4A} clones were grown allopatrically in DM250-glucose as described before, except that inoculation after acclimation cultures was performed in 50 mL of DM250-glucose (250 mL flask). One-milliliter samples were taken every hour until early stationary phase (~8h). We measured growth (OD_{450nm}) and recover the culture supernatants after centrifugation to then quantify the glucose consumed and the acetate produced for each clone using the Glucose Assay Kit (Merck Millipore) and Acetic Acid Assay Kit (Megazyme), respectively, following the manufacturers' recommendations.

Results

Abolishment of competition leads to a loss of frequency dependence.

Figure 1A shows a typical pattern of negative frequency dependence between S and L clones isolated from the Ara-2 population at 18,000 generations. In a competition between 18KL^{Sym} and 18KS^{Sym}, at two different ratios, any initially rare ecotype is able to invade the other (relative fitness >1 when rare but <1 when frequent, thus sustaining long-term coexistence of

the two lineages). To test the robustness of such ecological equilibrium, we repeated competition experiments using clones evolved from 18KS^{Sym} and 18KL^{Sym} clones but in allopatric conditions for 300 generations (Fig. 1D). When competed against 18KS^{Allo}, 4 out of the 6 tested 18KL^{Allo} clones were able to invade the other ecotype whatever their initial frequency. Reciprocally, these results indicate that 18KS^{Allo} clones are no longer able to invade 18KL^{Allo} populations when initially rare except for 18KL^{Allo1A} and 18KL^{Allo4A}. Accordingly, a short term allopatric evolution of the two previously coexisting ecotypes thus abolished their negative frequency dependent interaction. To further investigate the impact of allopatric propagation on the S/L interaction, we then competed 18KS^{Allo} clones against the 18KL^{Sym} clone (Fig. 1B). In 4 cases, the relative fitness of 18KS^{Allo} clones to 18KL^{Sym} did not appear significantly dependent on the initial frequency of the 2 competitors and 3 out of these 4 18KS^{Allo} clones evolved higher fitness than the L clone contemporary to their ancestor. For the two last 18KS^{Allo} clones, their interaction with the 18KL^{Sym} remains frequency dependent, negatively for 18KS^{Allo4A} that however is not significantly outcompeted when frequent, but positively for 18KS^{Allo6A} that outcompetes 18KL^{Sym} only when frequent. The reciprocal competition experiments were also conducted by competing 18KL^{Allo} clones against 18KS^{Sym} (Fig. 1C). Interestingly, in all, but one case the interactions appeared to remain dependent on frequency with the fitness of the 18KL^{Allo} relative to 18KS^{Sym} being systematically higher when initially rare. Except for the 18KL^{Allo6A}, it thus appeared that the allopatric propagation of the 18KL clone did not completely abolish the capacity of 18KS^{Sym} to invade the L ecotype when rare although with lower efficiency since 18KL^{Allo3A} and 18KL^{Allo5A} were not significantly outcompeted when frequent. All together, these results clearly demonstrate that independent propagation of S and L lineages in allopatric conditions rapidly disrupts pre-existing negative frequency dependent interaction between the two ecotypes. Moreover, they indicate that this lost most probably relies on some adaptation of 18KS derived clones that occurred during their propagation in allopatric conditions since 18KL^{Allo} populations retained their propensity to be

invaded by 18KS^{Sym} whereas 18KS^{Allo} no longer exhibit frequency dependent interactions with 18KL^{Sym}. Accordingly, competition in sympatric S/L evolving populations may predominantly consist in L lineage imposing character displacement in the S ecotype.

S and L lineages present an exaggerated divergence in sympatry but similar phenotype when lineages evolved in allopatry.

As reported previously, S and L lineages have specialized in the use of one carbon source, acetate or glucose, respectively [12]. Such character displacement is supposed to minimize competition between sympatric divergent populations thus promoting long-term coexistence. Accordingly, both S and L lineages have evolved more and more divergent metabolic properties (Fig. 2). The growth rate of the L lineage on glucose has increased by 20% compared to the ancestor during the first 18,000 generations and remains as high after 50,000 generations (Fig. 2A). On the other hand, the growth rate of the S lineage on acetate has increased by about 50% at 18,000 generations and even exhibits a further 2-fold increase between 18,000 and 50,000 generations (Fig. 2B). Character displacement in these sympatric evolving lineages is obvious since both lineages not only exhibit increased growth rate on its specific carbon source but also decreased ability to use the alternative resource *e.g* the growth rates of the S lineage on glucose and of the L lineage on acetate decreased between 18,000 and 50,000 generations. Because competition for resources is supposed to favor character displacement, we then investigated the impact of allopatric propagation on their metabolic properties. After a short term (300 generations) allopatric evolution, the growth rate of the L lineage on glucose was not significantly changed whereas the S lineage increased its growth rate reaching a value close to the maximal value exhibited by the L clones. On the contrary, the short term allopatric evolution has led to an increase in growth rate on acetate for both lineages, with S and L clones converging to a similar phenotype that resembles an intermediary phenotype between highly specialized S and L clones isolated after 50,000 generations. In the absence of competition, it thus clearly

appeared that both ecotypes converge to similar phenotypes regarding resources usage. Moreover, these results also suggest that the allopatric evolution of the L lineage mainly results in the improvement of acetate but not glucose consumption whereas the allopatrically evolved S clones improved both their glucose and acetate utilization.

S and L lineages converge in resource use once competition is abolished.

Compared to their common ancestor, clones from the S and L lineages after 6,500 generations have been previously shown to consume glucose faster and to produce acetate [12]. Here, we compared the 18KL^{Sym} and 18KS^{Sym} clones for these two properties. When cultivated in mono-culture, these two clones both exhaust glucose by 5h but with a small superiority for the 18KL^{Sym} with a Δ_{Gluc} at middle exponential of -645 uM compared to only -437 μ M for the 18KS^{Sym}. These two clones also appeared highly similar regarding acetate production that begins after 2h of growth and reaches a similar maximal concentration after 4h. However, they largely differ later on since acetate concentration remains high for many hours (~8h) in 18KL^{Sym} culture while 18KS^{Sym} rapidly consumes its own secreted acetate. Indeed, after exhausting the glucose by 5h, 18KS^{Sym} showed diauxic growth sustained by acetate consumption until it was depleted after 7h (Fig. 3A), whereas 18KL^{Sym} didn't even start to consume acetate at 8h (Fig 3D). These results confirm metabolic divergence between S and L cells with a more efficient acetate scavenging by the S lineage, owing to a faster metabolic switch from glucose to acetate. Because allopatric evolution was shown to promote phenotypic convergence of the S and L lineages, we then analyzed the metabolic kinetics of clones isolated from S and L populations evolved in allopatric conditions: 18KS^{Allo1A}, 18KS^{Allo4A}, 18KL^{Allo1A} and 18KL^{Allo4A} (Fig. 3). Regarding the two tested S clones (Fig. 3B and 3C), both improved glucose consumption with an increasing of the Δ_{Gluc} at middle exponential phase from -437 uM to -930 uM and -817 uM for 18KS^{Allo1A} and 18KS^{Allo4A}, respectively. These clones also improved on the alternative resource, acetate, which is faster depleted after glucose exhaustion, especially for 18KS^{Allo1A}.

Regarding the two tested 18KL^{Allo} clones (Fig. 3E and 3F), a similar improvement of glucose consumption leading to an earlier glucose depletion was observed. Most importantly, during allopatric evolution, these clones evolved the capacity to consume their own secreted acetate although less efficiently than any 18KS clone. Indeed 18KL^{Allo} clones start consuming acetate 2h after glucose is exhausted with a typical diauxic growth whereas 18KS clones exhibit an almost instantaneous switch from glucose to acetate utilization. In summary, 18KS^{Allo} and 18KL^{Allo} clones demonstrate strong phenotypic convergence for improvement of both glucose and acetate consumption. However, this detailed analysis of their metabolic properties also revealed that the S lineage maintains its specificity for faster switch from glucose to acetate consumption.

Discussion

Character displacement is an important ecological mechanism that can drive adaptive diversification in sympatry. Changes in the relative abundances of carbon sources including appearance of new resources owing to by-products secretion, may produce new selective pressures for divergence accompanied with shifts in characters including phenotype and metabolite uptake rates. These changes drove at first by competition will at last reduce competition between populations that will specialize for distinct resources and allow the establishment of ecotypes that occupy separated ecological niches. Interspecific competition is believed to be the cause of character displacement, since it may create resource-use partitioning, the ultimate cause of phenotypic divergence [21].

A previous study, using experimental bacterial populations evolving in the presence of two carbon sources, has reported competition as a cause of character displacement [13]. Here we studied the role of competition in character displacement and the long-term co-existence by negative-frequency dependent selection using as a model one population of the LTEE where

two ecotypes emerged in sympatry owing to niche construction [12]. These two *in-vitro* evolution systems differ in several aspects. The model of Friesen et al. uses both glucose and acetate as carbon sources in a relatively high concentration from the beginning of the experiment, and the polymorphism quickly emerged after ~1,000 generations [11, 13]. Each ecotype specialized on one resource and the genetic nature of the acetate specialist can be explained by only one mutation in the gene *iclR* [22]. In our model, the polymorphism emerged on the basis of niche construction, since only glucose was supplemented as sole carbon source. Indeed, through evolutionary time, the L ecotype evolves high glucose uptake rates accompanied with acetate overflow, which becomes a new ecological niche that is subsequently exploited by the S ecotype that emerged only after 6,000 generations owing to the successive fixation of three beneficial mutations in three global regulator genes [18, 23]. Later on, these ecotypes evolved to be niche specialists with fitness trade-off across the alternative niche [12]. Despite their differences, these two models where polymorphism emerged for one common ancestor also share fundamental similarities. In both cases, the causal role of competition in ecological character displacement was demonstrated: when interspecific competition is relaxed, exaggeratedly divergent ecotypes that arose in sympatry rapidly converge in phenotype and resource use and their frequency dependent interactions are lost. These studies also evidence the lack of robustness of polymorphisms arisen by character displacement owing to metabolic tradeoffs, since they can be reverted with a short cycle of allopatric evolution.

The long-term co-existence of the S and L lineages is due to negative-frequency dependent selection where sympatric evolution induces S and L specialization for only one carbon source, glucose for L and acetate for S. This specialization is further evident during the late course of sympatric evolution (from 18,000 to 50,000 generations) as the growth ability of each ecotype on its alternative substrate seems to degrade over evolutionary time (Fig. 2). The cause of this specialization trend can be attributed to the competition between sympatrically evolving

ecotypes since allopatric evolution of both S and L lineages rapidly gave rise to convergence to similar growth rates on both glucose and acetate. Indeed, a short-term allopatric evolution (300 generations), induces both the S and L lineages to converge: first to a higher growth rate on acetate but with a more pronounced improvement for the L lineages than for the S lineages that were already acetate specialists after the initial sympatric evolution (Fig. 2); second to an optimal growth rate on glucose that L lineages already reached before allopatric evolution whereas acetate-specialist S lineages took advantage of the allopatric evolution to evolve this trait. Accordingly, the absence of sympatric competition thus appeared to release some constraints on the evolutionary pathways accessible to the S and L lineages evolving in sympatry. In allopatry: the L lineages remained glucose specialists but rapidly evolved the capacity to exploit the acetate ecological niche which is no longer occupied by the S; the S lineages rapidly evolved more generalist traits by resuming initial acetate specialization trade-offs minimizing their glucose usage while concomitantly improving their growth rate on acetate. Going back to the selective pressures that sympatric competition imposes on S and L lineages evolution, we may assume that: L cells are mostly devoided of any pressure by the S cells owing to their already optimized carbon source usage in their own ecological niche (glucose); on the contrary, S cells always have to optimize their acetate-specialized traits to prevent colonization of their own ecological niche by the L lineage. In summary, it appears that sympatric competition between S and L lineages mainly imposes character displacement of the S cells to counteract highly probable invasion of their own ecological niche by the competing L ecotype. Such assumption is reinforced by the fact that S cells isolated after 18,000 generations of sympatric evolution (18KS^{sym} with an already so high acetate specialization) retain their ability to invade allopatrically evolved L clones (18KL^{Allo}) even if the later took advantage of allopatric evolution to improve their acetate scavenging. Accordingly, previous results have demonstrated that L cells encroached on S ecological niche [15], highly suggesting

that L cells mainly impose character displacement to S-cells that have to optimize their acetate usage to preserve their own ecological niche.

Conclusions

This work shows that long-term coexistence largely relies on continual competition between sympatric species since, when eliminated by means of allopatric evolution, frequency-dependent interactions between ecotypes are rapidly lost. The absence of competitors was shown to allow both the S and L lineages to converge to similar phenotypes on the two available carbon sources demonstrating that competition is the main driver of ecological character displacement. However, this convergence appears to evolve to the detriment of the S lineages which were systematically out-competed by the L lineages after allopatric propagation indicating that the L rapidly encroached on the S ecological niche. Accordingly, sympatric competition most probably imposed S cells to further improve their acetate scavenging to prevent exploitation of this ecological niche by the L whereas the optimal glucose usage of the L prevents any colonization of their niche.

List of abbreviations

NA

Ethics approval and consent to participate

NA

Consent for publication

NA

Availability of data and material

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' contributions

JC performed all experiments; DR performed the allopatric evolution experiments JC, DR, JG, TH, and DS analyzed the data; JC and DS designed the experiments; REL provided all strains; JC and DS wrote the manuscript, which was edited by all the authors

Acknowledgements

J.C. acknowledges the French ministry of education and research for her fellowship

References

1. Schluter D: **Ecological character displacement in adaptive radiation.** *American Naturalist* 2000, **156**:S4-S16.
2. Brown WL, Wilson EO: **CHARACTER DISPLACEMENT.** *Systematic Zoology* 1956, **5**(2):49-64.
3. Abrams PA, Cortez MH: **Is competition needed for ecological character displacement? Does displacement decrease competition?** *Evolution* 2015, **69**(12):3039-3053.
4. Losos JB: **Ecological character displacement and the study of adaptation.** *PNAS* 2000, **97**(11):5693-5695.
5. Kassen R: **Toward a General Theory of Adaptive Radiation Insights from Microbial Experimental Evolution.** In: *Year in Evolutionary Biology 2009*. Edited by Schlichting CD, Mousseau TA, vol. 1168. Malden: Wiley-Blackwell; 2009: 3-22.
6. Slatkin M: **ecological character displacement.** *Ecology* 1980, **61**(1):163-177.
7. Taper ML, Case TJ: **Quantitative genetic models for the coevolution of character displacement.** *Ecology* 1985, **66**(2):355-371.
8. Doebeli M: **An explicit genetic model for ecological character displacement.** *Ecology* 1996, **77**(2):510-520.

9. Melville J: **Competition and character displacement in two species of scincid lizards.** *Ecology Letters* 2002, **5**(3):386-393.
10. Grant PR, Grant BR: **Evolution of character displacement in Darwin's finches.** *Science* 2006, **313**(5784):224-226.
11. Friesen ML, Saxer G, Travisano M, Doebeli M: **Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*.** *Evolution* 2004, **58**(2):245-260.
12. Großkopf T, Consuegra J, Gaffé J, Willison JC, Lenski RE, Soyer OS, Schneider D: **Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment.** *BMC Evol Biol* 2016, **16**(1):163.
13. Tyerman JG, Bertrand M, Spencer CC, Doebeli M: **Experimental demonstration of ecological character displacement.** *BMC Evolutionary Biology* 2008, **8**:9.
14. Lenski RE, Rose MR, Simpson SC, Tadler SC: **long-term experimental evolution in *escherichia-coli* .1. Adaptation and divergence during 2,000 generations.** *American Naturalist* 1991, **138**(6):1315-1341.
15. Le Gac M, Plucain J, Hindre T, Lenski RE, Schneider D: **Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*.** *PNAS* 2012, **109**(24):9487-9492.
16. Rozen DE, Lenski RE: **Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism.** *American Naturalist* 2000, **155**(1):24-35.
17. Rozen DE, Schneider D, Lenski RE: **Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism.** *Journal of Molecular Evolution* 2005, **61**(2):171-180.
18. Plucain J, Hindre T, Le Gac M, Tenaillon O, Cruveiller S, Medigue C, Leiby N, Harcombe WR, Marx CJ, Lenski RE *et al*: **Epistasis and Allele Specificity in the**

- Emergence of a Stable Polymorphism in Escherichia coli.** *Science* 2014, **343**(6177):1366-1369.
19. Rozen DE, Philippe N, de Visser JA, Lenski RE, Schneider D: **Death and cannibalism in a seasonal environment facilitate bacterial coexistence.** *Ecology Letters* 2009, **12**(1):34-44.
 20. Hindre T, Le Gac M, Plucain J, Wielgoss S, Gaffe J, Schneider D: **Evolution in action: dream or reality?** *Biofutur* 2010(316):52-56.
 21. Stuart YE, Losos JB: **Ecological character displacement: glass half full or half empty?** *Trends in Ecology & Evolution* 2013, **28**(7):402-408.
 22. Spencer CC, Bertrand M, Trivisano M, Doebeli M: **Adaptive diversification in genes that regulate resource use in Escherichia coli.** *Plos Genetics* 2007, **3**(1):6.
 23. Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, Sniegowski PD: **Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations.** *Genetics* 2002, **162**(2):557-566.
 24. Cooper TF, Rozen DE, Lenski RE: **Parallel changes in gene expression after 20,000 generations of evolution in Escherichiacoli.** *Proc Natl Acad Sci U S A* 2003, **100**(3):1072-1077.
 25. Ostrowski EA, Woods RJ, Lenski RE: **The genetic basis of parallel and divergent phenotypic responses in evolving populations of Escherichia coli.** *Proceedings of the Royal Society B-Biological Sciences* 2008, **275**(1632):277-284.
 26. Koo MS, Lee JH, Rah SY, Yeo WS, Lee JW, Lee KL, Koh YS, Kang SO, Roe JH: **A reducing system of the superoxide sensor SoxR in Escherichia coli.** *Embo Journal* 2003, **22**(11):2614-2622.
 27. Li QX, Dowhan W: **structural characterization of escherichia-coli phosphatidylserine decarboxylase.** *Journal of Biological Chemistry* 1988, **263**(23):11516-11522.

Figures

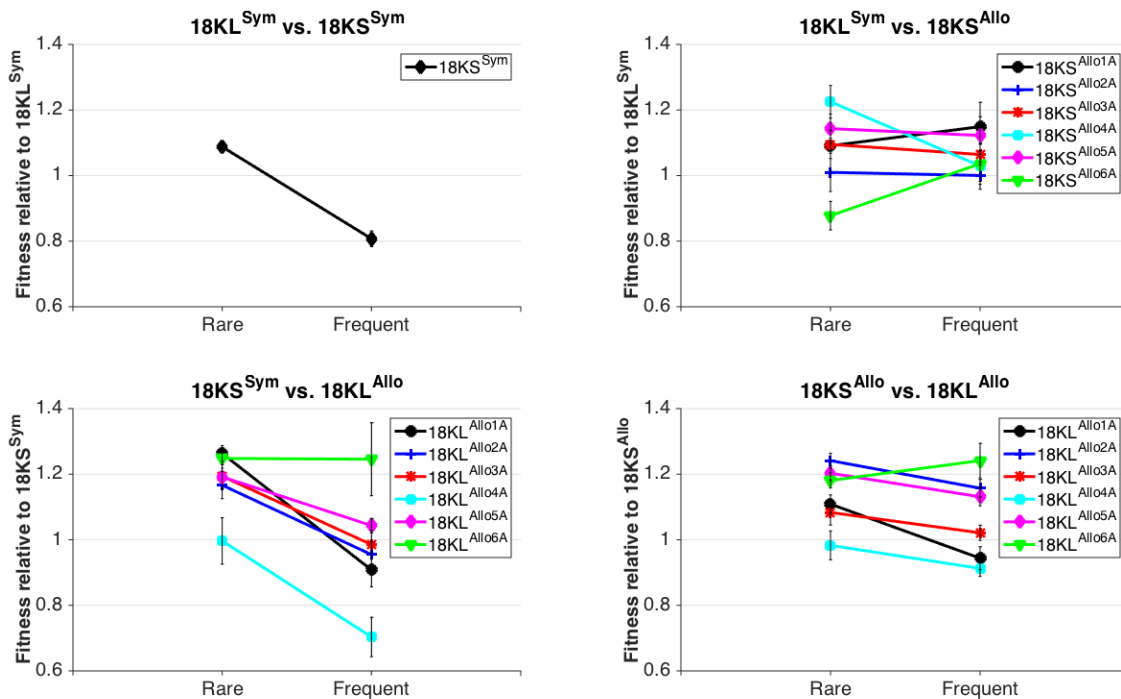


Fig.1 Frequency dependent interactions between 18KL^{Sym}, 18KS^{Sym}, 18KL^{Allo} and 18KS^{Allo} clones. (A) Typical pattern of negative-frequency dependent interaction between 18KL^{Sym} vs. 18KS^{Sym}. Results from competitions between (B) 18KL^{Sym} vs. each of the 18KS^{Allo} clones. (C) 18KS^{Sym} vs. each of the 18KL^{Allo} clones, and (D) 18KS^{Allo} clones vs. 18KL^{Allo} clones.

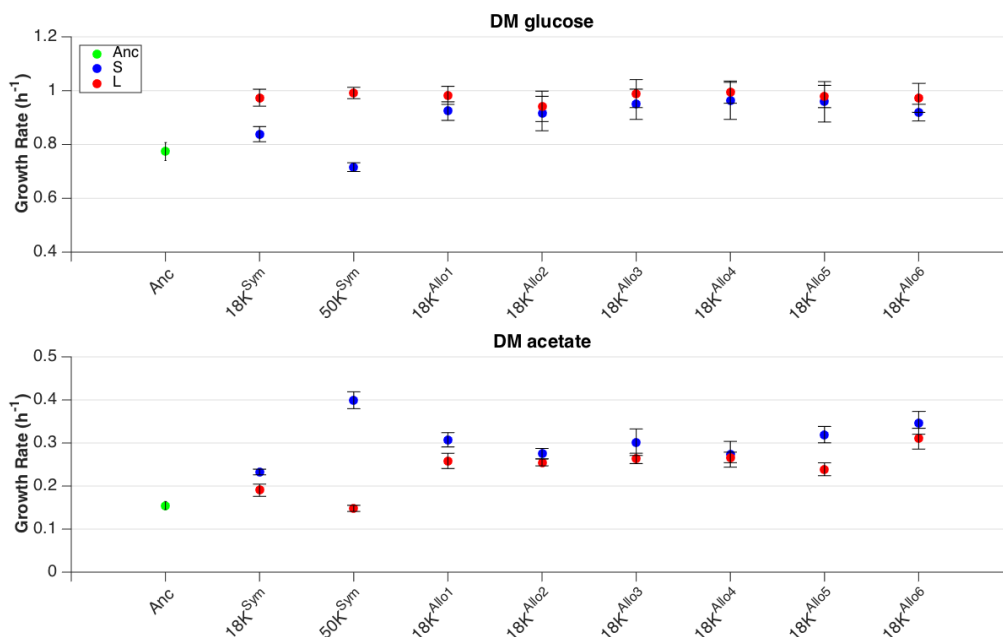


Fig. 2 Growth rates (h^{-1}) on glucose (A) and acetate (B) of the Anc (green), S (blue), L (red) evolved in sympatry or allopatry.

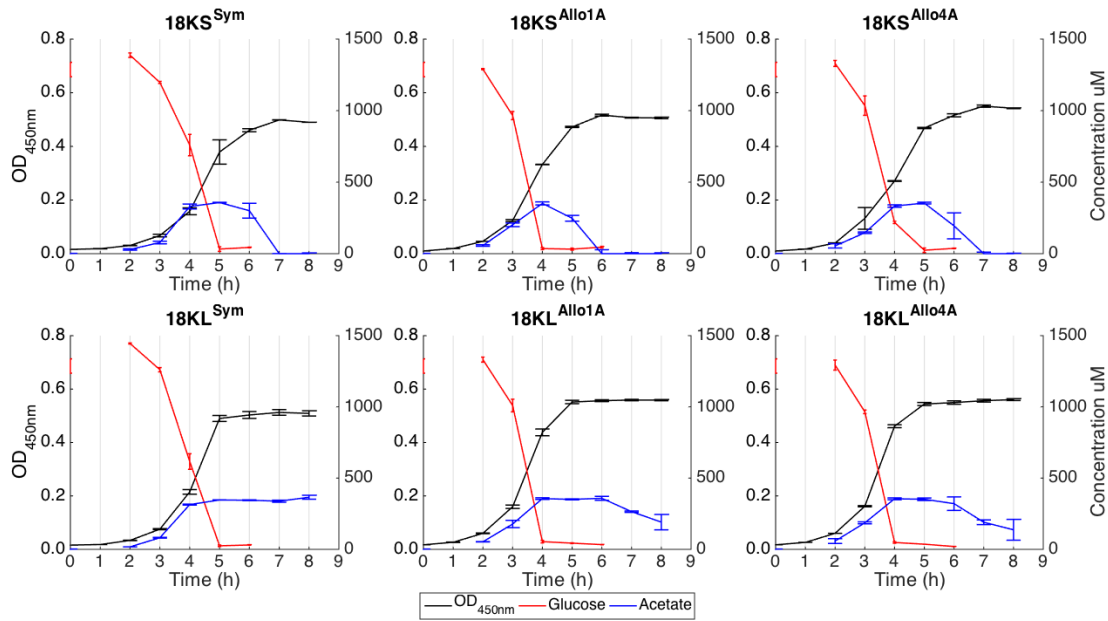


Fig.3 Metabolites and OD_{450nm} kinetics during 9 h growth on DM250-glucose. (A) 18KS^{Sym}, (B) 18KS^{Allo1A}, (C) 18KS^{Allo4A}, (D) 18KL^{Sym}, (E) 18KL^{Allo1A}, (F) 18KL^{Allo4A}. The black, red, and blue curves show OD_{450nm}, glucose concentration, and acetate concentration, respectively.

**Chapter 3: Molecular Genetics of a New
Ecological Opportunity Exploitation During
Long-Term Bacterial Sympatric Adaptive
Diversification**

As part of my interest for the understanding of the ecological characteristics of the S and L emergence and co-existence, I wanted to characterize the molecular mechanisms that drove this diversification event. This subject will be developed in this chapter, which is presented as a paper entitled, “*Molecular genetics of a new ecological opportunity exploitation during long-term bacterial sympatric adaptive diversification*” (Consuegra J, Plucain J, Gaffé J, Lenski RE, Hindré T, Schneider D) which will be submitted to the Journal *mBio*.

Previous work about the genetic bases of the S lineage emergence showed that three mutations in each of three global regulatory genes (*spoT*, *arcA*, *gntR*) were sufficient to reproduce the S-lineage specific traits in the ancestral background: growth on the L-secreted by-product (identified as acetate during my PhD work, see chapter 1), negative frequency-dependent interactions with L, and long-term maintenance (Plucain et al. 2014). In this previous study however, the involvement of each of the three mutations in the S-specific traits was not strictly established, although the *spoT* and *gntR* mutations were shown to be involved respectively in fitness increase and negative frequency-dependent interactions between L and S (Plucain et al. 2014). Moreover, all involved mutations were not identified since the combination of the three previous mutations in the ancestral genetic background only partially produced the S-specific phenotypic traits.

As described in chapter 1 (Großkopf et al. 2016), the S and L emergence and co-existence relied on acetate cross-feeding. Acetate secretion by L generated a new ecological niche that S exploited more efficiently. The S lineage is characterized by efficient acetate scavenging and faster metabolic switch from glucose to acetate consumption. The main objective of this chapter is: i) to study the molecular basis of acetate consumption by evolved clones from the S lineage in the population Ara-2 focusing on the role of ArcA in the polymorphism establishment, and ii) to identify new mutations and genes involved in the emergence and maintenance of the S/L polymorphism. We used an experimental approach including genome analyses, the construction

of isogenic strains of the ancestor and evolved clones from the S lineage, analysis of promoter activities and DNA-protein interactions.

We identified a new S-specific mutation, *acs^S*. We showed that, together with the evolved *arcA* allele from the S lineage (*arcA^S*), *acs^S* was involved in the ability of the S lineage to invade and then co-exist with the L lineage (Figs. 2 and 3 of the following paper). The two *arcA* and *acs* genes control central carbon metabolism and acetate consumption, respectively. We showed that: i) the two *arcA^S* and *acs^S* mutations were essential for the emergence of the adaptive diversification event and especially for acetate consumption, ii) they are involved in the early establishment of the S lineage, and iii) strong epistatic interactions between *arcA^S* and other mutations from the S lineage contributed to the exploitation of this new ecological opportunity (Fig. 6 of the following paper).

We also identified the molecular mechanism of *arcA^S*. Indeed, the corresponding L93F change in the evolved ArcA protein reduced its DNA-binding ability and thereby its regulatory properties (Fig. 4 of the following paper). These molecular changes drove an early restructuring of the transcriptional control of central metabolism leading to improved acetate consumption by the emerging S lineage (Figs. 5 and 6 of the following paper).

In conclusion, we showed that rewiring central metabolic pathways through transcriptional changes of target genes and epistatic interactions were essential in the establishment of bacterial diversification.

**Molecular Genetics of a New Ecological Opportunity Exploitation During Long-Term
Bacterial Sympatric Adaptive Diversification**

Jessika Consuegra,^{a,b} Jessica Plucain,^{a,b*} Joël Gaffé,^{a,b} Richard E. Lenski,^{c,d} Thomas
Hindré,^{a,b,#} Dominique Schneider^{a,b}

Univ. Grenoble Alpes, Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité -
Informatique, Mathématiques et Applications, Grenoble (TIMC-IMAG), F-38000 Grenoble,
France^a; Centre National de la Recherche Scientifique (CNRS), TIMC-IMAG, F-38000
Grenoble, France^b; Department of Microbiology and Molecular Genetics, Michigan State
University, East Lansing, USA^c; BEACON Center for the Study of Evolution in Action,
Michigan State University, East Lansing, USA^d

Running Head: Molecular Genetics of Sympatric Diversification

#Address correspondence to T. Hindré, thomas.hindre@univ-grenoble-alpes.fr

*Present address: ETH Zurich, Institute of Integrative biology, 8092 Zurich, Switzerland.

Abstract word count: 250 words

Text word count: 6045 words

Abstract

Adaptive diversification events are essential evolutionary processes since they produce phenotypic innovation including colonization of novel ecological niches and further adaptation. In bacteria, they may arise in sympatry owing to ecological mechanisms such as tradeoff and character displacement that generate new ecological opportunities leading to divergence. However, less information is available at the genetic level. The longest adaptive diversification event ever detected during laboratory evolution experiments occurred during the Long-Term Evolution Experiment (LTEE) where twelve populations are independently propagated since about 65,000 generations from the same common ancestor of *Escherichia coli*. In one population called Ara-2, two lineages, S and L, emerged from the same ancestor at about 6500 generations and dynamically coexisted ever since. Both acetate secretion by L, which provided a new ecological opportunity, and the establishment of frequency-dependent interactions resulted in the emergence of the S lineage. Here, we characterized two S-specific mutations that provided this lineage with the ability to invade and then coexist with the L lineage. They affect the *arcA* and *acs* genes which control central carbon metabolism and acetate consumption, respectively. We showed that these mutations are together essential for the emergence of the adaptive diversification event by driving early restructuring of the transcriptional control of central metabolism leading to improved acetate consumption by S. Moreover, strong epistatic interactions between the *arcA* allele substituted in Ara-2 and other mutations from the S lineage contributed to the exploitation of this new ecological opportunity. The emergence and maintenance is the result of a complex multi-step process.

Importance

In sexual organisms, geographical isolation may result in divergence and even speciation events. In bacteria, this process of adaptive diversification may also occur in sympatry, *i.e.* in the same environment. The emergence of new ecological niches is involved in this divergence but its genetic bases are still poorly understood. Here we investigated the longest adaptive diversification event that occurred during experimental evolution in *Escherichia coli* and that resulted in the emergence and coexistence of two phenotypically differentiated lineages. We characterized two mutations that allowed one of the diverging lineages to exploit a new ecological opportunity provided by the other lineage. We showed that rewiring central metabolic pathways through transcriptional changes of target genes and epistatic interactions were essential in the establishment of bacterial diversification.

Introduction

Adaptive diversification events are central issues in evolutionary biology (1). Organismal divergence produces phenotypic innovation including colonization of novel ecological niches, speciation at the long term and adaptation to environmental conditions. Divergence is also a public health concern since the control of infections is strongly related to the variability of bacterial populations (2). In sexual organisms, adaptive diversification may be driven by geographical isolation (allopatry), thereby restricting genetic exchange and producing reproductive barriers. In asexual organisms, evolution of divergent lineages is possible even in a single given environment (sympatry), if ecological opportunities exist including spatial organization (3) and/or nutritional diversity (4-8). Even in homogeneous environments, niche construction whereupon the bacterial organisms modify themselves their environment may drive the emergence of polymorphisms (4, 7, 9). Lineage divergence has been shown to involve tradeoffs for resources together with character displacement that result in negative frequency-dependent interactions between the different lineages (1, 3, 4, 7, 8, 10).

Besides these ecological causes, the genetic and molecular bases leading to sympatric divergence of organisms are mostly unknown except in the cases of heterogeneous environments and when single mutations resulted in the emergence of polymorphic states (3, 4). Even if entire genomes of clones belonging to diverging lineages have been sequenced, the identification of the relevant mutations is still missing (11, 12). In a single case that occurred during the Long-Term Evolution Experiment (LTEE) with *Escherichia coli*, a combination of mutations has been identified that resulted in sympatric divergence (13). Even then, the precise role of each mutation is mostly unknown and additional mutations are likely to be involved to produce the precise phenotypic traits of the diverging lineages, highlighting the need for more detailed genetic and molecular studies to fully dissect the different steps producing complex and dynamic polymorphisms.

During the LTEE, twelve independent populations are propagated from a common ancestor of *E. coli* B REL606 (14) by serial daily transfers in a glucose-limited environment (15) amounting for more than 60,000 generations of evolution. In one of the 12 populations, called Ara-2, a unique diversification event occurred where two different lineages, named L (Large) and S (Small) according to their colony and cell sizes, emerged from a common ancestor at ~6500 generations (7, 13, 16, 17). The two phenotypically-differentiated lineages coexist ever since, owing to negative frequency-dependent interactions whereby each lineage is favored and invades the other when rare, precluding the extinction of any of the two lineages. The L lineage consumes better the glucose with higher growth rates than the S lineage that exploits a resource corresponding to a by-product that is secreted by L (7). We recently showed that this secreted metabolite was acetate and that maintenance was favored by an ecological dynamics involving character displacement and trade-offs with the L and S lineages becoming more and more adapted to glucose and acetate, respectively (18). These ecological mechanisms prevented simultaneous evolution of either lineage on both carbon sources.

In *E. coli*, acetate is degraded through two different metabolic pathways depending of its extracellular availability. In the presence of large amounts of acetate, the low-affinity pathway ACKA/PTA produces acetyl-CoA through an acetyl-phosphate intermediate (19). To scavenge low acetate amounts, which is the case in the population Ara-2 of the LTEE (18), *E. coli* uses a high-affinity pathway in which acetate is activated to acetyl-CoA by the acetyl-CoA synthase enzyme encoded by the *acs* gene (19). Acetyl-CoA is then used to produce energy through the tricarboxylic acid (TCA) cycle in which the aconitase B enzyme encoded by *acnB* is mandatory for either acetate assimilation (20) or anabolism purposes via the glyoxylate shunt (19-21). In that case, malate synthase A encoded by *aceB* catalyzes the irreversible condensation of acetyl-CoA and glyoxylate to produce (S)-malate (21).

Combined analyses of genome sequences (13) and global transcription profiles (22) of early evolved clones from both L and S lineages revealed that 3 mutations affecting the *spoT*,

arcA and *gntR* genes were sufficient, when introduced together in the ancestral genetic background, to mimic the phenotypic traits of the S lineage (consumption of the L-secreted by-product *i.e.* acetate, negative frequency-dependent interactions with L, and coexistence with L), although to a lesser extent than typical evolved clones from the S lineage (13). Therefore, additional mutations are likely to be involved in the emergence of the L and S polymorphism. Moreover, the role of each of the 3 mutations in the specific S-associated phenotypic traits was not strictly established, although the *spoT* and *gntR* mutations were shown to be involved, respectively, in fitness increase and negative frequency-dependent interactions between L and S (13). The order of substitution and precise alleles of the mutations were also shown to be important in the establishment of the L and S lineages. Here, we investigated the role of the *arcA* mutation in the emergence of the L and S polymorphism and identified a fourth mutation involved in that evolutionary process.

The ArcA/ArcB two-component system plays a key role in the regulation of the central metabolism in response to oxygen availability (23). Upon activation by its cognate sensor ArcB and homodimerization, the ArcA response regulator modulates the transcription of hundreds of *E. coli* genes, including operons related to metabolism, respiratory control and chromosome replication (24, 25). The S-specific *arcA* mutant allele affects a position encoding an amino-acid in the protein that mediates ArcA dimerization upon activation by ArcB (26) but its functional consequence has not yet been investigated.

Here we investigated the molecular basis of acetate consumption by evolved clones from the S lineage in the population Ara-2 of the LTEE. As the three previously mentioned genes involved in acetate assimilation (*acs*, *acnB* and *aceB*) are known to be repressed by ArcA (25), we hypothesized that the evolved *arcA* allele substituted in the S lineage, called *arcA^S* here, would be a good candidate to cause this phenotypic innovation. We showed that this allele produced changes in the transcriptional control of key genes that are involved in acetate assimilation and that it occurred at an evolutionary time that allowed the emergence of the S

lineage by inducing higher acetate consumption. Moreover, we demonstrated pervasive epistatic interactions between *arcA^S* and other mutations that occurred in pre-S evolved clones that were on the line of descent leading to the S lineage. Finally, scrutinizing the available L and S genome sequences allowed us to identify an S-specific mutation in the transcriptional regulatory region of *acs* (*acs^S*) that is involved in the emergence of the S lineage.

Results

Improvement of acetate assimilation and the emergence of the S lineage. Improved acetate consumption by the S lineage was shown to be essential for the exploitation of the new ecological niche provided by the L lineage (18). We therefore tested whether this phenotypic innovation was concomitant with the establishment of negative frequency dependent interactions between L and S. We compared to the ancestor REL606 (called Anc) the consumption of acetate as a growth substrate of two evolved clones (13, Table 1): an S clone isolated at 6500 generations, called 6.5KS1, and a pre-S clone, called 6K3, isolated at 6000 generations which lies on the line of descent of the S lineage but does not yet show negative frequency-dependent interactions with early evolved L clones. With acetate as the sole carbon source, the pre-S clone 6K3 had a shorter lag phase, higher growth rate and higher yield than Anc (Fig. 1A). When grown in a mixture of glucose and acetate, 6K3 revealed a diauxic growth with faster switching to acetate consumption after exhaustion of glucose (Fig. 1B). These phenotypic traits were further slightly improved in 6.5KS1.

Despite these improved acetate-associated traits, the pre-S 6K3 clone did not exhibit negative frequency-dependent interactions with L clones by contrast to 6.5KS1 (13). Therefore, higher acetate consumption emerged at least 500 generations before the S lineage was fully established in the Ara-2 population and was further improved later on during evolution (18). Moreover, it confirms the involvement of several mutations for the emergence of the polymorphism.

Involvement of ArcA in the ability of the S lineage to consume acetate. To investigate the role of the evolved *arcA* allele from the S lineage (*arcA^S*) in acetate consumption, we compared the growth profiles, in a minimal medium containing acetate as the sole carbon source (DM250-acetate), of 10 strains (Table 1) including the 3 Anc, 6K3, 6.5KS1 clones and 7 isogenic strains except for the *arcA* allele: Anc with an in-frame *arcA* deletion (*Anc Δ arcA*), and with each of three *arcA* alleles from the LTEE (13), including the one from

the Ara-2 S lineage and two from the populations Ara+1 and Ara+2 (Anc::*arcA*^S, Anc::*arcA*^{Ara+1}, Anc::*arcA*^{Ara+2}, respectively); the 6K3 clone with the ancestral *arcA* allele (6K3::*arcA*^{Anc}); and the 6.5KS1 clone with each of the two ancestral and Ara+1 *arcA* alleles (6.5KS1::*arcA*^{Anc} and 6.5KS1::*arcA*^{Ara+1}, respectively). We confirmed that both evolved 6K3 and 6.5KS1 clones exhibited improved growth traits compared to Anc (Fig. 2).

Four main trends emerge from these data (Fig. 2): i) deletion of *arcA* in the ancestral genetic background improved the growth rate, consistent with the repression of the genes involved in acetate consumption by ArcA (25); ii) introduction of the three evolved *arcA* alleles in the ancestor chromosome improved the growth rate, suggesting decreased activity of the ArcA evolved proteins; iii) introduction of the ancestral *arcA* allele in the evolved clones abolished the improvement of acetate consumption back to the level of the ancestor strain for 6.5KS1 and even completely abolished growth for 6K3; iv) as 11 of the 12 populations of the LTEE have mutations in either *arcA* or *arcB* (13), we investigated the effect of other evolved *arcA* alleles on the growth profiles. The three evolved *arcA* alleles from Ara-2 S, Ara+1 and Ara+2 had the same effect on acetate consumption in the ancestral genetic background, whereas the Ara-2 S allele conferred a higher growth rate than the Ara+1 allele in the 6.5KS1 genetic background. This difference was significant ($P < 0.0001$, unpaired *t*-test).

We next checked the ability of the same strains to grow on L-conditioned supernatant which is a phenotypic hallmark of the S lineage (Fig. 3). The results were almost identical to the data presented in Fig. 2. We confirmed that 6.5KS1 consumed the by-product secreted by 6.5KL4 and demonstrated that the pre-S clone 6K3 also had this property. By contrast, Anc was unable to grow in the L-conditioned supernatant except when harboring *arcA*^S. Replacing *arcA*^S by its ancestral counterpart in 6K3 and 6.5KS1 abolished and reduced, respectively, their growth ability. Finally, whereas the impact of each of the three evolved *arcA* alleles in the ancestral background was indistinguishable, only *arcA*^S conferred an optimal growth rate on L-conditioned supernatant.

Therefore, the evolved *arcA^S* allele is mandatory for a better consumption of acetate in the S lineage, and resulted in reduced activity of the ArcA protein (see also next section). Moreover, several lines of evidence suggest that additional genetic determinants are necessary and interact epistatically with the evolved *arcA^S* allele: introducing *arcA^S* in the ancestral background did not confer optimal growth rate in the presence of acetate or L-conditioned supernatant, the effect of *arcA^S* was dependent on the evolved genetic background (compare 6K3::*arcA^{Anc}* and 6.5KS1::*arcA^{Anc}* in Fig. 2), and allelic divergence for different *arcA* mutations was not detected in all genetic backgrounds. The evolved *arcA^S* allele conferred therefore specificity only in an evolved background and was therefore contingent on other mutations. Indeed, for instance the Ara+1 and Ara+2 populations have been reported to have an improved growth rate in acetate, relative to the ancestor, at 20,000 and 50,000 generations (28) but no adaptive diversification and long-term coexistence of different lineages have been observed in these populations.

DNA-binding ability of ArcA^S. After phosphorylation by ArcB, ArcA dimerizes and binds to the promoter regions of its target genes (25, 26). Because the L93F change that occurred in the evolved ArcA^S protein affects an amino-acid localized in the dimerization domain, we purified both ArcA^{Anc} and ArcA^S proteins, and performed gel shift assays to compare their DNA-binding abilities (Fig. 4). As a target DNA fragment, we used the transcriptional regulatory region of *fadE*, a gene that is well known to be regulated by ArcA (29). A retarded band was observed only with the phosphorylated ArcA^{Anc} protein, showing its DNA binding ability to a known regulated promoter region. However, using the same protein concentration, no binding was detected for the phosphorylated ArcA^S protein indicating reduced affinity for DNA (Fig. 4A). This is consistent with the evolved *arcA^S* allele conferring phenotypic effects similar to a deletion allele (Fig. 2).

To estimate the dissociation constant (K_d) for each of the ArcA^{Anc} and ArcA^S proteins, we repeated the EMSA assay with different concentrations of each protein (Fig. 4B, C). We

calculated the Kd of the ArcA^{Anc}-P and ArcA^S-P proteins for the *fadE* promoter region to be 0.7 and 1.2 μ M, respectively. Therefore, the evolved *arcA*^S allele resulted in reduced DNA-binding ability of the corresponding protein suggesting that it may subtly affect the regulation at ArcA-regulated promoters. We therefore next investigated the effect of *arcA*^S on the transcription level of genes known to be involved in acetate metabolism and regulated by ArcA.

Overexpression of *acnB* and *aceB* in the S lineage. Acetate assimilation relies on both acetate activation by the acetyl-CoA synthase encoded by *acs* (see next section) and an active TCA cycle and glyoxylate shunt. We therefore investigated the impact of the evolved ArcA^S protein on the transcription level of the two genes *acnB* and *aceB* that are essential for acetate catabolism and that are known to be repressed by ArcA (24, 25). We assessed their promoter activities by using GFP transcriptional fusions (see Materials and Methods) in genetic backgrounds harboring or not *arcA*^S (Fig. 5). Both *acnB* (Fig. 5A) and *aceB* (Fig. 5B) promoters have a 2- to 4-fold higher activity in 6.5KS1 compared to Anc. Replacing the evolved *arcA*^S allele by its ancestral counterpart in 6.5KS1 eliminated this up-regulation for *acnB*, and even abolished the *aceB* promoter activity during the first 6 hours of growth. Increased transcription of *acnB* and *aceB* is therefore caused by *arcA*^S which is consistent with the previous results showing that the ArcA^S protein had reduced activity compared to ArcA^{Anc}. These results confirm the role of *arcA*^S in the transcriptional changes of metabolic genes that further sustains acetate assimilation in the S lineage.

In the ancestral genetic background, we detected different behavior for both *acnB* and *aceB* genes (Fig. 5). Indeed, while *arcA*^S had no obvious effect on the transcription level of the *acnB* promoter, it strongly reduced transcription from the *aceB* promoter. Therefore, consistently with our previous results, these data suggest the occurrence of epistatic interactions between *arcA*^S and other mutations that occurred during 6500 generations and that were essential to the emergence of the S lineage.

Temporal changes in *acs* transcription in the S lineage. By scrutinizing evolved genomes from population Ara-2 (13, 30), we detected an S-specific mutation in the promoter region of *acs* (C>A at position -86 with respect to the first nucleotide of the *acs* translation codon). This mutation was present in the pre-S clone 6K3 (Table 1) and then shared by all S clones from 6500 generations.

First, we tested the effect of the evolved *acs*^S allele on *acs* transcription with transcriptional fusions of the *acs*^{Anc} and *acs*^S promoter regions (pUA66::*acs*^{Anc} and pUA66::*acs*^S). We introduced each of the two plasmids in Anc and 6.5KS1, and monitored the activities of the two *acs* promoters during growth in minimal medium with glucose as the sole carbon source (Fig. 6). Comparing Anc and 6.5KS1, we observed two main differences for the two promoters: i) in 6.5KS1, transcription was activated earlier after 5.5 h of growth and only transiently for less than 2 h (versus 7 h and for at least 12 h in Anc, respectively, Fig. 6A, C); ii) *acs*^S increased transcription by 2-fold in Anc, but reduced it in 6.5KS1.

Second, we tested the effect of the evolved *arcA*^S allele on *acs* transcription by comparing the activity of both the *acs*^{Anc} promoter in Anc and Anc::*arcA*^S, and the *acs*^S promoter in 6.5KS1 and 6.5KS1::*arcA*^{Anc}. In the ancestral background, *arcA*^S resulted in earlier activation without tremendously affecting the transcription level (Fig. 6A). By contrast in 6.5KS1, *arcA*^S resulted in a 5-fold decrease of the transcription level.

Both *acs*^S and *arcA*^S had therefore positive effects on *acs* transcription only in the ancestral background and not anymore in 6.5KS1, suggesting that the positive effect of *arcA*^S, and possibly of *acs*^S, on acetate consumption was not associated with increased *acs* transcription. An alternative hypothesis may however be that *acs* transcription was only transiently increased by *acs*^S and *arcA*^S over evolutionary time.

We therefore analyzed the activity of the *acs*^{Anc} and *acs*^S promoters in the pre-S 6K3 clone, which carries both *acs*^S and *arcA*^S (Table 1), and its derivative 6K3::*arcA*^{Anc} (Fig. 6B). The evolved *acs*^S allele strongly increased, by ~5 fold, the duration of transcription activation

of *acs*, while *arcA^S* resulted in earlier transcription activation by 2 hours. This is likely to be important in the emergence of the S lineage as we recently demonstrated fast switching from glucose to acetate consumption for the S clones (18). Therefore, *acs^S* and *arcA^S* alleles had similar effects in the two Anc and 6K3 strains, but we detected several transcriptional differences between 6K3 and 6.5KS1 (Fig. 6B, C): i) although the *acs^S* promoter was activated after 5.5 h of growth as in 6.5KS1, its transcription was longer and maintained until 10 h of growth in 6K3; ii) *arcA^S* strongly decreased the activation time and increased both the transcription level and duration of the *acs^S* promoter in 6K3; iii) *acs^S* increased *acs* transcription in 6K3 while decreasing it in 6.5KS1. Therefore, at least transiently during the emergence of the S lineage, *acs^S* and *arcA^S* were likely to allow better acetate consumption by positively affecting *acs* transcription. Moreover, and consistently with the previous sections, these results suggest epistatic interactions between *arcA^S* and other mutations that arose between 6000 and 6500 generations in the S lineage.

To summarize, improved acetate consumption in pre-S clones at 6000 generations may partly rely on *acs* transcription changes (earlier activation and higher activity) when *acs^S* and *arcA^S* were fixed. These changes are further tuned owing to epistatic interactions with other mutations before the establishment of negative frequency-dependent interactions between L and S at 6500 generations.

Discussion

Understanding the molecular bases of the emergence of stable polymorphism resulting in the coexistence of bacterial ecotypes with new phenotypic traits is an important challenge, not only from a fundamental point of view but also in both public health for the fight against the emergence of virulence and antibiotic resistance and biotechnology for the emergence of inefficient strains. Our study focused on the molecular mechanisms that allowed the establishment of a stable polymorphism and in particular the emergence of one of the two

ecotypes that coexist since more than 50,000 generations in one of the 12 populations of the LTEE. Adaptive diversification is expected, and occurred, in other experimental evolution experiments, especially in heterogeneous environments when either two carbon sources were present in the environment from the beginning of evolution (6, 8, 11, 12) or spatial structuration existed (3). It has also been observed in sympatry owing to niche construction (9) in high-glucose chemostats, where significant acetate overflow is produced by *E. coli* strains (4). Here, the polymorphism observed in one of the 12 population of the LTEE was unexpected owing to the presence of very low concentrations of glucose (25 mg/L) as the sole carbon source. The emergence of this polymorphism was previously shown to rely on the exploitation by the S lineage of a new ecological niche resulting from acetate secretion by the L lineage owing to three S-specific mutations. These mutations affected regulatory genes and were shown to partially confer the S-specific phenotypes (7, 13, 18). Here, we showed that mutations, affecting both the *acs* and *arcA* genes were essential for the S lineage to evolve a higher capacity to consume acetate, but only at the time of emergence of the S lineage. Further epistatic interactions complexified this phenotype-to-genotype association.

A mutation in the promoter region of *acs* (*acs*^S) was identified in the 6K3 pre-S clone at 6000 generations and was further shared specifically by all evolved clones from the S lineage. We showed that this mutation resulted in an upregulation of *acs* that encodes acetyl-CoA synthase which is involved in acetate consumption by producing acetyl-CoA from acetate (19). This mutation affects the promoter region of *acs* (C to A at position -86) and is located in the binding site CRP I of the CRP activator protein that overlaps the binding site FIS III of the Fis inhibitor protein. Fis competes successfully with CRP for binding to these overlapping sites, thereby reducing CRP-dependent *acs* transcription (33). In that previous study (33), the exact same mutation *acs*^S has been generated and shown to induce a 10-fold reduction of Fis binding resulting in a 2-fold increase of *acs* transcription. Consistently, we detected a 2-fold increase in *acs* transcription after introducing the *acs*^S mutation in the ancestral genetic background (Fig.

6A). An even stronger transcriptional activation was associated with the *acs*^S mutation in the 6K3 clone that most likely helped to enhance acetate consumption in the pre-S clones, thereby setting the stage for the emergence of the S lineage. Involvement of *acs* in adaptive diversification events was previously reported during 1750 generations of experimental evolution of *E. coli* K-12 in chemostats with minimal medium containing 125 mg/L of glucose as a sole carbon source (4). In this study, a polymorphism emerged with three ecotypes coexisting owing to acetate and glycerol cross-feeding. The acetate cross-feeding phenotype was associated with semi-constitutive overexpression of *acs*, by either IS element insertions or a single nucleotide substitution in the transcriptional regulatory region of *acs* (4). In our study however, two important differences can be observed. First, the L/S polymorphism emerged only after 6500 generations while it was almost immediately detected in the chemostat study. Second, the *acs*^S allele was associated with a mutation within *arcA* (*arcA*^S) which was not the case in the chemostat. This might be explained either by initial mutations that occurred before *acs*^S and that might interfere with its substitution or by very different concentrations of acetate in the two studies: ~50 μM in our case (18) versus ~6 mM in the chemostat study (34). The much lower acetate concentrations detected in the LTEE might explain that the acetate ecological niche was more difficult to access, thereby “needing” more time. In particular, we previously showed that a first mutation in *spoT* was mandatory for the polymorphism to emerge (13).

We showed that changes in *acs* transcription included earlier and increased promoter activity in the ancestral genetic background and, in addition, semi-constitutive activity in the pre-S clone 6K3 (Fig. 6). These changes not only relied on *acs*^S but also on *arcA*^S, an S-specific mutation affecting the *arcA* global regulatory gene and already present in 6K3. Moreover, *arcA*^S also resulted in increased transcription in the 6.5KS1 clone of the two *acnB* and *aceB* genes that are also involved in acetate consumption (Fig. 5). Since it is well-known that the ArcA protein is a repressor of the transcription of these three genes, these data suggested decreased

activity of ArcA^S. We indeed showed this was the case since ArcA^S had a decreased DNA-binding activity compared to the ancestral protein (Fig. 4). The evolved *arcA^S* allele resulted in the L93F change in the evolved ArcA^S protein, altering its DNA-binding capacity (Fig. 4). Indeed, it has been shown that the L93 residue is localized in the dimerization domain of ArcA which is requested for its activity (23, 26, 35). We further showed that *arcA^S* improved growth of Anc, 6K3 and 6.5KS1 on both acetate and a supernatant from an L clone culture (Figs. 2 and 3). Therefore, both *acs^S* and *arcA^S* were involved in the emergence of the S lineage by improving the growth of the evolved clones from that lineage on acetate, the byproduct secreted by L. This was performed by more efficient transcription of genes involved in acetate consumption (*acs*, *acnB* and *aceB*), the evolved ArcA^S protein thereby promoting a restructuring of the metabolic network that resulted in the later emergence of the S lineage.

However, we showed that overexpression of *acs* did not explain by itself the establishment of the L/S polymorphism in the Ara-2 population for two main reasons: i) even if the pre-S clone 6K3, sampled at 6000 generations, harbored both *acs^S* and *arcA^S* alleles and exhibited higher acetate consumption, it was unable to invade when rare an L clone through negative frequency-dependent interactions (13); ii) while *arcA^S* was still essential for improved acetate consumption in the 6.5KS1 clone (Fig. 2), it did not improve anymore the transcription of *acs* in that clone and even strongly repressed it (Fig. 6). The same hold true for *acs^S* which had almost no effect anymore on *acs* transcription in 6.5KS1. Therefore, additional mutations that occurred between 6000 and 6500 generations in the S lineage interacted epistatically with both *acs^S* and *arcA^S* and further modified the metabolic abilities of S clones that culminated into the negative frequency-dependent interactions with L clones. Altogether, our results indicate that improved acetate consumption occurred in pre-S clones owing to both evolved *acs^S* and *arcA^S* alleles by more efficient transcription of *acs*. Emergence of the S lineage including the acquisition of negative frequency-dependent interactions with the L lineage then involved additional mutations that abolished the effect of *acs^S* and *arcA^S* on *acs* transcription.

At least one of these additional mutations has been shown previously to affect *gntR* that encodes the repressor of the genes encoding the enzymes of the Entner-Doudoroff pathway (13). Identification of all these additional mutations will necessitate a careful analysis since the Ara-2 population has evolved a hypermutator phenotype owing to a mutation in *mutL* that was substituted well before the establishment of the two L and S lineages (36). Therefore, hundreds of mutations are already present in evolved clones sampled after 6000 and 6500 generations.

We detected several evidences of pervasive epistatic interactions among mutations in the Ara-2 population that were involved in the emergence of the S lineage. First, we previously showed epistatic interactions between the three mutations in *spoT*, *arcA* and *gntR* that are involved in the emergence of the S lineage (13). Second, we showed here that the improvement of *acs* transcription that relied on both *acs^S* and *arcA^S* occurred in pre-S clones and was then abolished in the S lineage at 6500 generations. Third, we previously showed that only the evolved *arcA^S* allele conferred S-specific phenotypic traits and not evolved *arcA* alleles from other populations from the LTEE (13). We further showed here that this allelic specificity occurred only in an evolved background from the S lineage and not in the ancestor (Figs. 2 and 3). This may explain why other populations from the LTEE have been reported to have improved growth rates in acetate relative to the ancestor at 20,000 and 50,000 generations (28) without any adaptive diversification and long-term coexistence of different lineages in these populations. Fourth, *arcA^S* activated the transcription of *acnB* and *aceB* in 6.5KS1 which was not the case when introduced in the ancestral genetic background (Fig. 5). Therefore, epistatic interactions have occurred between *arcA^S* and mutations that were substituted before 6500 generations. By contrast, in all cases where stable polymorphisms in experimental bacterial populations have been analyzed in details, single mutations were responsible for the emergence of the co-existing ecotypes (4, 6, 12). The Ara-2 population has an additional complexity since three mutations have been previously shown to be requested for the emergence of the S lineage, and to interact epistatically (13). Here, we showed that at least one additional mutation (and

more likely several) and more epistatic interactions are involved. Therefore, polymorphisms arising from niche construction processes in environments with low concentrations of resources like glucose in the LTEE may involve complex genetic pathways to be activated.

Together with previous studies, we showed here that the emergence of the S lineage in the population Ara-2 from the LTEE is a multi-step process. The first step involved a mutation in the *spoT* gene which was involved in fitness increase through a yet unknown mechanism (13). Second, the evolved *acs*^S and *arcA*^S alleles allowed pre-S clones to improve acetate consumption by improving the transcription of at least *acs*, *acnB* and *aceB* genes, allowing the exploitation of this new ecological opportunity. In a third step, at least the evolved *gntR* allele conferred invasion abilities to the S lineage by establishing negative frequency-dependent interactions between L and S clones. Additional mutations are likely to be involved. Pervasive epistatic interactions have been detected between these mutations and other mutations that were substituted before 6500 generations, the time of phenotypic differentiation of the S lineage.

Materials and Methods

Bacterial strains and mutant construction. All strains derived from the LTEE, including the REL 606 ancestor (Anc) and the following evolved clones sampled from the polymorphic population Ara-2 (Table 1): 6K3, a pre-S evolved clone isolated at 6000 generations that lies on the line of descent leading to the S lineage; 6.5KS1, an S clone isolated at 6500 generations; and 6.5KL4, an L clone isolated at 6500 generations. Isogenic strains except for the two *Anc*Δ*arcA* and 6K3::*arcA*^{Anc} were constructed by allelic exchange during a previous study (13), using the suicide plasmid pKO3 (31).

Briefly, the construction of the strain *Anc*Δ*arcA* requested two PCR steps. First, two DNA fragments of ~500 bp each from each of the flanking regions of the *arcA* coding sequence were PCR-amplified from the ancestral genomic DNA using the two primer pairs ODS906: 5'

-ACGCCATTCTGCTGATTGCA-3' / ODS907: 5' -
 CGATGTCGACTAATCGGCTTTACCACCGTCAA-3', and ODS908: 5' -
 ATCGGTCGACCATGTTTGCTACCTAAATTGCC-3' / ODS909: 5' -
 TTCCTGACTGTACTAACGG-3'. Second, an overlapping PCR product of ~1000 bp was produced using the two primers ODS906 and ODS909. For the construction of 6K3::*arcA*^{Anc}, the *arcA*^{Anc} allele was PCR-amplified from the ancestral genomic DNA as an ~1000-bp DNA fragment with ~500 bp on each side of the position corresponding to the *arcA*^S mutation, using the two primers 5' -GCCAGTAAAGAAGTTACAATGGAC-3' and 5' -CGAAGCGTAGTTTTATTGGGTGT-3'.

The two PCR products allowing the construction of each of the two strains (*AncΔarcA* and 6K3::*arcA*^{Anc}) were then first cloned into pCRII-Topo (Invitrogen, Carlsbad, USA), and then sub-cloned into pKO3. After electro-transformation of the relevant strain (*Anc* or 6K3) with the appropriate suicide plasmid, chloramphenicol-resistant cells with a copy of this non-replicative plasmid integrated into the chromosome were selected at 43°C (31). As previously described (31), subsequent plating on 5% sucrose-LB plates without NaCl allowed the selection of plasmid loss, because pKO3 carries *sacB*, which renders cells sensitive to killing by sucrose. After this selection step, clones were screened by PCR (for the *AncΔarcA* strain) and sequencing (for the 6K3::*arcA*^{Anc} strain) to isolate the desired isogenic constructs.

Media and culture conditions. Bacteria were grown in Davis minimal (DM) medium as in the LTEE (15). The DM medium was supplemented with different carbon sources at various concentrations: glucose alone at 1000 mg/L (DM1000-glucose), 250 mg/L (DM250-glucose) or 25 mg/L (DM25-glucose), glucose (10% v/v) and acetate (90% v/v) at 250 mg/L (DM glucose:acetate), or acetate alone at 250 mg/L (DM250-acetate). These higher concentrations, instead of the DM25-glucose medium used during the LTEE (15), were used to increase cell density and improve the accuracy of measurements of cell growth. After overnight growth in DM1000-glucose, each strain was acclimated in triplicate cultures into DM250-

glucose for 24 h at 37°C and shaking at 120 rpm after a 10,000-fold dilution. Each acclimation culture was then inoculated as duplicate, at a 1:100 dilution, into either a 6.5KL4 filtrate prepared from a 24-h culture of 6.5KL4 grown in DM250-glucose, DM glucose:acetate, or DM250-acetate. The cultures were then incubated in 96-well microtiter plates at 37°C for 24 h. Growth was monitored by scoring the OD_{450nm} every 10 min with an Infinite M200 microplate reader (Tecan®). Growth rates were computed by fitting an exponential growth model using the custom analysis software, Curve Fitter (N. F. Delaney, CJM, unpublished; <http://www.evolvedmicrobe.com/Software.html>).

Promoter activities. We measured the activities of the promoters of the two genes *acnB* and *aceB* that encode two enzymes from the TCA cycle, aconitase B and malate synthase B, respectively, by using the corresponding reporter plasmids (derived from pUA66) from the *E. coli* library of *gfp* transcriptional fusions (32). These two genes are known to be repressed by ArcA. We also measured the promoter activity of *acs* that encode acetyl-CoA synthase. In this particular case, we constructed the corresponding pUA66-derived reporter plasmid as it was not available from the transcriptional fusion library. Briefly, the *acs* intergenic region was PCR-amplified from the ancestral (*acs*^{Anc}) or 6.5KS1 (*acs*^S) genomic DNA using the two primers 5'-TGCTGAGGGTTTATCAGGCA-3' and 5'-AAGCTGAAGATACGGCGTGC-3'. The PCR products were then inserted into the pUA66 plasmid using T4 DNA ligase (New England Biolabs, Ipswich, USA), resulting in the two plasmids pUA66::*acs*^{Anc} and pUA66::*acs*^S. Each plasmid, as well as the empty pUA66 reference plasmid as a standard, was introduced by electro-transformation into Anc, 6.5KS1, 6.5KL4 and derived constructed strains with different *arcA* alleles (Table 1). Plasmid-bearing clones were grown in DM250-glucose supplemented with 25 µg/µl kanamycin. Both OD_{450nm} and GFP fluorescence emission were measured every 10 min for 24 h with an Infinite M200 microplate reader (Tecan®). Promoter activities were estimated as rates of GFP production from each promoter region (32). They were computed using MATLAB as (dGFP/dt)/OD_{450nm}, where GFP is the fluorescence signal after subtracting

the value for the empty plasmid. All figures show the average of three biological and two technical replicates, and error bars show standard deviations.

Purification of ArcA proteins. The ancestral ArcA^{Anc} and evolved ArcA^S proteins were produced in their His-tag recombinant forms after cloning each of the corresponding genes into the expression plasmid pQE-80L (Qiagen, Hilden, Germany). Briefly, overexpression of both genes was performed in an *E. coli* MG1655 Δ slyD strain, since the SlyD protein possesses a natural His-tag and has approximately the same size than ArcA. The use of the MG1655 Δ slyD strain therefore avoided co-purification of SlyD with ArcA. Each of the two recombinant strains was grown in LB broth supplemented with 100 μ g/ml of ampicillin at 30°C. At an OD_{600nm} = 0.6, protein production was induced with 1 mM IPTG and the culture was further incubated for 6 h at 30°C with vigorous shaking (200 rpm). Cells were then harvested and protein overexpression was verified by SDS-PAGE and Western-blot using a His-tag antibody (Thermo-Scientific Pierce, Rockford, USA). Protein purification was performed in native conditions from a clone with the highest and lowest level of protein expression after and before IPTG induction, respectively, by using Ni-NTA agarose columns following protocols 9 and 12 of the QIAexpressionistTM handbook (Qiagen, 2003).

DNA probe preparation. Using the ancestral genomic DNA as template, a cold-labeled DNA probe was prepared by PCR from the promoter region of *fadE*, a gene known to be repressed by ArcA (29). The PCR product was obtained with a DIG-labeled primer (*fadE*5'DIG – CAGGACTTTTTGACCTGAAG-3', the DIG-labeled nucleotide being indicated in bold) and an unlabeled primer (*fadE*5' – GTATTCTCGCTACGGTTGTC -3'). The PCR product was verified by electrophoresis in a 1% agarose gel and purified with the NucleoSpin[®] PCR Clean-up Kit (Macherey-Nagel GmbH & Co., Hoerd, France) following the manufacturer's recommendations.

Electrophoretic mobility shift assay (EMSA). ArcA proteins were phosphorylated in a buffer containing 100 mM Tris-HCl (pH 7.4), 100 mM KCl, 10 mM MgCl₂, 2 mM

dithiothreitol, 10% glycerol, and 50 mM acetyl-phosphate for 45 min at 30°C. The DIG-labeled probe (21 nM) was allowed to interact with 1.2 µM of non-phosphorylated and phosphorylated ArcA proteins for 30 min at room temperature and immediately subjected to electrophoresis in a 4% acrylamide gel at 4°C. The gel was electro-transferred to a Nylon membrane and DIG immuno-detection was performed using Fab fragments from an anti-DIG antibody conjugated with alkaline phosphatase (Roche®), using CSPD (Roche®) as dephosphorylation substrate. To estimate dissociation constants (Kd), we used increasing concentrations of ancestral and evolved ArcA-P proteins (0 – 9 µM) that were incubated with 10 nM of the DIG-labeled *fadE* probe. The relative intensities of the retarded bands were measured with Image J (Maryland, USA) and plotted against protein concentrations. The Kd was defined as the protein concentration required to bind half of the labeled probes.

Acknowledgments

J.C. and J.P. acknowledge the French Ministry of Research and the Université Grenoble Alpes for PhD fellowships.

J.C. performed almost all experiments; J.P. purified the ArcA proteins and constructed isogenic strains; J.G. helped for the EMSA experiments; JC, JG, TH, and DS analyzed the data; J.C., T.H. and D.S. designed the experiments; R.E.L. provided all strains; J.C. and D.S. wrote the manuscript, which was edited by all the authors.

Funding information

This work was funded by Université Grenoble Alpes (to Dominique Schneider), Centre National de la Recherche Scientifique (CNRS, to Dominique Schneider), and a grant from the European Commission 7th Framework Programme (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427, to Dominique Schneider).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

1. **Schluter D.** 2000. The ecology of adaptive radiations. Oxford University Press.
2. **Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ, Priebe GP, Kishony R.** 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* **43**:1275–1280.
3. **Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB.** 2002. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**:33–46.
4. **Treves DS, 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.
5. **Friesen ML, Saxer G, Travisano M, Doebeli M.** 2004. Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution* **58**:245–260.
6. **Spencer CC, Tyerman J, Bertrand M, Doebeli M.** 2008. Adaptation increases the likelihood of diversification in an experimental bacterial lineage. *Proc Natl Acad Sci U S A* **105**:1585–1589.
7. **Rozen DE, Lenski RE.** 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am Nat* **155**:24–35.
8. **Blount ZD, Borland CZ, Lenski RE.** 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A* **105**:7899–7906.
9. **Laland KN, Odling-Smee FJ, Feldman MW.** 1999. Evolutionary consequences of niche construction and their implications for ecology. *Proc Natl Acad Sci U S A* **96**:10242–10247.
10. **Abrams PA, Cortez MH.** 2015. Is competition needed for ecological character displacement? Does displacement decrease competition? *Evolution* **69**:3039–3053.
11. **Blount ZD, Barrick JE, Davidson CJ, Lenski RE.** 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**:513–518.

12. **Herron MD, Doebeli M.** 2013. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol* **11**:e1001490.
13. **Plucain J, Hindré T, Le Gac M, Tenailon O, Cruveiller S, Medigue C, Leiby N, Harcombe WR, Marx CJ, Lenski RE, Schneider D.** 2014. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* **343**:1366–1369.
14. **Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF.** 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3). *J Mol Biol* **394**:644–652.
15. **Lenski RE, Rose MR, Simpson SC, Tadler SC.** 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* **138**:1315–1341.
16. **Rozen DE, Schneider D, Lenski RE.** 2005. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *J Mol Evol* **61**:171–180.
17. **Rozen DE, Philippe N, de Visser JA, Lenski RE, Schneider D.** 2009. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecol Lett* **12**:34–44.
18. **Großkopf T, Consuegra J, Gaffé J, Willison JC, Lenski RE, Soyer OS, Schneider D.** 2016. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC Evol Biol* In press.
19. **Kumari S, Beatty CM, Browning DF, Busby SJW, Simel EJ, Hovel-Miner G, Wolfe AJ.** 2000. Regulation of acetyl coenzyme A synthetase in *Escherichia coli*. *J Bacteriol* **182**:4173–4179.
20. **Brock M, Maerker C, Schutz A, Volker U, Buckel W.** 2002. Oxidation of propionate to pyruvate in *Escherichia coli* - Involvement of methylcitrate dehydratase and aconitase. *Eur J Biochem* **269**:6184–6194.
21. **Molina I, Pellicer MT, Badia J, Aguilar J, Baldoma L.** 1994. Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme. *Eur J Biochem* **224**:541–548.
22. **Le Gac M, Plucain J, Hindré T, Lenski RE, Schneider D.** 2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci U S A* **109**:9487–9492.

23. **Malpica R, Sandoval GR, Rodriguez C, Franco B, Georgellis D.** 2006. Signaling by the Arc two-component system provides a link between the redox state of the quinone pool and gene expression. *Antioxid Redox Signal* **8**:781–795.
24. **Liu X, De Wulf P.** 2004. Probing the ArcA-P modulon of *Escherichia coli* by whole genome transcriptional analysis and sequence recognition profiling. *J Biol Chem* **279**:12588–12597.
25. **Park DM, Akhtar MS, Ansari AZ, Landick R, Kiley PJ.** 2013. The bacterial response regulator ArcA uses a diverse binding site architecture to regulate carbon oxidation globally. *PLoS Genet* **9(10)**:e1003839.
26. **Toro-Roman A, Mack TR, Stock AM.** 2005. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. *J Mol Biol* **349**:11–26.
27. **Kumari S, Tishel R, Eisenbach M, Wolfe AJ.** 1995. Cloning, characterization, and functional expression of *acs*, the gene which encodes acetyl-coenzyme A synthetase in *Escherichia coli*. *J Bacteriol* **177**:2878–2886.
28. **Leiby N, Marx CJ.** 2014. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS Biol* **12**:10.
29. **Cho BK, Knight EM, Palsson B.** 2006. Transcriptional regulation of the *fad* regulon genes of *Escherichia coli* by ArcA. *Microbiology* **152**:2207–2219.
30. **Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C, Schneider D, Lenski RE.** 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**:165–170.
31. **Link AJ, Phillips D, Church GM.** 1997. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: Application to open reading frame characterization. *J Bacteriol* **179**:6228–6237.
32. **Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U.** 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods* **3**:623–628.
33. **Browning DF, Beatty CM, Sanstad EA, Gunn KE, Busby SJW, Wolfe AJ.** 2004. Modulation of CRP-dependent transcription at the *Escherichia coli acsP2* promoter by nucleoprotein complexes: anti-activation by the nucleoid proteins FIS and IHF. *Mol Microbiol* **51**:241–254.

34. **Rosenzweig RF, Sharp RR, Treves DS, Adams J.** 1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**:903–917.
35. **Jeon Y, Lee YS, Han JS, Kim JB, Hwang DS.** 2001. Multimerization of phosphorylated and non-phosphorylated ArcA is necessary for the response regulator function of the Arc two-component signal transduction system. *J Biol Chem* **276**:40873–40879.
36. **Sniegowski PD, Gerrish PJ, Lenski RE.** 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**:703–705.

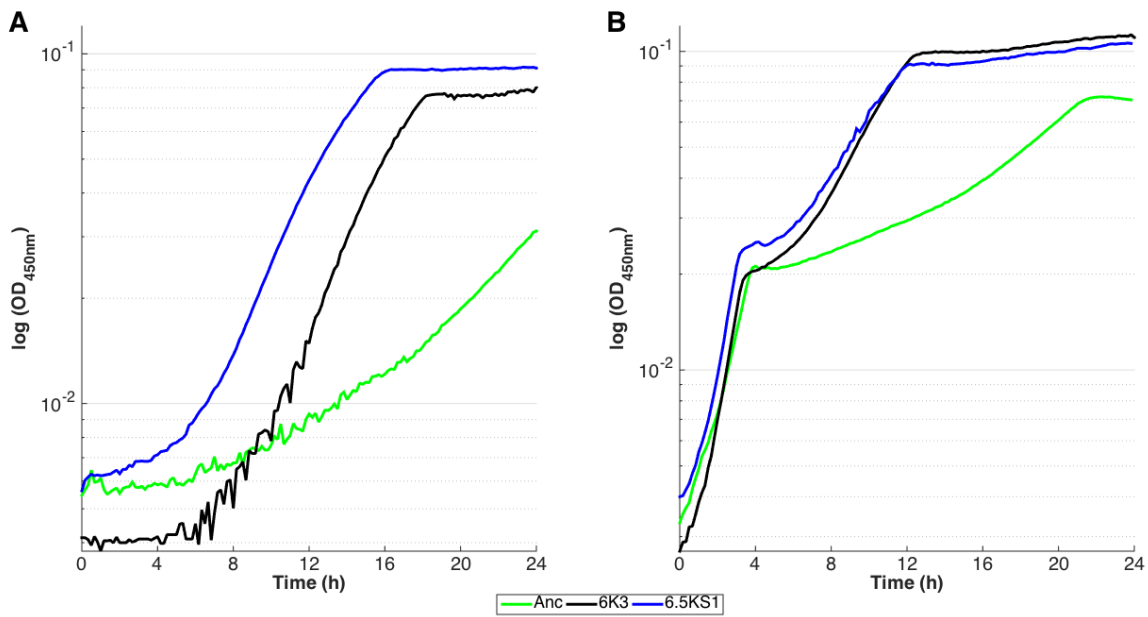
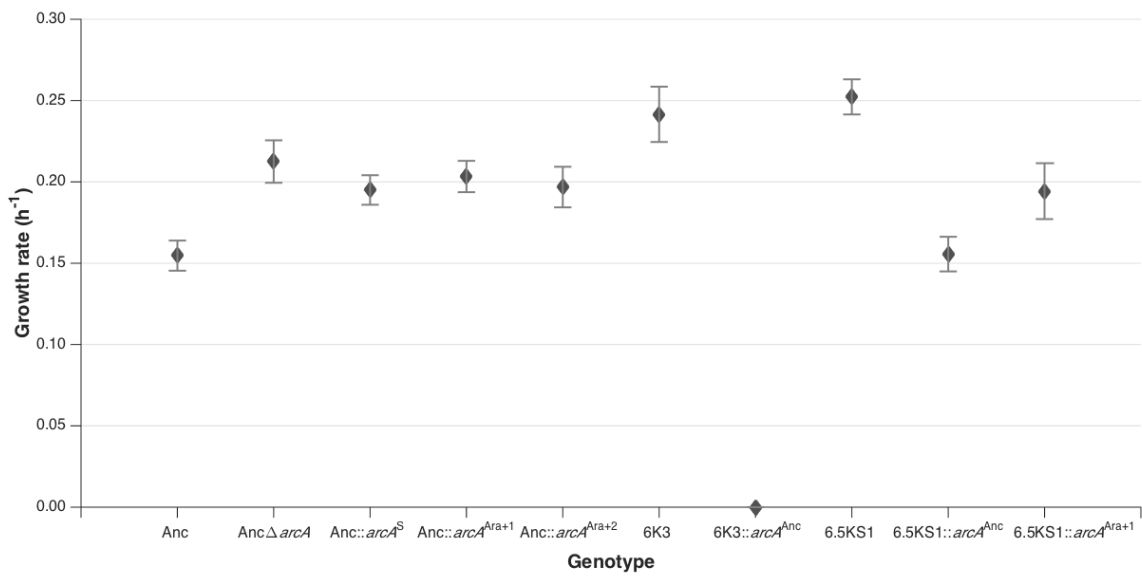


FIG 1 Growth profiles of Anc (green), 6K3 (black) and 6.5KS1 (blue) clones in DM250-acetate (A) and DM-glucose:acetate (B). The curves shown are means of three biological replicates, each performed in duplicates.



FIG

2

Growth rates (h⁻¹) of Anc, 6K3, 6.5KS1 and constructed isogenic *arcA* mutants in DM250-acetate medium. Data are means of three biological replicates, each performed in duplicates, and error bars show standard deviations.

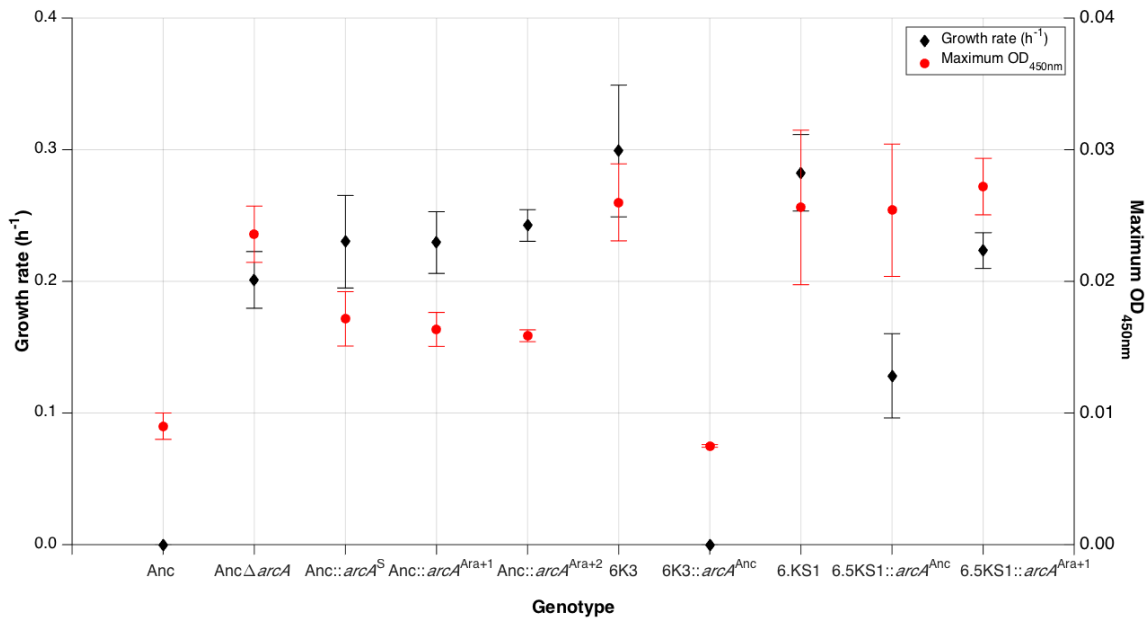


FIG 3

Growth rates (h^{-1} , black diamonds) and maximum OD_{450nm} (red dots) of Anc, 6K3, 6.5KS1 and constructed isogenic *arcA* mutants in a 24-h filtrate from a culture of 6.5KL4 (Table 1) grown in DM250-glucose. Data are means of three biological replicates, each performed in duplicates, and error bars show standard deviations.

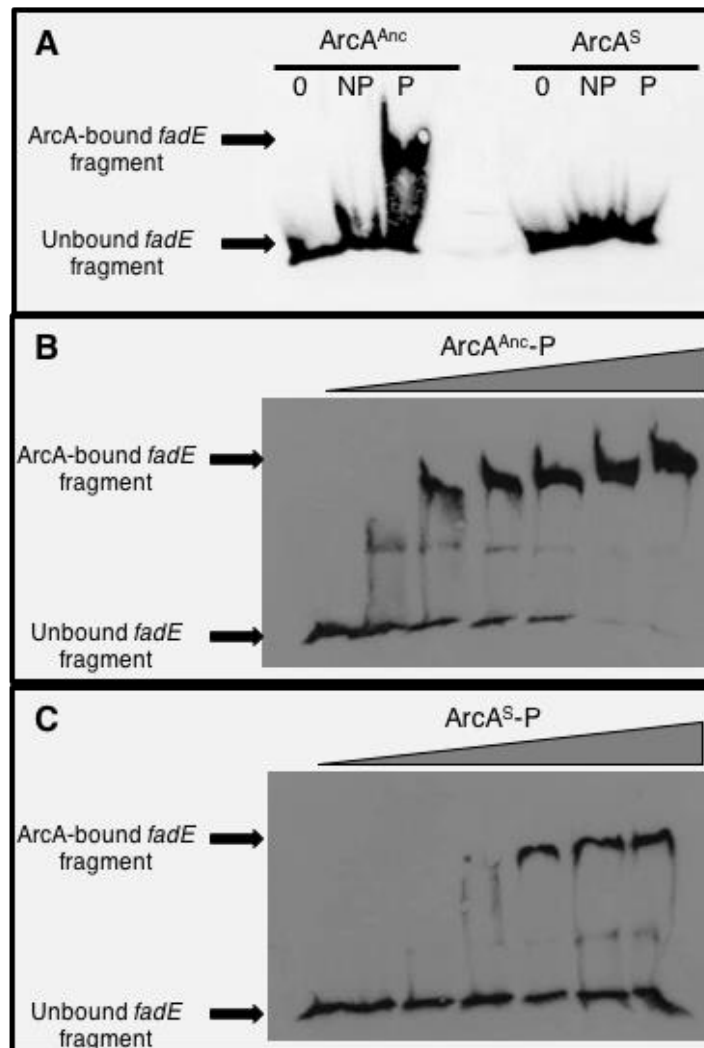


FIG 4 DNA-binding ability of ArcA proteins to the promoter region of *fadE*. (A) Electrophoretic mobility shift assay (EMSA) was performed using both ArcA^{Anc} (left) and ArcA^S (right) proteins. Lines 0, no protein; lines NP, non-phosphorylated ArcA proteins; lines P, phosphorylated ArcA proteins. (B and C) To estimate dissociation constants (K_d), the *fadE* DNA probe was incubated with increasing concentrations of ArcA^{Anc}-P (B) and ArcA^S-P (C). Both proteins were able to bind to the target promoter (retarded band), although with different binding affinities ($K_d = 0.7$ and $1.2 \mu\text{M}$, respectively). The unbound and ArcA-bound DNA fragments are indicated with arrows.

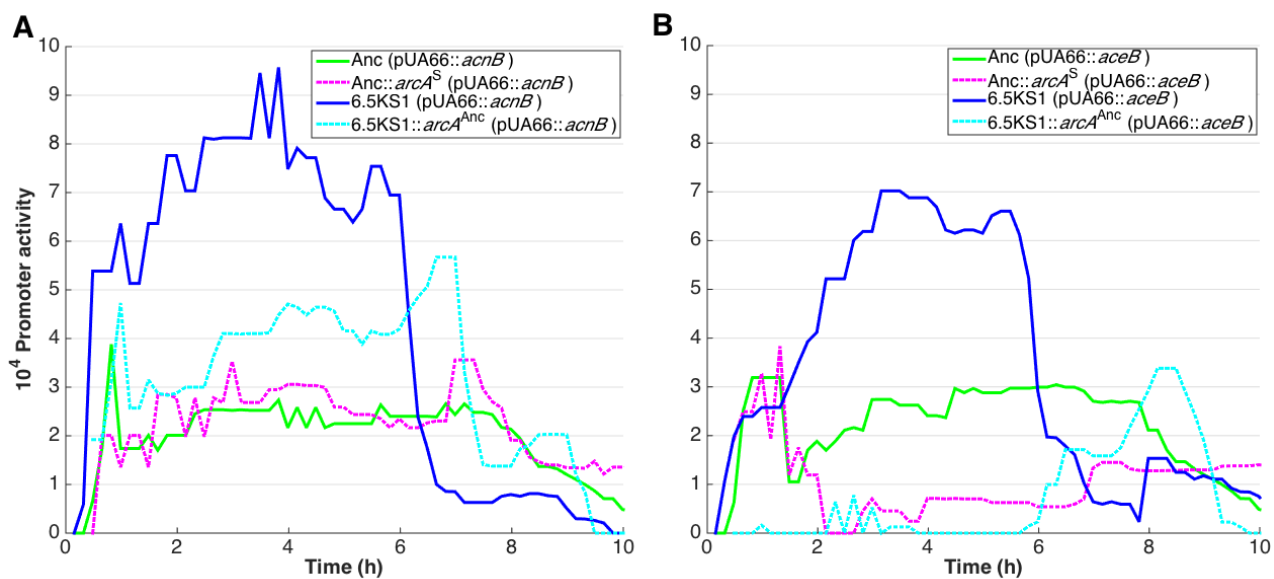


FIG 5 Transcription level of *acnB*, encoding aconitase B (A) and *aceB*, encoding malate synthase A (B) as a function of *arcA* alleles and genetic backgrounds. The promoter activities of the two genes were measured by transcriptional fusions as (dGFP/dt)/OD_{450nm} during the first 10 h of growth in DM250-glucose medium using as a reporter the pUA66-GFP plasmid introduced in Anc (green), Anc::*arcA*^S (dotted, magenta), 6.5KS1 (blue), and 6.5KS1::*arcA*^{Anc} (dotted, cyan) strains (Table 1). The curves shown are means of three biological replicates, each performed in duplicates.

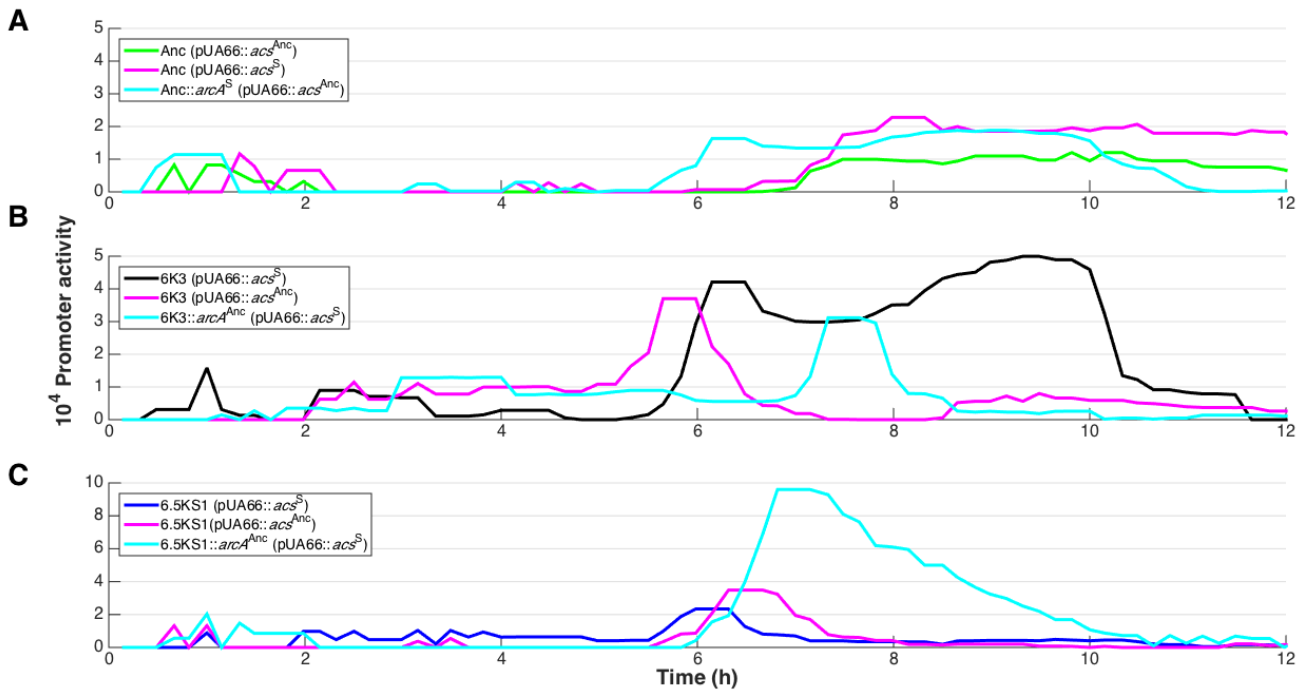


FIG 6 Transcription level of *acs* promoter alleles. The promoter activities of *acs*^{Anc} or *acs*^S were measured by transcriptional fusions as (dGFP/dt)/OD_{450nm} during the first 12 h of growth in DM250-glucose medium using as a reporter the pUA66-GFP plasmid introduced in Anc with the *arcA*^{Anc} and *arcA*^S alleles (A), in 6K3 and 6K3::*arcA*^{Anc} (B), and in 6.5KS1 and 6.5KS1::*arcA*^{Anc} (C). The curves shown are means of three biological replicates, each performed in duplicates.

TABLE 1 Strains used in this study

Strain	Genotype ^a				Reference
	<i>spoT</i>	<i>arcA</i>	<i>gntR</i>	<i>acs</i>	
Ancestor REL606 (Anc)	Anc	Anc	Anc	Anc	(14)
Anc Δ <i>arcA</i>	Anc	Del	Anc	Anc	This study
Anc:: <i>arcA</i> ^S	Anc	Evol (S)	Anc	Anc	(13)
Anc:: <i>arcA</i> ^{Ara+1}	Anc	Evol (Ara+1)	Anc	Anc	(13)
Anc:: <i>arcA</i> ^{Ara+2}	Anc	Evol (Ara+2)	Anc	Anc	(13)
6K3	Evol	Evol	Anc	Evol	(13)
6K3:: <i>arcA</i> ^{Anc}	Evol	Anc	Anc	Evol	This study
6.5KS1	Evol	Evol	Evol	Evol	(13)
6.KS1:: <i>arcA</i> ^{Anc}	Evol	Anc	Evol	Evol	(13)
6.5KS1:: <i>arcA</i> ^{Ara+1}	Evol	Evol (Ara+1)	Evol	Evol	(13)
6.5KL4	Evol	Anc	Anc	Anc	(13)

^a The allelic status is given for 4 genes (*spoT*, *arcA*, *gntR* and *acs*). Anc, ancestral allele; Del, in-frame deletion; Evol, evolved allele with the name of the lineage (S) or population (Ara+1, Ara+2) in which it was substituted.

Chapter 4: Long-Term Dynamics of Mutation Rates in the Ara-2 Population

This section of my PhD is part of a bigger project aiming at the complete characterization of the dynamics of mutation rates during the Long-Term Evolution Experiment with *E. coli*. During the LTEE, 6 of the 12 populations evolved a mutator phenotype owing to mutations in DNA repair genes. The project combines experimental characterization of mutation rates and a theoretical model of the rise and fall of mutation rates developed by Olivier Tenaillon (INSERM, Université Paris Diderot, Hôpital Bichat, Paris). The experimental characterization started in 2013 with the publication in *PNAS* of the mutation rate dynamics of the Ara-1 population by Sébastien Wielgoss, a former post-doctoral researcher in our team (Wielgoss et al. 2013). It was followed by the doctoral work of Larissa Viraphong in 2015 who studied the dynamics of mutation rates in the hypermutator Ara+6 population (Viraphong 2015). Both Ara-1 and Ara+6 populations evolved a hypermutator phenotype after 26,000 and 4000 generations, respectively, owing to mutations in *mutT*, a gene involved in the detoxification of the oxidized nucleotide 8-oxo-G (Maki and Sekiguchi 1992). After this first rise of mutation rates, compensatory mutations (in *mutY* and other genes) were subsequently fixed in each population, such that the mutation rate decreased again to reduce the genetic load due to deleterious mutations (Wielgoss et al. 2013, Viraphong 2015). In each case however, the mutation rates remained higher than the ancestral one despite the compensatory mutations.

To complete the experimental analyses in the LTEE populations, I studied the long-term dynamics of mutation rates in the Ara-2 population. Three main differences made this analysis important: first, the Ara-2 population evolved a hypermutator phenotype much earlier at 2000 generations (therefore before the split between the S and L lineages); second the hypermutator phenotype is due to a mutation in *mutL*, one of the genes involved in methyl-mismatch repair (Jayaraman 2009) and not in *mutT*; third, Ara-2 has the particularity to harbor a long-term polymorphism with the two lineages S and L co-existing for more than 50,000 generations. During my PhD, we performed genomic analyses and experimental measurements of mutation

rates. The results are presented in this chapter and are included in a manuscript that is in preparation and that combines the dynamics of mutation rates in Ara+6, Ara-2, as well as in the three other mutator populations of the LTEE (Ara-3, Ara-4, and Ara+3). Analyses in the last three populations involve metagenomics analyses over evolutionary time that were performed in the laboratory of Michael Desai (Harvard University, USA).

4.1. Introduction

Mutations of DNA sequences are the fundamental source of long-term evolutionary adaptation. (In microbes, horizontal transfer is also a substrate for evolution, but we will focus here on mutations). The extraordinary high level of adaptability of microorganisms allows them to colonize most, if not all, ecological niches on Earth and to resist to the strong selective pressures imposed by human activities, *e.g.* global warming, pollution, antibiotics,... This ability is related to the production of novel adaptive mutants, a capacity called evolvability, that depends on genetic features either by changes on mutation rates, mutational robustness or epistasis (Arenas and Cooper 2013).

Mutations are essential to produce genetic diversity. They arise from DNA damages or replication errors that escape from internal DNA repair systems of organisms (Jayaraman 2009). The evolutionary potential of mutations is related to their effects on organismal fitness in the environment where the organism evolves. Mutations can thus be beneficial, deleterious or neutral if they increase, decrease or have no measurable effect on fitness in a specific environment, respectively. The differential impact of mutations on fitness imposes therefore an evolutionary trade-off between evolvability (access to new genotypes with the potential of phenotypic innovation) and genome stability (capacity to resist to the potential deleterious effects of mutations).

In order to preserve their genome integrity, organisms have evolved mechanisms that increase replication fidelity, such as base selection, *in situ* proof reading of errors (editing) and post-replicative mismatch repair (MMR) (Jayaraman 2009). Alterations of these mechanisms through mutations in genes involved in DNA repair may entail modifications in mutation rates defined as the number of mutations per genome per generation. The implementation of replication fidelity mechanisms, and thus high fidelity, limits the number of mutations in an

organismal genome, preserving it from deleterious mutations but at the same time reducing the ability of generating genetic diversity in a population. Conversely, a high mutation rate although increasing the frequency of beneficial mutations also has a cost by increasing the genetic load. The consequence of this relationship between fidelity and mutation availability may be considered as a trade-off between short-term fitness with few mutations and long-term evolvability (Ferenci 2016).

However, selection can favor the emergence of mutator alleles in genes involved in the mechanisms of replication fidelity, resulting in hypermutator strains. The hypermutator state is characterized by an increase of 10- to 100-fold of the mutation rate. The emergence of a mutator phenotype in bacteria is frequently reported in both environmental and clinical settings (Trobner and Piechocki 1984, LeClerc et al. 1996, Matic et al. 1997, Oliver et al. 2000, Denamur et al. 2002), as well as during laboratory evolution experiments (Sniegowski, Gerrish, and Lenski 1997, Giraud et al. 2001, Wielgoss et al. 2013).

Because asexual populations cannot generate variation by recombination, mutations arise and fix sequentially owing to selective sweeps. Selection of a mutator phenotype therefore relies on the association between the mutator allele and mutations at other loci that affect fitness (Tenaillon et al. 2000, Chao and Cox 1983). This process, named “genetic hitchhiking”, implies that the rise of a mutator allele results mainly from its linkage with beneficial mutations. Thereafter, the mutator allele acts as an evolvability allele, thereby increasing the capacity of an organism to adapt without conferring any direct fitness benefit (Arenas and Cooper 2013, Tenaillon et al. 1999, Shaver et al. 2002). However, even if a high mutation rate confers evolvability in an asexual population in the short term, over the long term indirect selection to reduce the genetic load linked to deleterious mutations will eventually favor a decrease in mutation rates (Sniegowski et al. 2000).

The role of mutator alleles in adaptive evolution has been studied using computational models. Taddei et al., proposed that a transient increase in mutation rates in natural asexual populations is frequent because mutator genotypes generate adaptive mutations more efficiently (Taddei et al. 1997). In their simulations, they showed that on average the fitness of populations with a high frequency of mutators increased faster during the process of adaptation and that once the adaptive peak has been reached, the mutator frequency will decline. Two selective pressures may favor lower genomic mutation rates: the concomitant influx of deleterious mutations and the energetic cost of fidelity (Sniegowski et al. 2000). This tension between adaptation and genetic load leads to an interesting debate about the evolutionary dynamics of mutation rates, especially as a function of the environmental conditions.

Evidence for mutation rate fluctuation exists in nature. Indeed, signatures of past inactivation and re-functionalization through horizontal gene transfer of mismatch repair genes have been detected in *E. coli* (Denamur et al. 2000). Laboratory evolution experiments are usually performed with asexual populations in which gene transfer cannot occur; even though, decay of the mutator phenotype has been frequently observed, in most cases by compensatory mutations (Trobner and Piechocki 1984, McDonald et al. 2012, Wielgoss et al. 2013).

The long-term dynamics of mutations rates is an important issue in evolutionary biology, being frequently studied in *in vitro* evolved populations. Indeed, during the Long-Term Evolution Experiment with *E.coli* (LTEE), 6 out of the 12 populations evolved mutator phenotypes (Sniegowski, Gerrish, and Lenski 1997, Wielgoss et al. 2013, Viraphong 2015, Cooper and Lenski 2000). In this experiment, twelve populations are independently propagated since about 65,000 generations from the same common ancestor of *Escherichia coli* in a glucose-limited minimal medium by daily 1:100 dilution transfers (Lenski et al. 1991). The time of emergence of the mutator phenotype varies among the mutator populations, with four of them evolving it early during the experiment, before 10,000 generations, and two much later

after 20,000 generations. The long-term fate of the mutator phenotype has been studied in two of the six mutator populations, Ara-1 and Ara+6 (Viraphong 2015, Wielgoss et al. 2013). The Ara-1 population is characterized by both the late emergence of the mutator state and two independent compensation events sharing a similar mechanism (Wielgoss et al. 2013). Indeed, this population evolved a mutator state after ~26,000 generations owing to a mutation in *mutT* encoding an enzyme that purges the pool of intracellular oxidized guanine nucleotides (8-oxo-dGTP). The increase in mutation rate was ~100-fold. After the emergence of this mutator state, at least two independent lineages with an antimutator phenotype swept to fixation in less than 10,000 generations. Each of these two antimutator lineages had a two-fold reduced mutation rate owing to mutations in *mutY* encoding a DNA repair glycosylase. Genomic analyses showed that these antimutators were selected for their reduced genetic load at an evolutionary time where the bacterial population was already well adapted to its environment (Wielgoss et al. 2013).

Population Ara+6 evolved a mutator state much earlier during evolution, before 4000 generations, after a mutation in *mutT* which was characterized by a 100-fold increase in the mutation rate compared to the ancestor. This high mutation rate decreased successively owing, (i) first, to a compensatory mutation in *dnaE* that occurred at 5000 generations, (ii) and second, to two compensation events involving *mutY* mutations in lineages at 20,000 and 40,000 generations. These steps presumably reduced the genetic load imposed by the high mutation rate that became too strong when the population was highly adapted to its environment. This dynamics suggests that mutation rate, and consequently evolvability, can be fine-tuned as a continuous trait, according to the needs for adaptation (Viraphong 2015).

A third population of the LTEE, the population Ara-2, experienced an early increase in mutation rates owing to a mutation in *mutL* (Sniegowski, Gerrish, and Lenski 1997). This population is polymorphic. After 6500 generations, two lineages, called L (Large) and S (Small)

according to their cell sizes, emerged from the same ancestor at ~6500 generations and dynamically coexisted ever since owing to negative frequency-dependent interactions whereby each lineage is favored and invades the other when rare, precluding the extinction of any of them (Le Gac et al. 2012, Rozen and Lenski 2000, Rozen, Schneider, and Lenski 2005). We recently showed that the ecological mechanisms involved in this polymorphism implied niche construction by the L lineage through the secretion of acetate and the subsequent character displacement and trade-offs with the L and S lineages becoming more and more adapted to glucose and acetate, respectively. These ecological mechanisms prevented simultaneous evolution of either lineage on both carbon sources (Großkopf et al. 2016).

The mutator state of the Ara-2 population was initially reported in a survey of the 12 LTEE populations at 10,000 generations (Sniegowski, Gerrish, and Lenski 1997). The mutator phenotype was due to a 6-bp insertion in a repeated region of the *mutL* gene encoding MutL a member of the MTHLS complex that participates in the methyl-directed mismatch repair (MMR) pathway in *E. coli* (Jayaraman 2009). Indeed, the mutator phenotype was reverted by complementation with a plasmid bearing the ancestral allele (Shaver and Sniegowski 2003). The mutator allele increased in frequency at about 2000 generations and was substituted after 3500 generations (Shaver et al. 2002, Shaver and Sniegowski 2003). Therefore, the mutator phenotype emerged 3000 generations before the diversification event and is shared by both S and L lineages.

Here, we studied the long-term dynamics of the mutation rates in the polymorphic Ara-2 population by analysing genome sequences of S and L clones sampled during the co-existence of the two S and L lineages. Moreover, we performed fluctuation tests to experimentally quantify mutation rates in both lineages and determine their dynamics during the 50,000 generations of co-existence. Finally, the genome sequences allowed us to characterize the genetics of the dynamics of the mutator state.

4.2. Materials and Methods

4.2.1. Long-Term Evolution Experiment (LTEE), strains and media

The LTEE consists in 12 populations founded from the same ancestral strain of *E. coli*, REL606 (Jeong *et al*, 2009). They are propagated since 1988 by daily 1:100 dilutions in Davis minimal medium (Lenski *et al*, 1991) supplemented with glucose at 25 mg/L (DM25). Here, we focused on one population, called Ara-2, in which two lineages, S and L, diverged before 6500 generations from a common ancestor and co-exist ever since (Rozen & Lenski, 2000; Plucain *et al*, 2014). All strains used in this study are derived from population Ara-2. We used mixed population samples from 2000, 5000, 6000, 6500, 11,000, 18,000, 20,000, 30,000, 40,000 and 50,000 generations. One evolved clone was sampled at each of the timepoints 2000 and 5000 generations before the diversification event. They were named by their sampling time followed by an arbitrary numeral for a given clone. For example, 2K4 is a pre-diversification evolved clone that was sampled at 2000 generations. After the diversification event, at 6500 generations, one evolved clone from each of the two S and L lineages was sampled at each of the following timepoints: 6500, 11,000, 18,000, 20,000, 30,000, 40,000 and 50,000 generations. Each evolved clone was named by its generation followed by S or L according to its lineage and an arbitrary numeral for a given clone. For example, 6.5KS1 is a clone from the S lineage that was sampled at 6500 generations.

Strains were grown in rich LB medium (Bertani 1951) or minimal DM medium supplemented with different concentrations of glucose (25 mg/L for DM25, 250mg/L for DM250, 1g/L for DM1000, 2g/L for DM2000, and 4g/L for DM4000). Strains were plated on solid DM4000 medium supplemented with 100 g/L agar. When necessary, rifampicin (DM-Rif) was added at 100 mg/L.

4.2.2. Identification of S clones

During the drop test (see section 4.2.4), clones belonging to the S lineage of the Ara-2 population were identified by taking advantage of the evolved *arcA* allele (C>T at position 4628191 of the REL606 genome) that is known to be diagnostic of this lineage (Plucain et al. 2014). The *arcA* region bearing the evolved allele was PCR-amplified using the two primers 5'-ACGCCATTCTGCTGATTGCA-3' and 5'-TTTCCTGACTGTACTAACGG-3'. The amplified PCR-product was then sequenced using Sanger technology (GATC Biotech AG, Konstanz, Germany).

4.2.3. Genome analyses

All the genome sequences used in this study have been already published and are publicly available (Tenaillon et al. 2016). The dynamics of the number of mutations was determined using the genomes of one clone from each pre-diversification timepoint (500, 1000, 1500, 2000 and 5000 generations) and one clone from each S and L lineage after diversification (6500, 15,000, 20,000, 30,000, 40,000, and 50,000 generations). It was computed as the number of single nucleotide polymorphisms (SNP) by generation; they were then sorted as synonymous or non-synonymous mutations.

4.2.4. Qualitative determination of the mutator phenotype: Drop test

Drop tests were performed at least three times independently for each mixed population and isolated clone. Briefly, after overnight growth at 37°C in DM1000 with shaking, strains were inoculated in triplicate by a 10,000-fold dilution into DM1000, and grown for 24 h at 37°C with shaking at 160 rpm as an acclimation step. Cultures were then concentrated 10-fold and a 10- μ L drop was deposited onto selective DM-Rif plates and also, as a growth control, onto DM4000 plates. Plates were then incubated for at least 72h at 37°C. The presence of significant

growth, compared to the ancestor, on DM-Rif represented a positive mutator phenotype for a given strain/population.

4.2.5. Quantitative determination of mutation rates: Fluctuation test

Fluctuation tests (Luria and Delbrück 1943) were performed at least three times independently for each isolated clone as described (<http://barricklab.org/twiki/bin/view/Lab/ProtocolsFluctuationTests>). Briefly, after overnight growth at 37°C in LB with shaking, strains were inoculated by a 10,000-fold dilution into DM25, in which they grew for 24 h at 37°C with shaking at 120 rpm as an acclimation step. For each strain, 30 replicate cultures were grown each in 200 µL of DM2000 during 24 h at 37°C with shaking at 160 rpm. Twenty-four replicates were then plated onto selective DM-Rif plates while the six others were diluted 10⁶-fold and plated onto DM4000 plates to compute the total cell number. Mutation rates were then estimated using the Fluctuation Analysis Calculator (FALCOR, (Hall et al. 2009) with the Ma-Sandri-Sarkar Likelihood Estimation method (Ma, Sandri, and Sarkar 1992).

4.3. Results

4.3.1. Increase of mutation accumulation corresponds to the phenotypic increase of mutation rates

We analyzed the genomes from 18 clones isolated from population Ara-2 at different timepoints during evolution (Tenaillon et al. 2016). We investigated only single nucleotide polymorphisms (SNP), since it is the type of mutation expected from the *mutL* defect responsible for the mutator phenotype in this population. As expected (Shaver et al. 2002, Shaver and Sniegowski 2003), a strong increase in the number of mutations was observed between 2000 and 5000 generations, *i.e.* before the diversification event (Fig. 25). Three SNP

mutations occurred in the 2000-generation clone 2K4 while 98 additional SNPs were identified in the 5000-generation clone 5K (Fig. 25A). The 3 SNPs in 2K4 were also present in 5K.

The number of mutations increased almost linearly for each of the two S and L lineages, with twice as much non-synonymous mutations compared to synonymous mutations and without significant difference between S and L clones (Fig. 25B and C). Before the mutator state, all SNP mutations were non-synonymous and synonymous mutations increased after the rise in mutation rates. This trend was previously shown for population Ara-1 (Wielgoss et al. 2013). Surprisingly, a transient plateau was detected in the number of: i) non-synonymous mutations in the S clones between 15,000 and 30,000 generations, ii) synonymous mutations in the L clones between 15,000 and 30,000 generations, and iii) total mutations in both S and L at 40,000 generations, with only three synonymous mutations appearing between 40,000 and 50,000 generations in the S clones and none in the L clones.

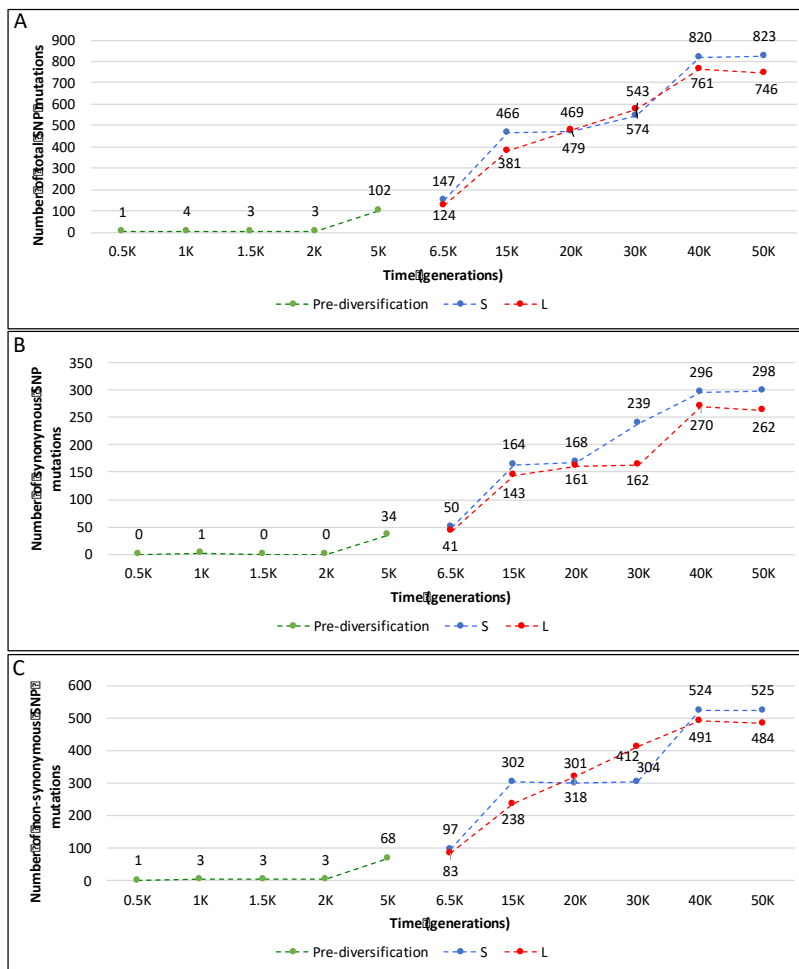


Figure 25. Number of mutations in the Ara-2 population over evolutionary time.

(A) Total number of SNPs. (B) Number of synonymous SNP mutations. (C) Number of non-synonymous SNP mutations. The genomes from pre-diversification, S and L clones are shown in green, blue and red, respectively. The number of relevant SNP mutations is given above each point for the pre-diversification and S clones, and under the point for the L clones.

4.3.2. Reversion of the mutator phenotype

We qualitatively tested the mutator phenotype of the mixed populations sampled at different timepoints during evolution by plating 10 μ L of a concentrated culture of each mixed population on DM-Rif. The presence of significant growth, compared to the ancestor, was diagnostic of a hypermutator phenotype for a given population (Fig. 26).

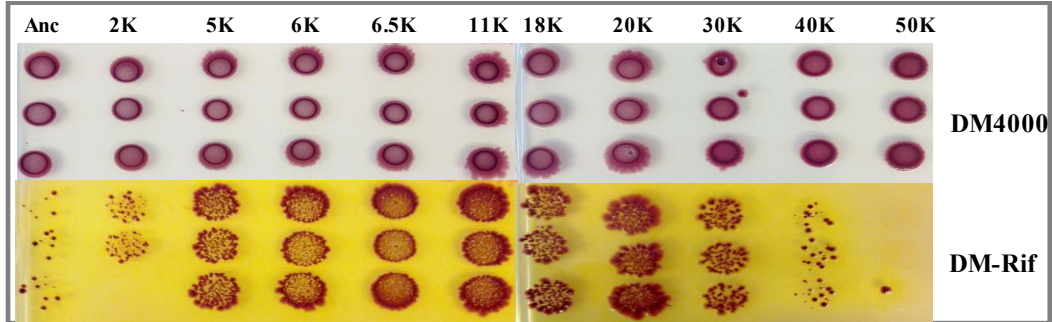


Figure 26. Drop test of mixed Ara-2 populations sampled over evolutionary time.

Cultures of mixed populations of the Anc, S and L lineages from different timepoints (2K to 50K) were plated on DM-Rif and DM4000 as a growth control. The presence of significant growth, compared to the ancestor, indicates a mutator phenotype.

It was previously shown that the hypermutator phenotype emerged at ~2000 generations and was fixed at ~3500 generations (Shaver et al. 2002, Shaver and Sniegowski 2003). We confirmed this dynamics with the increasing growth on DM-Rif between 2000 and 5000 generations (Fig. 26). The mutator phenotype was conserved up to 30,000 generations and subsequently decreased significantly at 40,000 generations, indicative of a reduced mutation rate either in the entire population (therefore in both S and L lineages) or in only one of the two lineages. We therefore grew the mixed population at 40,000 generations in DM1000 liquid culture and, after plating on DM-Rif, sampled 10 isolated clones from which we PCR-amplified and sequenced the *arcA* gene. Indeed, a mutation was detected in *arcA* that is diagnostic of the S lineage (Plucain et al. 2014). All 10 clones harbored the ancestral *arcA* allele, suggesting that they were L clones, implying that the reduction of the mutator rate in the entire population at 40,000 generations most probably occurred in the S lineage while the L lineage was still mutator. To confirm this hypothesis, we performed drop tests with evolved clones from each

of the two S and L lineages that were sampled at different timepoints during evolution (Fig. 27).

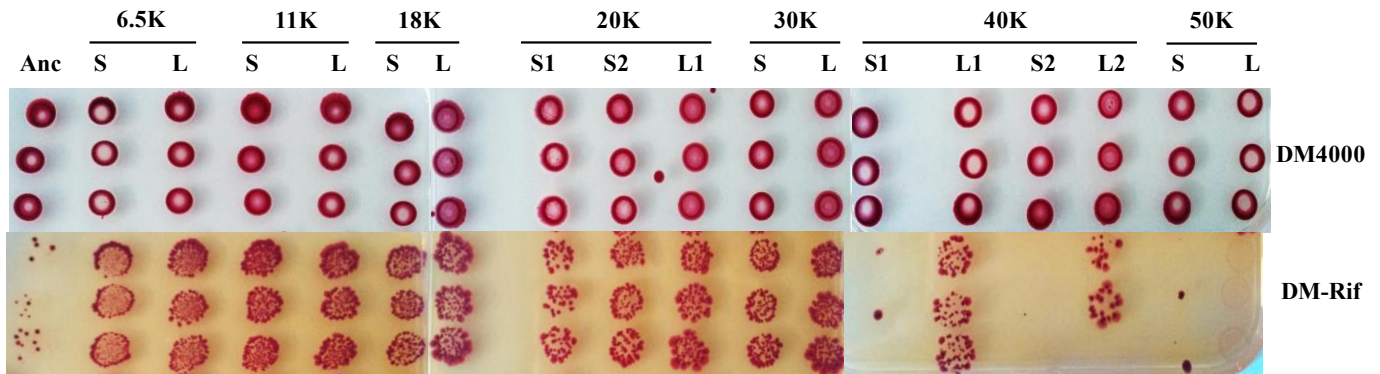


Figure 27. Drop test of isolated Ara-2 S and L clones sampled over evolutionary time.

Cultures of the Anc, S and L isolated clones from different timepoints (6.5K to 50K) were plated on DM-Rif and DM4000 as a growth control. The presence of significant growth, compared to the ancestor, indicates a mutator phenotype.

As for the mixed populations, the mutator phenotype was maintained in both S and L lineages up to 30,000 generations and was then reduced successively, first in the S lineage between 30,000 and 40,000 generations and second in the L lineage between 40,000 and 50,000 generations. These successive changes in mutation rates did not alter the co-existence between the two S and L lineages.

4.3.3. Dynamics of experimental mutation rates over evolutionary time

We experimentally measured mutation rates by fluctuation tests (Luria and Delbrück 1943), using individual evolved clones sampled from the S and L lineages over evolutionary time (Fig. 28).

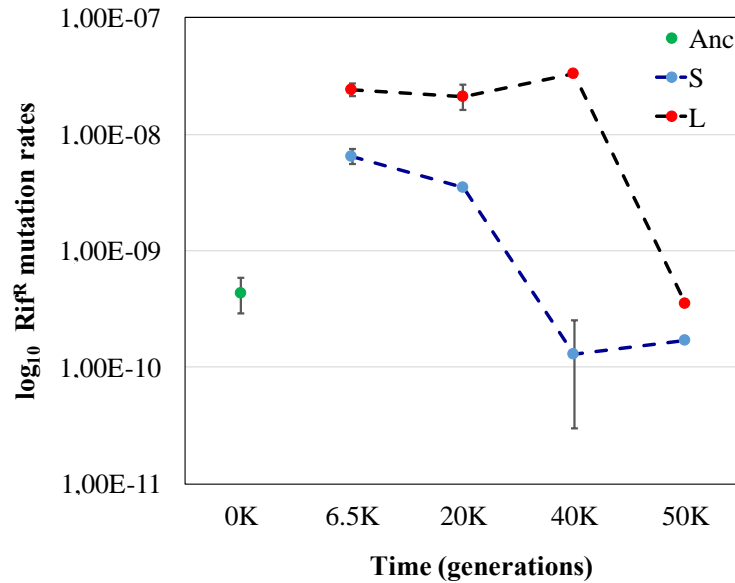


Figure 28. *In vivo* quantification by fluctuation tests of mutation rates in isolated clones from the S and L lineages of population Ara-2.

Mutation rates were determined by fluctuation tests on DM-Rif for the ancestor and isolated clones from the S and L lineages. Error bars show 95% confidence intervals.

These results are consistent with the qualitative determination of the mutator phenotype by drop tests on DM-Rif (Figs 26 and 27). At the time of their phenotypic differentiation, both S and L lineages revealed a 10-fold increased mutation rate compared to the ancestor owing to the *mutL* mutator allele (Sniegowski, Gerrish, and Lenski 1997) that emerged at ~2000 generations (see above). These high mutation rates were altered in the S lineage at 40,000 generations and in the L lineage at 50,000 generations. It is very interesting to emphasize that the mutation rates not only decreased owing to compensatory mechanisms, as we already noticed in at least two other mutator populations from the LTEE (Wielgoss et al. 2013, Viraphong 2015), but went back to the ancestral mutation rate. This suggests that the phenotypic reversion of mutation rates in Ara-2 may be due to a genetic reversion.

4.3.4. Genetic reversion of the *mutL* mutator allele

The mutation involved in the increased mutation rates in population Ara-2 is a 6-bp insertion in the *mutL* gene. The MutL protein is a member of the MutHLS complex that participates in the methyl-directed mismatch repair (MMR) pathway in *E. coli* (Jayaraman 2009). The 6-bp insertion is located in a highly repeated region essential for the ATPase activity of MutL, and has been shown to explain the Ara-2 mutator phenotype (Shaver and Sniegowski 2003).

We analyzed the genomes sequences of S and L clones sampled at 6500 (3 L clones, 2 S clones), 11,000 (one clone each), 18,000 (one clone each), 20,000 (1 L clone, 2 S clones), 30,000 (one clone each), 40,000 (one clone each), and 50,000 (one clone each) generations (Table 1). We detected the genetic reversion of the mutator *mutL* allele, first in the S lineage at 40,000 generations and second in the L lineage at 50,000 generations, consistent with our phenotypic analyses.

Table 1. Dynamics of the mutator *mutL* allele over evolutionary time in the Ara-2 population.

Position	Type	Change	Gene	Effect	Codon	Lineage	Generation
4375768	Insertion	6-bp	<i>mutL</i>	In Frame	68	Anc	5000
							6000
							6500
							11,000
							18,000
4375768	Insertion	6-bp	<i>mutL</i>	In Frame	68	L	20,000
							30,000
							40,000
							6000
							6500
4375768	Insertion	6-bp	<i>mutL</i>	In Frame	68	S	11,000
							18,000
							20,000
							30,000

4.4. Discussion

The Ara-2 population of the LTEE evolved a mutator phenotype very early during evolution, before its major hallmark, the sympatric diversification event leading to the long-term co-existence of the S and L lineages. This population provides therefore a unique opportunity to study the long-term dynamics of mutation rates in a divergent population both at the phenotypic and genomic levels. We observed a multi-step dynamics of mutation rates over 50,000 generations of evolution. First, a hypermutator phenotype emerged at ~2000 generations that was fixed at 3500 generations (Shaver et al. 2002). It was related to a 6-bp insertion in the *mutL* gene, which resulted in a 10-fold increase in the mutation rate compared to the ancestor. After the rise of the hypermutator genotype, the population experienced the emergence of a stable and dynamic polymorphism at ~6500 generations. The two nascent lineages shared the *mutL* hypermutator allele and its associated 10-fold increase in the mutation rate. Second, the high mutation rates were reverted later during evolution, at both the phenotypic and genetic levels. Indeed, the S lineage reverted first between 30,000 and 40,000 generations, followed by the L lineage between 40,000 and 50,000 generations. The evolved S and L clones therefore reverted back to the ancestral mutation rate. To the best of our knowledge, a complete reversion of the hypermutator genotype was not previously reported during evolution experiments and even in natural conditions. Indeed, previous studies on the dynamics of mutations rates in *in vitro*-evolving populations have reported mutation rate compensation events that reduced the mutation rates at a level that was intermediate between the hypermutator and ancestral states (McDonald et al. 2012, Wielgoss et al. 2013). For example, in the Ara-1 population, mutation rate compensation occurred within a few thousands of generations after the fixation of a mutator phenotype owing to further changes in mismatch repair genes (Wielgoss et al. 2013), while two successive steps of mutation rate compensation occurred in Ara+6 (Viraphong 2015).

One may ask the role of the hypermutator trait in the emergence of the sympatric divergence event resulting in the two S and L lineages. This is particularly relevant since three mutations have been shown to be necessary for the emergence of the S lineage (Plucain et al. 2014) and we showed during this PhD thesis the involvement of at least a fourth mutation (Consuegra et al. submitted to mBio). The role of mutation rates in the emergence of adaptive diversification events can be studied both *in silico* and *in vivo*. *In silico* approaches may include computational models of evolution, as the one that we implemented (Großkopf et al. 2016). Implementations using varying mutations rates and/or varying times of appearance of the hypermutator phenotype might give us valuable information about the positive impact of high mutation rates in the evolvability of sympatric diversification in an unstructured environment. Similar approaches may be used *in vivo*, for example by replaying evolution with, as ancestors, isogenic strains except for the mutator alleles and analyzing its effect in the emergence of adaptive diversification. However, it must be emphasized that this type of experiments are time-consuming and, most of the times, their duration is uncertain to see the emergence of the relevant phenotypic traits. Computer simulations have predicted that the fitness of populations with high mutation rates increased faster during the process of adaptation and that once the adaptive peak was reached, the mutator phenotype will decline (Taddei et al. 1997). The fitness trajectories measured in the long-term evolution experiment show that although adaptation slows down it continues to increase even after 50,000 generations of evolution (Lenski and Travisano 1994, Wisser, Ribbeck, and Lenski 2013, Barrick et al. 2009). Mutator populations also tended to adapt faster than non-mutator populations (Wisser, Ribbeck, and Lenski 2013).

Genotypic reversion of the mutator allele in the Ara-2 population is “relatively easy” in the genetic sense, since the 6-bp insertion occurred in a mutation hotspot that consists in a 6-bp motif that is repeated 3 times and where homologous recombination may therefore occur during chromosomal replication. This reversion however occurred only late in evolutionary time, probably due to the complex ecological interactions of the S and L lineages. Indeed, adaptation

in Ara-2 is much intricate than in the other populations owing to the selective pressures exerted by the co-existence of both S and L lineages in the respective ecological niches where they adapted most, acetate and glucose, respectively. Hence, it has been shown that compensatory reduction of mutation rates occurred when populations were adapted to their environment (Wielgoss et al. 2013). Therefore, full adaptation to their respective favorite niches may take longer than the adaptation only to glucose in the other populations. This may explain the late reversion of the mutator phenotype, after tens of thousands of generations for both the S and L lineages.

Another fascinating feature of the reversion of mutation rates in Ara-2 is that it did not occur simultaneously in the two S and L lineages. Two hypotheses may be considered: (i) the first reversion of the mutator phenotype in the S lineage exerted a selective pressure for the reversion over the L lineage because the speed of evolution was then different between the two lineages, (ii) the reversion events were independent with the high mutation rates implying a too high genetic load in each of the two lineages when they were well adapted to their respective ecological niche. In favor of the second hypothesis, we detected a correlation between the growth rates in glucose or acetate (see Chapter 1; (Großkopf et al. 2016) and mutation rates. Hence, at the time of the S mutator reversion, at ~40,000 generations, the S growth rates in glucose stabilized and those in acetate were at their highest values. In contrast, at the time of mutator reversion in the L lineage at 50,000 generations, the L growth rates in glucose and acetate were at their maximum and minimal values, respectively.

In conclusion, the independent reversion of mutation rates in each of the two S and L lineages presumably occurred to reduce the genetic load imposed by a high mutation rate that became too strong when the flux of beneficial mutations of high effect was limiting, *i.e.* when each lineage was highly adapted to its favorite ecological niche. These dynamics suggests that mutation rate plays an important role in the evolvability of sympatric divergent lineages and that it can be fine-tuned as a continuous trait, according to the needs for adaptation.

Chapter 5: Role of Environmental Seasonality on Sympatric Bacterial Adaptive Diversification

In this chapter, I present data that resulted from a collaboration between our research team and the Beagle team of LIRIS/INRIA headed by Prof. Guillaume Beslon. During more than 5 years, the two teams have worked together to understand the mechanisms underlying evolutionary processes using a combination of *in vivo* and *in silico* experimental evolution. This collaboration includes a joint project called “Evo-Evo” that is funded by the European Commission under the FP7 program. This project has the objective to characterize the “evolution of evolution” through experimental evolution, bioinformatics and a computational framework for the design of individual-based *in silico* evolution models.

One of the interests of “Evo-Evo” is to model sympatric adaptive diversification to understand the underlying driving forces. I participated in this project during the three years of my PhD with the results of my research that were used to implement an *in silico* model intending to simulate the ecological bases of sympatric adaptive diversification. A first product of these combined approaches is presented in the following paper entitled, “*Beware batch culture: Seasonality and niche construction predicted to favor bacterial adaptive diversification*” (Rocabert C, Knibbe C, Consuegra J, Schneider D, Beslon G) which has been submitted to *PLoS Computational Biology*.

Adaptive diversification is an essential evolutionary processes since it produces the phenotypic innovation that allows the colonization of novel ecological niches. However, the ecological, genetic and molecular mechanisms explaining these phenomena are still poorly understood. Experimental evolution is a powerful tool to study evolutionary processes including sympatric adaptive diversification. However, experiments are still long, costly and include a high level of uncertainty to obtain exploitable results. To overcome this, computational models of *in silico* evolution, where artificial organisms evolve in a computer program for many generations, have been designed. These models usually include several scales (genome, phenotype and/or environment), allowing to mimic *in vivo* experiments,

although it has intrinsic limitation since evolution of real organisms implies the interaction of a wider range of biological structures and levels.

Here, we present a multiscale framework of *in silico* evolution. Bacteria-like organisms with their own genome encoding a regulatory and a metabolic network, evolve on a virtual medium for tens of thousands of generations. The virtual organisms are able to modify their environment by up-taking nutrients and releasing by-products, which leads to complex ecosystem evolution (Fig. 1 of the following paper). Using this model, we performed simulations in order to study the environmental conditions and driving forces that allow the occurrence of an adaptive diversification event as the one described for the Ara-2 population.

Random viable populations were evolved during ~40,000 generations in three different environments:

- *Seasonal*. Mimicking the LTEE, the organisms grew on a unique resource that was periodically provided every ~6-7 generations, and the rest of the environment is eliminated at the same time. Therefore, this was the equivalent of serial transfers in fresh medium as in the LTEE.
- *Continuous*. Mimicking chemostat evolution experiments, the organisms grew on the same resource that was provided in a continuous flow. The organisms produced free metabolites that were degraded through time. This simulation mimics experiments like those performed by Tom Ferenci (Maharjan et al. 2006).
- *Poisson*. As the seasonal one, except that the resource was provided (and the remaining environment was eliminated) at random time points following a Poisson law.

Comparison of the evolutionary outcome in the three experiments revealed important differences in the structure of the population. Organisms evolving in the seasonal LTEE-like environment often diverged into two sub-populations that co-existed for a very long period of

time. Conversely, organisms evolving in the continuous chemostat-like or in the Poisson environment evolved a single quasi-species (Figs. 2 to 6 of the following paper).

As observed for the adaptive diversification event in the Ara-2 population, the two *in silico* co-existing lineages in the seasonal environment corresponded to two different ecotypes. Their emergence was related to niche construction. The ecotype A (L-like) mainly consumed the primary resource present in the environment, while ecotype B (S-like) only consumed secondary resources that were produced and secreted by ecotype A (and possibly by ecotype B themselves). The long-term co-existence of the two ecotypes was due to negative frequency-dependent interactions (Fig. 8 of the following paper).

This study supports the essential role of seasonality in both temporal niche partitioning and promoting the emergence and long-term co-existence of ecotypes through cross-feeding and negative frequency-dependent interactions.

Beware Batch Culture: Seasonality and Niche Construction Predicted to Favor Bacterial Adaptive Diversification

Charles Rocabert¹, Carole Knibbe¹, Jessika Consuegra^{2,3}, Dominique Schneider^{2,3}, Guillaume Beslon^{1*}

1 Univ. de Lyon, CNRS, INRIA, INSA-Lyon, UCB Lyon 1, LIRIS UMR5205, F-69622 Lyon, France.

2 Univ. Grenoble Alpes, Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications, Grenoble (TIMC-IMAG), F-38000 Grenoble, France.

3 Centre National de la Recherche Scientifique (CNRS), TIMC-IMAG, F-38000 Grenoble, France.

* guillaume.beslon@inria.fr

Abstract

Metabolic cross-feeding interactions between microbial strains are common in nature, and emerge during evolution experiments in the laboratory, even in homogeneous environments providing a single limiting carbon source. In sympatry, when the environment is well-mixed, the reasons why emerging cross-feeding interactions may sometimes become stable and lead to monophyletic genotypic clusters occupying specific niches, named ecotypes, remain unclear. As an alternative to evolution experiments in the laboratory, we developed a multi-scale model of *in silico* experimental evolution, equipped with the whole tool case of experimental setups, competition assays, phylogenetic analysis, and, most importantly, allowing for evolvable ecological interactions. Digital organisms with an evolvable genome structure encoding an evolvable metabolic network evolved for tens of thousands of generations in environments mimicking the dynamics of real controlled environments, including chemostat or batch culture providing a single limiting resource. We show here that the evolution of stable cross-feeding interactions requires seasonal batch conditions. In this case, adaptive diversification events result in two stably co-existing ecotypes, with one feeding on the primary food and the other on by-products. We show that the regularity of the serial transfers is essential for the maintenance of the polymorphism, as it allows for at least two stable seasons and thus two temporal niches. A first season is externally generated by the transfer into fresh medium, while a second one is internally generated by niche construction as the provided nutrient is replaced by the secreted by-products of bacterial growth. In chemostat conditions, even if cross-feeding interactions emerge, they are not stable on the long-term because fitter mutants eventually invade the whole population. This study supports the crucial role played by seasonality in temporal niche partitioning and in promoting cross-feeding subgroups into stable ecotypes, a premise to sympatric speciation.

Author Summary

Stable bacterial cross-feeding interactions, where one strain feeds on the waste of the other, are important to understand, as they can be a first step toward bacterial

speciation. Their emergence is commonly observed in laboratory experiments using *Escherichia coli* as a model organism. Yet it is not clear how cross-feeding interactions can resist the invasion of a fitter mutant in a well-mixed environment with a single food, where there seems to be a single ecological niche. Here, we used another kind of “model organism”, a digital one, allowing for detailed and fast investigations, and providing a way to disentangle generic evolutionary mechanisms from the specifics of *E. coli*. We let digital organisms with evolvable genomes and metabolic networks compete for resources in conditions mimicking laboratory evolution experiments. In chemostat simulations, although cross-feeding interactions regularly emerge, selective sweeps regularly purge the population of its diversity. By contrast, batch culture allows for much more stable cross-feeding interactions, because it creates seasons and thus distinct temporal niches, thereby favoring the adaptive diversification of proto-species even in homogeneous environments.

Introduction

Stable metabolic cross-feeding interactions between microbial strains are commonly observed in nature [1–4]. For example, nitrification, an important step of the nitrogen cycle, is carried out in consecutive steps by several bacterial species maintaining cross-feeding interactions [3]. In laboratory experiments, microbial populations also demonstrated their ability to quickly establish metabolic cross-feeding interactions between morphotypes [5–13].

An important question, at the crossroads between ecology and evolution, is the evolutionary stability of such cross-feeding polymorphisms, because they are often considered as the first steps toward speciation. According to Cohan [14], the species concept in bacteria should not rely on the named species of systematics but on the notion of *ecotype*, which itself relies on the ecological and evolutionary dynamics of the subpopulations. Two bacterial subpopulations may be considered as different ecotypes if they form monophyletic clusters, occupy different ecological niches and if periodic selection purges diversity in one subpopulation independently from the other [14]. A cross-feeding polymorphism therefore leads to adaptive diversification and ultimately to speciation when it is stable enough to resist the invasion of a mutant that would otherwise take over the whole population.

If the environment is spatially structured, the stabilization of new ecotypes that emerged after an adaptive diversification event is facilitated by the locality of environmental conditions and frequency-dependent interactions. This mechanism of allopatric (or micro-allopatric) divergence is a well-known mechanism, since ecotypes can escape competitive exclusion in their local niches [14]. For example, *Pseudomonas fluorescens* populations have been shown to produce adaptive diversification events in spatially heterogeneous environments, but not in homogenized conditions [5,6].

However, microbial populations can also exhibit adaptive diversification in sympatry, when the environment is homogeneous with a single carbon resource. In this case, the stability of the ecotypes is maintained by frequency-dependent interactions, often due to cross-feeding interactions, as observed in the Long-Term Evolution Experiment with *Escherichia coli* (LTEE [15]). In this ongoing experiment, 12 populations are independently propagated in a constant glucose-limited environment in batch culture since 1988. The experiment reached 65,000 generations at the time of this writing. Everyday, 1% of the population is transferred in fresh medium such that each population experiences each day a succession of feast and famine phases before the next cycle. In one of the 12 populations, a long-term polymorphism has been observed [11]. Two ecotypes, named S and L (for Small and Large, related to their respective colony sizes on plate), evolved from a common ancestor before generation 6,500. The L ecotype

grows efficiently on glucose, while the S ecotype mainly grows on acetate, a by-product secreted by L. Experiments showed that the interaction between S and L ecotypes relies on negative frequency-dependent selection, meaning that each ecotype has a selective advantage when rare. This balanced polymorphism is stable for more than 55,000 generations [11]. It was also shown that S and L ecotypes specialized on their own niches, the L ecotype increasing its ability to grow on glucose but not on acetate, and conversely for the S ecotype [16].

The evolutionary stability of this polymorphism may be explained by the temporal niche partitioning that arises from the periodic transfers into fresh medium [17]. A first season starts immediately after a transfer, when the environment contains mostly glucose. The L ecotype grows during this season and transforms glucose into acetate, thereby generating a second season where the environment contains acetate and supports the growth of the S ecotype.

Yet several experiments have shown that microbial populations can also evolve cross-feeding interactions in a chemostat in a few tens of generations [7, 8, 10]. Those interactions appear to be stable over a few hundreds of generations [7, 8, 10]. In chemostat, there is no obvious spatial or temporal niche partitioning and it is thus intriguing that the dynamics predicted by the competitive exclusion principle has not been observed so far. Indeed, one would expect a mutant to eventually appear, which would either completely degrade glucose or feed on both glucose and acetate, thereby outcompeting the specialized ecotypes. It has been proposed that energy constraints and fluxes optimization principles prevented competitive exclusion, thereby stabilizing the polymorphism [18, 19]. However, experimental evolution in chemostat has generally been performed for only a few hundreds of generations (up to 1,900 generations in [7]), precluding the possibility to confirm this statement on a longer term.

Thus, as a step to better understand how cross-feeding, niche construction and seasonality contribute to microbial diversification, we addressed here the following question: What makes emerging cross-feeding interactions stable in the long-term, in single carbon source batch culture or chemostat experiments?

While experimental evolution provides a very precise picture of evolution, it remains a long and costly process. An alternative approach consists in simulating evolution in a computer. *In Silico* Experimental Evolution (ISEE), where digital organisms are evolved for tens of thousands of generations, reproduces environmental conditions of experimental evolution [20]. Like in the wet approach, it is possible to simulate several independent populations to understand the respective importance of general laws and historical contingencies. In addition, ISEE provides an exhaustive fossil record and, more importantly, allows for 'impossible experiments' [21], like saving the fitness at full resolution for tens of thousands of generations, or changing any parameter (mutation rates, environment fluxes) at will.

Computational models of *in silico* experimental evolution have already been used to explore the evolution of cross-feeding interactions. Johnson and Wilke [22] studied the evolution of resource competition between two digital species coexisting via a mutualistic cross-feeding. Williams and Lenton [23] used an individual-based evolutionary model to explore the stability of connected ecosystems undergoing cross-feeding and 'evolutionary regime shifts'. Crombach and Hogeweg [24] and Boyle et al. [25] studied the evolution of resource cycling and its stability. Chow and colleagues [26] explored the relation between productivity and diversity in a digital ecosystem under mixed resources influx, while Pfeiffer and Bonhoeffer [18] and Gerlee and Lundh [19] explained the maintenance of cross-feeding interactions in a microbial population by energy and efficiency constraints on metabolic fluxes. Gerlee and Lundh [27] also related ecosystem productivity to energy-uptake efficiency. Finally, Großkopf et al. [16] predicted the adaptive diversification event leading to S and L

ecotypes in the LTEE, by mixing flux balance analysis and *in silico* evolution in a single model. While those models helped to decipher the evolution of cross-feeding interactions, most of them included only two or three scales (typically the genome, phenotype and environment), limiting their possibility to mimic *in vivo* experiments, since evolution of real microorganisms implies the interaction of a wide range of biological structures and levels.

We developed a new multi-scale computational model of *in silico* experimental evolution. This model allows us to study many questions raised by experimental evolution [20]. Typically, we can use it to investigate how evolution shapes the different structures of an organism (*e.g.*, genome size, complexity of the regulation network and of the metabolic network) and of an ecosystem (polymorphism, speciation) depending on global parameters such as environmental conditions or mutation rates.

Here, we studied the environmental conditions in which cross-feeding may lead to stable adaptive diversification events, by reproducing the resource dynamics of experimental evolution setups like chemostat or batch culture. Our results show that *stable* cross-feeding interactions are favored in batch culture, owing to the seasonality of the environment. In continuous culture, the absence of seasonality forbids niche construction and leads to competitive exclusion, even if the population is initially composed of two ecotypes maintaining frequency-dependent interactions.

Model

Our multi-scale computational model is individual-based. Digital bacterial-like organisms own a coarse-grained genome made of units encoding for a simplified metabolic network and possibly a genetic regulatory network. Those organisms evolve on a two dimensional toroidal grid, uptaking, transforming, and releasing metabolites, dividing in empty spots or dying. In this model, metabolites are implicit molecules identified by a tag $\in \mathbb{N}^*$. The model is described in more details below, and summarized in Fig 1. The source code is written in C++. All the material necessary to replay experiments (software, parameter files, strain backups, ...) is freely available at <http://www.evoevo.eu/adaptive-diversification-simulations/>.

Genome structure.

The genome is a circular single-strand sequence of units, inspired from [28, 29]. Units belong to five different types: non-coding (NC), promoter (P), binding site (BS), enzyme coding unit (E) and transcription factor coding unit (TF). The order of the units in the genome determines the existence of functional regions, meaning that not all sequences of units are functional. The functional regions of a genome are those that have the following pattern: a promoter (P), optionally flanked upstream or downstream by one or more binding sites (BS), followed by one or more coding units (E or TF). A promoter can thus control several coding units, like in bacterial operons. The first unit that is not a coding one interrupts the transcription and marks the end of the functional region.

Non-coding units have no particular function. They compose the non-coding part of the genome. Promoter units contain a floating-point number $\beta \in [0.0, 1.0]$ representing the basal production rate of the protein(s) controlled by the promoter. Enzyme units contain two integers s and $p \in \mathbb{N}^*$, indicating the tag of the substrate and product respectively, and two floating-point numbers $k_{cat} \in [10^{-3}, 10^{-1}]$ and $k_m \in [10^1, 10^3]$ describing the enzymatic kinetics (see the description of the metabolic network below). In the special case where $s = p$, the enzyme is considered as a pump, actively pumping in (resp. out) the metabolite s if k_{cat} is positive (resp. negative). For sake of simplicity, we do not describe here the details of the binding site and transcription factor units, as regulatory networks did not evolve in the present simulations. Initial genomes

Fig 1. Presentation of the model. The genotype-to-phenotype mapping, as well as the population and environment, are schematized here. **(A)** Description of the genotype-to-phenotype mapping. Organisms own a coarse-grained genome made of units. Non coding (NC) units are not functional (A.1). The arrangement of the units on the circular single strand defines functional regions, where a promoter (blue cross, A.2) controls the expression of all contiguous enzyme units (E, red circles), thereby allowing for operons. When enzyme units are expressed (A.3), they contribute to the metabolic network. Enzymes perform metabolic reactions in the cytoplasm, or pump metabolites in or out (A.4). The score of an organism is computed from its ‘essential metabolites’ (usually the score is the sum of essential metabolite concentrations). Lethal toxicity thresholds are applied to each metabolic concentration and forbid organisms to accumulate resources. **(B)** Description of the population and environment levels. Organisms are placed on a 2D toroidal grid, and compete for resources and space. When an organism dies, it leaves its grid cell empty and organisms in the Moore neighborhood (if any) compete to divide in available space. The competition is based on scores, a minimal threshold being applied on scores to forbid worst organisms to divide. At division, daughters share cytoplasm content (enzymes and metabolites). At death, metabolites from the cytoplasm are released in the local environment, and diffuse on the grid (B.1). On the largest scale, the population evolves on the environment by uptaking, transforming and releasing metabolites. Metabolites then diffuse and are degraded. This strong interaction between the population and the environment allows for the evolution of complex ecological situations, depending on environmental properties (B.2).

containing 50 units are generated by creating ten units of each type with random attribute values. To start an evolutionary run, different initial populations with such random genomes are tested until a viable population is found.

Upon cell division, the parental genome is replicated with mutations in the two daughter cells. Each genomic unit can undergo point mutations, meaning here changes in the numbers it contains, like the s , p , k_{cat} and k_m values for an enzyme unit. Each unit attribute mutates at a rate 10^{-3} . s and p mutate in a discrete uniform law $\in [-1, 1]$. k_{cat} and k_m mutate in log-scale, in a Gaussian law of standard deviation 0.01. β mutates in a Gaussian law of standard deviation 0.1. A unit can also undergo a type transition from any unit type to any other at a predefined rate, set here to 10^{-3} . The genome can also undergo large rearrangements affecting segments of several units (duplications, deletions, translocations and inversions). For each type of rearrangement, breakpoints are drawn at random in the whole genome (*e.g.*, 3 breakpoints for a duplication, 2 breakpoints for a deletion...). Breakpoints are drawn between two units, but are considered to break one of the two neighboring units. The randomly selected one then undergoes a point mutation and a type transition with rate 1.0. All rearrangement rates are set to 10^{-3} per unit, hence the number of rearrangements depends on the genome size thus limiting genome expansion [30].

Metabolic network.

Expressed enzymes perform catalytic reactions in the metabolic space. Each enzyme, at concentration $[E]$ in the cytoplasm, catalyzes one specific reaction $s \rightarrow p$, with $s \in \mathbb{N}^*$ and $p \in \mathbb{N}^*$ being the substrate and the product of a Michaelis-Menten-like reaction, respectively. In the absence of regulation of gene expression (as this is the case in the runs presented here), the evolution of the concentrations $[E]$, $[s]$ and $[p]$ over time is then defined by Eq 1:

$$\begin{cases} \frac{d[E]}{dt} = \beta[E] - \phi \\ \frac{d[s]}{dt} = -\frac{k_{cat}[E][s]}{k_m + [s]} \\ \frac{d[p]}{dt} = \frac{k_{cat}[E][s]}{k_m + [s]} \end{cases} \quad (1)$$

where β is the basal production rate specified in the promoter unit, ϕ is the enzyme degradation rate (set to 0.1 per minute for all enzymes here), k_m and k_{cat} are the kinetic attributes of the enzyme. At each time-step, for each organism, the internal concentrations of enzymes and metabolites are updated by computing the dynamics of its metabolic network during 100 simulated minutes using the adaptive Runge-Kutta-Cash-Karp method (RKCK). Thus, when the 32x32 grid is full of organisms, a time-step involves the computation of about a thousand of ODE systems. The parameter values of each ODE system are potentially unique, as they are encoded in the organism's genome and thus result from the mutation process. Those ODE systems can also differ by their number of equations, which depends on the number of functional enzymes found in the organism's genome.

Score function.

Some metabolites are essential for an organism's replication. Here, we arbitrarily defined metabolites whose tag is a prime number as essential. The score of an organism is then simply defined as the sum of its internal concentrations of essential metabolites. However, to prevent organisms from producing a single specific prime number in huge quantities, we also predefined lethal toxicity thresholds for both essential and non essential metabolites. Here these toxicity thresholds were set to 1 for all metabolites. We set the score of an organism to 0 if at least one metabolite exceeds its toxicity threshold. Upon cell division, the two daughter cells share the cytoplasmic content of the mother cell (enzymes and metabolites).

Population and environment.

Organisms evolve on a two-dimensional toroidal grid, each spot containing at most one organism. The physical environment is described at the grid level: each grid spot contains external metabolites (outside of organism's cytoplasm), each with its concentration. Those external metabolites diffuse with a diffusion rate D and are degraded with a degradation rate D_g . Organisms compete for the external metabolites and to produce offspring in empty spots. They interact with their local environment by pumping metabolites in and out and releasing their metabolic content at death. At each time-step, organisms are evaluated and either killed, updated or replicated depending on their current state:

1. Death probability follows a Poisson law of parameter P_{death} . At death, the metabolic content is released in the local environment,
2. If the organism does not die and cannot divide (*e.g.*, because there is no free space in its neighborhood), its metabolic network is updated, and its score is computed,
3. For each empty grid spot, all alive organisms in the Moore neighborhood whose score is higher than a minimum score compete, depending on relative scores, to select the replicating organism. Here, only the organism having the best score in the neighborhood is allowed to divide. The minimum score was set to 10^{-3} .

Experimental protocol

In all simulations, the environment provides one primary food which tag $m_{exo} = 10$. Initial organisms are generated at random, by bootstrapping the simulation until a viable population is found, *i.e.*, a population containing organisms possessing at least one pump to pump some 10 in, and (because 10 is not a prime number) at least one enzyme to transform 10 into a prime number, thereby producing an ‘essential metabolite’. Those organisms then grow on the primary food and start to release by-products (mostly at death), hence modifying their environment. We evolved populations in two different environments:

1. In the **periodic environment**, the resource dynamics of the LTEE [15] environment is mimicked. The environment is periodically refreshed by removing all the metabolites it contains and introducing m_{exo} at concentration $f_{in} = 10.0$ per grid spot. Metabolites that are inside the organisms (internal metabolites) are not affected by the refresh event. The refresh period is $\Delta t = 333$ time-steps, and corresponds to one cycle.
2. In the **continuous environment**, the resource dynamics of a chemostat environment is mimicked. The medium is constantly provided with a small influx of the primary resource and all the external metabolites are slowly degraded. Specifically, at each time-step, a concentration $\Delta f_{in} = 0.03$ of m_{exo} is added in every grid spot, and external metabolites disappear at rate $D_g = 0.003$.

For each type of environment, 12 independent populations have been propagated for 500,000 time-steps (approximately 50,000 generations). On the long-term, the quantity of resources available in the system is equivalent in both environments. The grid size is 32×32 . Complementary experiments have also been run in a randomized batch environment, similar to the periodic environment but where the environment reset intervals follow a Poisson law of parameter $\Delta t = 333$ time-steps, rather than occurring with an exact regular period of 333 time-steps. In all simulations, the death probability is set to $P_{death} = 0.02$ and the diffusion rate to $D = 0.1$. The simulation parameters common to all the simulations are described in S1 Table.

Trophic networks.

In order to recover the metabolic activity and the potential cross-feeding interactions in the population, the trophic network is computed at every time-step. For each organism, a ‘trophic profile’ is computed from its metabolic network activity. The trophic profile is a bit string summarizing the uptake, production, and release activity of an organism. The length of the bit string is defined by the largest metabolite tag present in the system at time t . For example, if an organism uptakes metabolite 4, produces 3 from 4 and releases 3, knowing that the largest metabolite tag in the system is 5, then its profile is $|00010|00100|00100|$. Organisms with identical trophic profiles are grouped together, and the trophic network is computed depending on profile relationships. For example, if organisms of a profile i pumps in a metabolite produced by a profile j , then a directed link is created from i to j . We classified the trophic profiles in two trophic groups:

1. “Group A” feeds on m_{exo} , and possibly on other metabolites,
2. “Group B” feeds on group A by-products, and possibly on other metabolites, but not on the primary food m_{exo} .

In some cases, we also distinguished a subgroup of group A: “pure group A”, that exclusively feeds on the exogenous nutrient m_{exo} provided as primary food. As defined in the introduction, a trophic group is considered as an ecotype if the organisms of the group form a monophyletic cluster (see below).

Phylogenetic relationships.

Phylogenetic relationships were exhaustively recorded during each simulation. Hence, it is possible to recover the line of descent of any organism, and to compare the phylogenetic tree structure with the distribution of the trophic groups in the population. In particular, we can determine if groups A and B are monophyletic, and thus can be considered as ecotypes. To this aim, we computed a phylogenetic structure score (PS score) to identify the degree of monophyly of both groups. This phylogenetic structure score is defined as $PS = |f_1 - f_2|$, where f_1 and f_2 are the relative frequencies of group B in both subtrees rooted to the last common ancestor of the whole final population. A high PS value indicates a strong clustering of groups A and B on the phylogenetic tree, *i.e.*, groups A and B are ecotypes. A low PS value indicates a random distribution or the absence of polymorphism.

Results

Starting from random organisms, we did not observe the emergence of genetic regulation networks in those simulations, in coherence with Weisse et al. [31], who showed that regulation is triggered by internal trade-offs linked to limitations in the levels of cellular energy, free ribosomes and proteins. We thus focused on the study of metabolic networks and ecological interactions. First, the global evolutionary dynamics of the system can be analyzed by looking at main simulation statistics. The evolution of the mean score, the environmental richness (the number of different metabolites available in the environment), the number of trophic profiles in the trophic network, and the proportion of group A or B organisms are represented in Fig 2. The score and the environmental richness are of the same order of magnitude in the continuous and in the periodic environments, but they are more stable in the continuous environment. The number of trophic profiles shows no striking difference between the periodic and the continuous environment (Fig 2A.3 and 2B.3), indicating that polymorphism is common in both situations. However, the dynamics of groups A and B are completely different. In the periodic environment, groups coexist, even if they present long-term frequency variations (Fig 2A.4). In the continuous environment, group B disappears in all cases (Fig 2B.4). Thus, even if the diversity of trophic profiles is similar in both environments, there is no group exclusively specialized on by-products in the continuous environment, while they are common in the periodic one.

Fig 2. Evolution of some typical variables. The evolution of the mean score (A.1 and B.1), the environmental richness (the number of different metabolites present in the environment, A.2 and B.2), the number of trophic profiles in the trophic network (A.3 and B.3), and the proportion of groups A or B organisms are represented (A.4 and B.4). (A) Evolution in the periodic environment. (B) Evolution in the continuous environment. In the groups A/B proportion figures, group A is represented in blue, and group B in green.

Final state of the simulations

For each simulation, we analyzed the final phylogenetic tree, and compared it to the distribution of groups A and B. All the phylogenetic trees are represented in Fig 3. Leaves are colored depending on their trophic group (group A in blue, group B in green). It appears that the structure of the trees strongly depends on the type of environment. In the periodic environment (Fig 3A), 6 phylogenetic trees among 12 (repetitions 1, 2, 3, 6, 10 and 12) show two well-separated clusters, each belonging to

one group. In those repetitions, two ecotypes evolved separately and remained stable on the long-term, showing that a stable cross-feeding interaction evolved. In the six other cases, trees are less deep, have no well separated clusters and there is no clear correlation between trophic groups and phylogenetic structure. In the continuous environment (Fig 3B), trees are much shorter than in the periodic environment. Group A went to fixation in all repetitions. Then, while polymorphism and cross-feeding existed at a similar level in both periodic and continuous environments (Figs 2A.3 and 2B.3), this polymorphism was not stable in the continuous environment.

Fig 3. Final phylogenetic trees of each simulation. (A) Phylogenetic trees of the 12 repetitions in the periodic environment. (B) Phylogenetic trees of the 12 repetitions in the continuous environment. Tree leaves are colored depending on their trophic group: group A in blue, group B in green. Phylogenetic trees are numbered by repetition.

Evolution of phylogenetic structure and trophic groups

To get more insight into the evolutionary dynamics, we computed the distribution of the Most Recent Common Ancestor (MRCA) age at each time-step and for all the simulations. The MRCA age reflects the stability of the adaptive diversification of ecotypes in a population. As shown in Fig 4, distributions confirm that deepest trees evolved in the periodic environment, with a mean coalescence time of 90,198 time-steps, and a large distribution tail (some trees having almost the same depth than the total simulation time - 500,000 time-steps). The mean coalescence time is 18,910 time-steps in the continuous environment. All pairwise Student tests are significant (p-value < 0.001). The MRCA age has also been computed for the complementary experiments in the random environment, the mean coalescence time being 16,703 time-steps. This result indicates that environmental variations must be regular to favor stable cross-feeding interactions. The evolution of the MRCA age during simulations is also represented in S1 Fig, for the three types of environment. This figure gives a better idea of the phylogenetic trees evolution dynamics, and shows that the MRCA age regularly collapses in the random and continuous environments, but is maintained for some simulations in the periodic environment.

Fig 4. Distribution of the Most Recent Common Ancestor age in all the simulations. For each type of environment, distributions of the Most Recent Common Ancestor (MRCA) age are computed, for each simulation time-step. For each environment, all the repetitions are included. All pairwise Student tests are significant, with Bonferroni correction (p-value < 0.001/3).

We then compared the phylogenetic structure with the distribution of groups A and B on tree leaves by computing the phylogenetic structure score *PS* (see Experimental Protocol). In Fig 5, this *PS* score is plotted against the coalescence time every 5,000 time-steps and for all the repetitions. In the periodic environment (Fig 5A), the deepest trees are also the most structured, with two well separated monophyletic ecotypes A and B. In the random environment (Fig 5B), the situation is contrasted, with a large distribution of the *PS* score, ranging from monomorphic trees (A or B groups being fixed), to polymorphic trees. However, the coalescence time is very short relatively to the periodic environment, revealing the instability of the phylogenetic structure (note that the random environment is the only one where we observed population extinction, in 4 populations out of 15. In the continuous environment (Fig 5C), the population is mostly monomorphic (the group A being fixed), with short coalescence times.

Fig 5. Phylogenetic structure score against the coalescence time. For each type of environment, the phylogenetic structure score (PS score) is plotted against the coalescence time every 5,000 time-steps, for all the repetitions. A high *PS* value indicates a strong clustering of groups A and B on the phylogenetic tree (*i.e.*, A and B are monophyletic ecotypes). A low *PS* value indicates a random repartition or the absence of polymorphism. **(A)** Periodic environment. **(B)** Random environment. **(C)** Continuous environment.

Once again, those results confirm that the periodic environment strongly favors the evolution of stable cross-feeding interactions, and that it is not the case for the random or the continuous environments, in apparent contradiction with the results of wet experiments in chemostat.

Evolution of trophic profiles

We then recovered the proportion of trophic profiles over time (at every 5,000 time-steps), in all the simulations of the periodic and continuous environments (Figs 6A and 6B, respectively). Trophic profiles belonging to group A are colored in shades of blue (purple indicating pure group A), those belonging to group B in shades of green. Those figures show that evolution in the model is ruled by periodic selection in a highly polymorphic population. This polymorphism is mainly due to the competition for resources, organisms constantly competing for the primary food but also for the by-products available in the environment (Fig 2).

However, in the periodic environment (Fig 6A), trophic profiles from groups A and B coexist over time, with periodic selection events occurring independently in both groups. This dynamics is typical from multiple niche selection, where beneficial mutations do not spread in all the population owing to competitive exclusion, but are confined in one specific niche. In the continuous environment (Fig 6B), group A is predominant in all the simulations, periodic selection affecting the whole population. In these conditions, the level of cross-feeding is maintained but the interactions are not stable.

These results reinforce the fact that stable cross-feeding interactions are only detected in the periodic environment. Specifically, in the periodic environment, evolution is driven by multiple niche selection, with periodic selection events independently occurring in ecotypes A and B. On the opposite, in the continuous environment, evolution is driven by periodic selection and competitive exclusion, indicating that there is less opportunity for niche construction.

Fig 6. Evolution of trophic profiles in the population, for the continuous and the periodic environments. Trophic profiles gather organisms that own the exact same metabolic activity (see methods). Blue profiles belong to trophic group A (purple profile being pure A). Green profiles belong to trophic group B. **(A)** Continuous environment simulations. **(B)** Periodic environment simulations.

Ecological dynamics in the periodic environment

Comparative analysis of phylogenetic structure in the different environments revealed that the periodic environment especially favors the evolution of stable cross-feeding interactions, leading to two monophyletic ecotypes A and B in 6 of 12 repetitions, ecotype A feeding on the primary resource and possibly on some by-products, while ecotype B only feeds on by-products. In the LTEE, it has been shown that the coexistence of S and L ecotypes is driven by negative frequency-dependent

interactions [12,13]. We analyzed in details the 6 populations to see whether the stable cross-feeding interactions were comparable to the S/L interaction.

Mutational history of ecotypes A and B.

In the 6 populations that evolved monophyletic ecotypes at this end of the simulations, we recovered the mutational history of ecotypes A and B lineages. Final phylogenetic trees of the 6 populations are represented in Fig 7. For each tree, the trophic group of the MRCA, as well as the generation at which one of the monophyletic ecotypes switched from the ancestor group to the other one (*i.e.*, when one ecotype lost or gained inflow pumps for the primary food), are shown. In most of the 6 populations, the same general pattern emerged: the population was primarily of group A, but niche construction on by-products led to an adaptive diversification, with one ecotype strongly specializing on by-products, such that it lost the ability to feed on the primary food.

Interestingly, in all the simulations, the loss of this ability is not the source of the adaptive diversification. The diversification event occurred a few hundreds of generations *before* the loss of the pump provoking the change of trophic group. In the LTEE, S ecotype specialized on acetate, but is still able to grow on glucose. However, recent work has shown that while S ecotype improved its ability to grow on acetate since the diversification event, it was not the case on glucose, presaging a possible complete loss of its ability to grow on glucose on a longer term [16]. Reversely, L ecotype improved its ability to grow on glucose, but not on acetate, also presaging a loss of its ability to grow on acetate.

Fig 7. Analysis of the adaptive diversification event leading to monophyletic ecotypes A and B. In the phylogenetic trees of final evolved populations, the colored circles indicate the trophic group of the common ancestor. The colored triangles indicate the moment when one monophyletic ecotype moved from one trophic group to the other (*i.e.*, losing or gaining pumps to feed on external nutrient). Group A (blue) grows on the primary food and possibly on by-products. Group B (green) exclusively grows on by-products.

Ecotype B frequency-dependent fitness in short term competition experiments.

To test whether ecotypes A and B coexistence is maintained by a negative frequency-dependent interaction, we performed short term competition experiments with the 6 populations of the periodic environment that evolved a stable cross-feeding interaction at the end of the simulations (repetitions 1, 2, 3, 6, 10 and 12). Initial populations were seeded with 9 different initial frequencies of B (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 – each with 10 repetitions) and have been propagated in the same periodic environment than evolved ecotypes during 10 cycles (*i.e.*, 3,330 time-steps). Then, we computed the log-fitness [32] of ecotype B, taking into account its initial frequency and its frequency after 10 cycles. Since the relative frequencies of ecotypes A and B depend on the advancement of the cycle, we measured the log-fitness all along the last cycle. Fig 8 demonstrates that the ecotypes A and B interaction is frequency-dependent, the ecotype B being favored when initially rare, and penalized when initially abundant. However, even if the B relative fitness always negatively depends on its initial frequency (the slope of the curve is constant), the global height varies during the cycle, with a minimum at approximately 117 time-steps (35% of a cycle, Fig 8A), and a maximum at the end of a cycle (333 time-steps, Fig 8B). Indeed, at each cycle, B ecotype growth is delayed compared to A ecotype, the former growing

on by-products during the second season, while the latter grows on fresh primary resource during the first season. S1 Video shows the variation of B relative fitness all over the 333 time-steps of the last cycle, at a full temporal resolution. Those results are in full agreement with the LTEE results [11,33].

Fig 8. Frequency-dependent relative fitness in short-term competition experiments. The frequency-dependent fitness has been computed using log-fitness [32,33] in short term competition experiments, starting with different initial frequencies of B ecotype and running for 10 cycles. For each of the 6 populations that evolved monophyletic ecotypes at the end of the simulations, 10 repetitions have been run per initial frequency of B (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9). The global mean frequency-dependent fitness is represented in black. Mean fitness per population are shown in shaded colors. Each individual experiment is plotted in shaded color dots, related to their mean color. **(A)** B relative fitness measured 117 time-steps after the last environmental refresh, when B relative fitness is at its lowest state. **(B)** B relative fitness measured at the end of the last cycle (highest B relative fitness state of the cycle).

Convergence to an oscillatory dynamics.

Owing to their negative frequency-dependence interaction, the relative frequencies of ecotypes A and B should stabilize over time, as in the LTEE [11]. We extracted from the previous competition experiments the evolution of ecotypes A and B proportions, at each time-step during the 10 cycles. The result is presented in Fig 9. Trajectories show that for all initial frequencies of B, a stable oscillatory dynamics is reached, whatever the considered repetition (Fig 9A for repetition 1, Fig 9B for rep. 2, Fig 9C for rep. 3, Fig 9D for rep. 6, Fig 9E for rep. 10 and Fig 9F for rep. 12). The observed variability is due to contingent evolutionary differences between the 6 populations, and to a sampling effect when the initial frequency of B is low. Here again we observed exactly the dynamics observed in the LTEE [11,33], even if the small population size increases oscillations due to organisms mortality.

Fig 9. Convergence to an oscillatory dynamics. Convergence to an oscillatory dynamics over 10 serial transfer cycles. Ecotype B is advantaged when rare, but is penalized when initially common, leading to a balanced polymorphism. Nine different initial frequencies of B have been tested (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9). Each trajectory is the mean of the B frequency among the 10 repetitions of each of the 6 populations. The standard deviation of each trajectory is plotted in shadowed colors. **(A)** Repetition 1. **(B)** Repetition 2. **(C)** Repetition 3. **(D)** Repetition 6. **(E)** Repetition 10. **(F)** Repetition 12.

Stability of the A/B cross-feeding interactions in the continuous environment

Our previous results have shown that in the continuous environment, no stable cross-feeding interaction evolved. This result is in apparent contradiction with wet experiments evolving *E. coli* populations in a continuous culture with glucose as a single limiting resource [7,8,10]. In those experiments, cross-feeding interactions emerged over a few hundreds of generations. Nonetheless, our results showed that digital populations evolving in the continuous environment resulted in cross-feeding interactions (Fig 2), but these interactions were not stable (Fig 4).

To test whether a population with two stable A and B ecotypes (evolved in the periodic environment) could persist in the continuous one, we let populations from the periodic environment evolve in chemostat-like environments for 50,000 time-steps. Each of the 6 populations that evolved a stable polymorphism at the end of the simulations in the periodic environment were transferred in a continuous environment at two different stages of their evolution: **(i)** just after the adaptive diversification (early populations, see Fig 7), **(ii)** and at the end of the simulations (late populations, after 500,000 time-steps). As a control, these populations were also propagated in the periodic environment. For each population, 10 repetitions were run in each environment. Then, we evaluated the stability of the A/B cross-feeding interaction by counting the number of simulations where the interaction persisted, the number of simulations where the interaction failed, and the time before interaction failure.

The proportion of simulations where the interaction persisted are displayed in Table 1 for the continuous environment, and in Table 2 for the periodic environment. The evolution of the proportions of groups A and B is also shown in Figs 10A and 10C for the continuous environment (early and late populations, respectively), and in Figs 10B and 10D for the periodic environment (early and late populations, respectively). First, Table 1 shows that, for early populations in the continuous environment, the interaction was not robust and persisted in only 30% of the assays. For late populations in the continuous environment, the interaction was more robust, as the polymorphism persisted in 49% of the assays. Population 6 is an exception, as ecotype B went extinct in all repetitions. However, Fig 7 indicates that in this population, the adaptive diversification occurred late in the simulation, and the interaction probably did not evolve enough to be stable at the end of the simulation (late populations). In the periodic environment, most of the assays kept the polymorphism stable, whenever the ecotypes evolved on the long-term or not, indicating that seasonality is of primary importance to stabilize the interaction.

Table 1. Proportion of assays where polymorphism persisted in chemostat conditions.

	Pop 1	Pop 2	Pop 3	Pop 6	Pop 10	Pop 12
Early populations	0%	60%	0%	0%	90%	30%
Late populations	40%	30%	80%	0%	100%	40%

For each population, 10 assays were simulated. The polymorphism was considered to be lost if the proportion of one of the ecotypes reached 0.0 during the simulation (the corresponding monophyletic group was outcompeted).

Table 2. Proportion of assays where polymorphism persisted in batch conditions.

	Pop 1	Pop 2	Pop 3	Pop 6	Pop 10	Pop 12
Early populations	90%	80%	100%	90%	100%	100%
Late populations	80%	100%	70%	100%	100%	90%

For each population, 10 assays were simulated. The polymorphism was considered to be lost if the proportion of one of the ecotypes reached 0.0 during the simulation (the corresponding monophyletic group was outcompeted).

Fig 11 shows the distribution of the time before A/B interaction failure for early and late populations, in the continuous environment. It appears that late populations were much more robust, since extinctions happened significantly later (with a mean of 23858.9 time-steps) than for early populations (with a mean of 12687.24 time-steps). Student test gives a p-value < 0.001.

In order to understand why the A/B interaction failed in half of the continuous environment experiments (51% of the assays for the late populations), and why the A/B interaction failures imply the extinction of ecotype B in most of the cases (90% of the failures for the late populations), we studied in details the evolution of digital organisms.

Fig 10. Stability of the A/B interaction evolved in the periodic environment, when placed in the continuous one. Early populations have been transferred just after the adaptive diversification. Late populations have been transferred at the end of the simulations (500,000 time-steps). For each repetition that evolved two ecotypes A and B (rep.1, 2, 3, 6, 10, and 12), 10 repetitions were run. If the A/B interaction is maintained, the simulation is colored in green. The interaction was considered to be lost (in red) if the proportion of one ecotype reached 0.0 during a simulation (the corresponding monophyletic group was outcompeted). **(A)** Early populations transferred in the continuous environment. **(B)** Early populations transferred in the periodic environment. **(C)** Late populations transferred in the continuous environment. **(D)** Late populations transferred in the periodic environment.

Fig 11. Time before A/B interaction failure in the continuous environment. This time is measured for all the competition experiments in the continuous environment (60 simulations). Early populations lose the interaction significantly quicker than the late ones (Student test is significant with a p-value < 0.001).

For each type of environment, we first computed the distribution of the variation in the number of essential metabolites produced and pumped in by group A organisms, between the beginning and the end of the assays, over all the assays. For each measure, we performed a one-sample Wilcoxon test against a theoretical mean of 0.0 (observed distributions being not Gaussian). Because two tests were performed per population (one for the production activity, another for the uptake activity), a Bonferroni correction ($n = 2$) was also applied on each test. After 50,000 time-steps of evolution in chemostat conditions, ecotype A organisms significantly modified their metabolism, by diminishing their production (mean variation of -1.4264, with a p-value of 1.278e-09 ***). However, there is no significant variation in the number of essential metabolites pumped in (mean variation of +0.5784, with a p-value of 0.036). This variation is also observed in periodic environment assays (mean of -0.6049, with a p-value of 7.067e-05 *** for the production variation ; mean of +0.2986, with a p-value of 0.1103 for the uptake variation), but the loss of production is significantly higher in the continuous environment (comparing production distributions between the periodic and the continuous environments with a Wilcoxon test for two samples gives a p-value of 6.685e-05 ***).

We next computed the same distributions in two situations: **(i)** When the ecotype B went extinct (between the beginning of the assay and 1,000 time-steps before ecotype B extinction) **(ii)** When the ecotype B survived. We then compared both distributions, in the continuous and the periodic environments, by performing a two-samples Wilcoxon test (with a Bonferroni correction of $n=2$). Results are striking: in the continuous environment, the loss of production is significantly higher in ecotype A organisms just before ecotype B extinction (mean variation of -1.6615) than when ecotype B survived (mean variation of -0.598), with a p-value of 0.001 (**). The difference of variation is not significant in the case of the variation of the uptake activity (p-value of 0.1932). The extinction of ecotype B in most of the continuous environment assays may thus be explained by the loss of essential metabolites produced by ecotype A. Since ecotype B relies on A's metabolites production, this deprivation is highly deleterious for ecotype B organisms, except if they themselves modify their metabolism to adapt to the new conditions. Indeed, we computed the distributions of the variation of essential metabolites produced or pumped in by ecotype B organisms, when the A/B interaction persisted. In this case, ecotype B organisms significantly reduced their production activity (mean variation of -0.5276, with a p-value of 0.0089 *), and increased their uptake activity (mean variation of +0.5121, with a p-value of 0.0015 **). This significant change in the metabolism of surviving ecotype B organisms indicates a

restructuring toward more generalist organisms. In the periodic environment, the low number of ecotype B extinctions (6 assays out of 60) precludes any statistical analysis.

To exemplify these statistical results, we studied in details the evolution of ecotypes A and B in the 10 repetitions of the population 1, when propagated in the continuous environment. In this population, at the beginning of the assays, ecotypes A and B interact via a negative frequency-dependent cross-feeding: ecotype A organisms produce essential metabolites 2, 3, 5, 7, 13 and 19; ecotype B organisms feed on metabolites 3 and 5 (produced by ecotype A organisms) and produce essential metabolites 2, 7, 13 and 19 from them. We precisely evaluated the evolution of the 6 essential metabolites that were produced by ecotype A organisms at the beginning of the assays. The result is presented in Fig 12. With the exception of the repetition 3, where the whole population went extinct, ecotype A organisms reduced their essential metabolites production in all but one of the assays. However, when ecotype A organisms stopped producing metabolites 3 and/or 5, ecotype B organisms systematically went to extinction. On the opposite, when ecotype A organisms stopped producing metabolites 17 and/or 19 (but maintaining the production of 3 and 5), ecotype B was not affected. These results confirm the mechanism of B extinction: when placed in continuous conditions, ecotype A organisms reorganized their metabolism and produced fewer essential metabolites. Now, while doing so they may have randomly lost metabolites that where necessary for the survival of ecotype B organisms, leading to their extinction.

Fig 12. Loss of essential metabolites production of ecotype A organisms, in the 10 repetitions of the population 1 in the continuous environment assays. The 10 repetitions of the population 1 are displayed. The 6 essential metabolites (2, 3, 5, 7, 17, and 19) that were produced by ecotype A organisms at the beginning of the assays are represented vertically for each repetition. Colored metabolites indicate a production loss. Essential metabolites that are consumed by ecotype B organisms are colored in red, the other in green. On the top, the evolution of groups A and B proportions is represented, and is colored in green when the A/B interaction persisted, or in red when the interaction failed. Repetition 3 is shaded in grey because the whole population went extinct. In all simulations where A have ceased to produce a metabolite pumped-in by B, B has gone to extinction.

Moreover, the fact that beneficial mutations from ecotype A spread all over the population indicates that competitive exclusion applies in the system. In this case, according to [14] A and B groups cannot be considered as separated ecotypes in the continuous environment. Note that in 4 assays over all the experiments, ecotype B fixed in the population (3 assays for late populations in the continuous environment, 1 assay for early populations in the periodic environment). When ecotype B invades the population, by-products are no more produced by ecotype A, therefore devoting the whole population to extinction.

Hence, it appears that the stability of the A/B cross-feeding interaction in continuous environment depends on its evolution time in the periodic environment since the adaptive diversification. This shows that the co-evolution of ecotypes A and B in the periodic environment reinforced their interaction, meaning that niche specialization stabilized the cross-feeding and fosters a robust negative frequency-dependence. However, even if the cross-feeding interaction seems stable over few thousands of generations, in the continuous environment a beneficial mutation in ecotype A lineage can lead to the extinction of ecotype B lineage, meaning that the stability of the A/B system in the periodic environment does not only depend on their cross-feeding interaction, but also on the seasonality of the environment.

Discussion

Using *in silico* experimental evolution, we have shown that the long-term maintenance of cross-feeding interactions is favored in a seasonal environment, where the environment is reset and primary food is supplied at regular intervals. In this environment, 6 simulations over 12 evolved a stable cross-feeding interaction at the end of the simulations, with two monophyletic ecotypes coexisting via a negative frequency-dependent interaction. At each cycle, Ecotype A grows during the first season, feeding on the primary food and releasing by-products, while ecotype B exclusively feeds on by-products during the second season. The stable coexistence of ecotypes A and B is then based on niche construction, followed by a negative frequency-dependent interaction, as the S and L ecotypes in the LTEE. According to our model, batch culture experiments seem to especially favor the evolution of stable cross-feeding polymorphisms, because the cyclical nature of the environment generates the conditions for the existence of at least two stable seasons: a first season is externally generated by the cyclic mechanism (thus being intrinsically stable) while the second one is generated by the replacement of the exogenously-provided nutrient by the secreted by-products through a mechanism of niche construction.

In the continuous environment, where the primary food is constantly provided (like in a chemostat), cross-feeding interactions emerge, but are not stable because of competitive exclusion. In this case, organisms enrich their environment via their metabolic activity, such that mutants are temporarily able to feed on by-products. But the absence of seasonality precludes any possibility for the stabilization of cross-feeding interactions. Previous wet experiments in chemostat demonstrated the emergence of cross-feeding interactions [7,8,10]. In those experiments, *E. coli* populations have been evolved in a chemostat with glucose as a single limiting resource for at most 1,900 generations. When isolated and evolved together in competition experiments, the different mutants identified to contribute to the cross-feeding interactions reached a stable equilibrium owing to frequency-dependent interactions [8]. Several reasons are invoked to explain why cross-feeding interactions could be stable in chemostat, despite the competitive exclusion principle. According to Pfeiffer and Bonhoeffer [18], cross-feeding may evolve in microbial populations as a consequence of the maximization of ATP production, and the minimization of enzyme concentrations and intermediate products. Those constraints may hinder the emergence of mutants completely degrading glucose (or uptaking glucose and acetate), and outcompeting other cells by competitive exclusion. In our model, organisms do not need to explicitly produce energy carriers. However, competition for resources and division imposes metabolic flux optimization. Based on the same conclusions, Doebeli [34] also suggested that this trade-off between uptake efficiency on the primary and the secondary resources should favor the emergence of cross-feeding polymorphism in chemostat but not in batch culture, because in a chemostat, by-products are more abundant and constantly provided. On the contrary, a more recent theoretical work concluded that, in a continuous and well-mixed environment, the diversity of cross-feeding polymorphism was negatively correlated with primary resource abundance [27].

Our results shed a new light on this question. First, in our model, cross-feeding polymorphisms emerge both in the periodic and the continuous environments. However the stabilization of the cross-feeding interactions is favored in the periodic environment, leading to the evolution of specialized ecotypes. [14] defined an ecotype as an independent monophyletic cluster occupying a specific ecological niche. Ecotypes are at the heart of the bacterial species concept: what makes the genetic cohesion of an asexual bacterial species is periodic selection that regularly purges the genetic diversity in a same ecological niche [14]. As a consequence, ecotypes occupying different niches independently undergo selective sweeps, the mutants from one niche not invading the

other niche. Thus, the stability of a cross-feeding polymorphism should only be analyzed in the light of the robustness of each ecotype against selective sweeps in other ecotypes [14]. This mechanism is observed in the LTEE, as well as in our model. In the periodic environment, ecotypes A and B independently undergo periodic selection events. In the continuous environment, competitive exclusion imposes us to conclude that only one ecotype evolved in this environment even though cross-feeding interactions are continuously observed in the population.

Secondly, when ecotypes A and B evolved in the periodic environment are transferred in the continuous environment, they maintain their negative frequency-dependent interaction for hundreds of generations, until a selective sweep purges the whole population diversity, and destroys the cross-feeding interaction. Moreover, ecotypes A and B that evolved for a long time in the periodic environment have a more robust interaction in continuous conditions, because of niche specialization on the long-term. In the light of those results, we suggest to distinguish between ecological stability and evolutionary stability. Even if different monophyletic clusters, related by cross-feeding interactions, have frequency-dependent interactions, they are not necessarily robust to competitive exclusion on the long-term. In this sense, ecotypes A and B are no longer ecotypes in the continuous environment. On the contrary, in the periodic environment, A and B ecotypes can be considered as proto-species.

Those remarks lead us to hypothesize that the S and L interaction observed in the LTEE, which is still at an early stage, should not be stable in a chemostat on the long-term, even if it could become more and more stable. We also hypothesize that the S/L polymorphism is an ongoing speciation event. On the long run, the S ecotype could even lose the ability to consume glucose.

In a more general view, what we observed is strongly related to known results about temporal niche partitioning in ecology [17]. Bacterial communities commonly undergo adaptive diversification or niche specialization in sympatry, when the environment is seasonal. For example, this mechanism has been observed in marine microbial communities [35], and in lake phytoplankton [36]. In the LTEE [13] and in our model, seasonality for glucose originates from the serial transfer, but the seasonality for acetate is due to cross-feeding and niche construction. Moreover, we demonstrated in our model that negative frequency-dependent cross-feeding is not enough to stabilize the interaction between multiple ecotypes. External factors are necessary, such a regular serial transfer. While the environment is intentionally simplified in those experiments, we can expect much more complex environmental conditions in nature.

Such complex interactions between external factors, emergent cross-feeding interactions and niche construction are therefore of primary importance to understand the evolution of microbial communities in well-mixed environments. Using a computational model of ISEE to decipher those interactions seems to be a rich complementary approach to wet experiments and mathematical modeling.

Conclusion

Using a multi-scale computational model of ISEE, we studied the evolution and stability of cross-feeding interactions in well-mixed environments, providing a single limiting resource periodically or continuously, as in batch cultures or chemostat devices. Our results led us to consider a stable cross-feeding polymorphism as the stable coexistence of different ecotypes, defined as different monophyletic clusters undergoing independent periodic selection events in their own ecological niche [14]. We observed that, even if cross-feeding polymorphism systematically appears in all the simulations, the evolution of stable ecotypes coexisting via cross-feeding is favored in the periodic environment, similarly to the S/L polymorphism observed in the LTEE [11]. In the continuous

environment, competitive exclusion precludes the stabilization of cross-feeding interactions, in apparent contradiction with wet experiments. Indeed, while ecotypes interacting via cross-feeding can temporarily reach an equilibrium, a mutant always eventually outcompetes them. Then, we suggest to study the evolution of cross-feeding polymorphism by fully integrating the notion of ecotype, and distinguishing between ecological stability and evolutionary stability, the latter including long-term evolutionary dynamics such as periodic selection. Our results contributed to understand temporal niche partitioning, by modeling various mechanisms such as cross-feeding, niche construction and seasonality. At a more general scale, our results may contribute to the study of evolution of bacterial communities, by deciphering the conditions of sympatric speciation in asexual populations.

Supporting Information

S1 Table. Common simulation parameters of the entire experimental protocol.

S1 Fig. Evolution of the MRCA age during simulations, for the three types of environments. For each environment, all the repetitions are represented in different colors. **(A)** Periodic environment. **(B)** Random environment. **(C)** Continuous environment.

S1 Video. Variation of the relative fitness of ecotype B during an entire cycle. This video shows the evolution of the ecotype B relative fitness during one cycle, measured at the last of the 10 cycles of the competition experiment. Each of the 333 frames corresponds to one time-step, the whole video presenting the entire cycle.

Acknowledgements

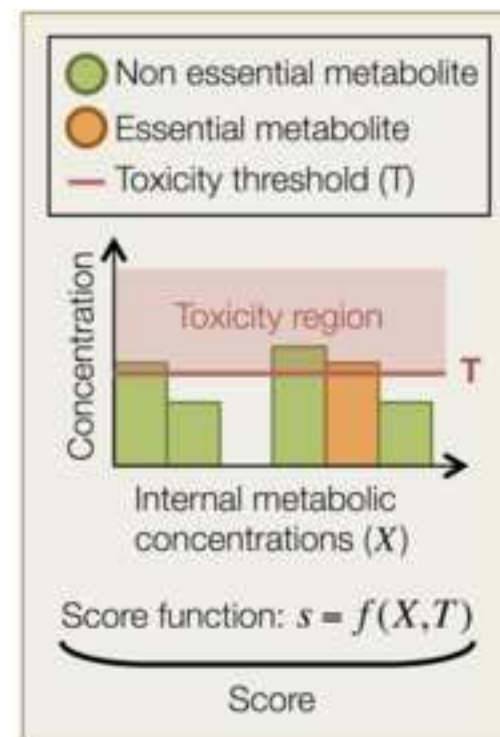
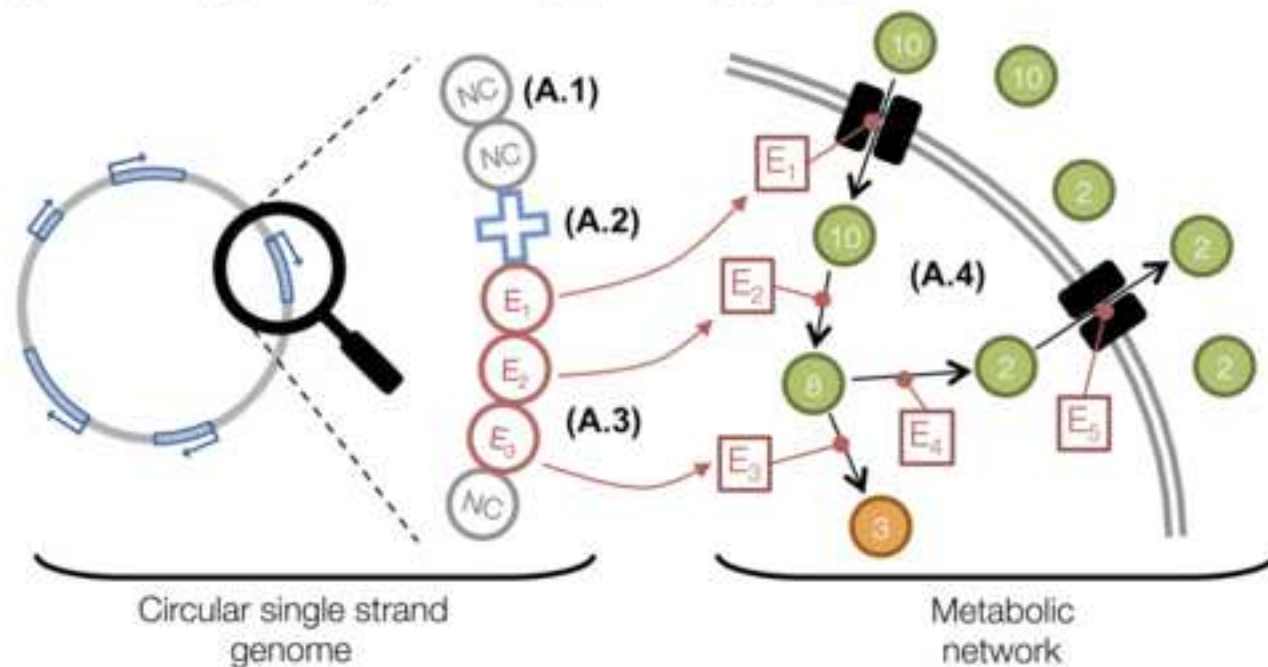
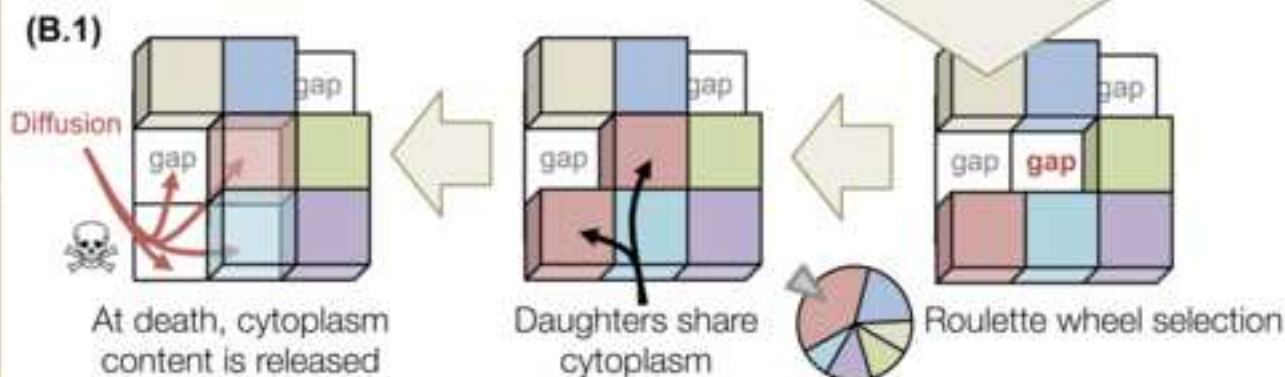
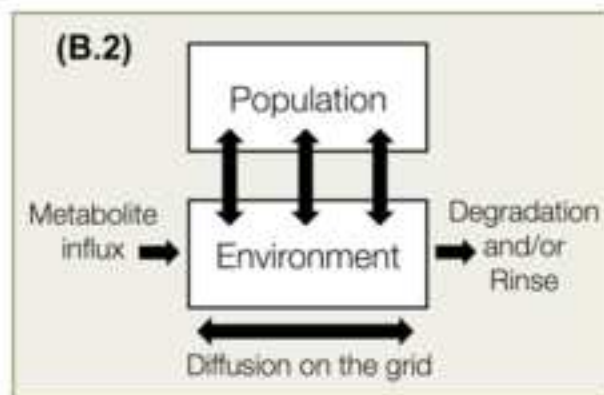
This work was supported by the European Commission 7th Framework Programme (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427). The authors want to thank Noah Ribeck (Michigan State University), Richard Lenski (Michigan State University), and all the partners of the EvoEvo project for fruitful discussions. JC thanks the French Ministry of Research for a PhD fellowship.

References

1. Stams AJ. Metabolic interactions between anaerobic bacteria in methanogenic environments. *Antonie van Leeuwenhoek*. 1994;66(1-3):271–294.
2. Dejonghe W, Berteloot E, Goris J, Boon N, Crul K, Maertens S, et al. Synergistic degradation of linuron by a bacterial consortium and isolation of a single linuron-degrading *Variovorax* strain. *Applied and Environmental Microbiology*. 2003;69(3):1532–1541.
3. Costa E, Pérez J, Kreft JU. Why is metabolic labour divided in nitrification? *Trends in microbiology*. 2006;14(5):213–219.
4. Katsuyama C, Nakaoka S, Takeuchi Y, Tago K, Hayatsu M, Kato K. Complementary cooperation between two syntrophic bacteria in pesticide degradation. *Journal of Theoretical Biology*. 2009;256(4):644–654.

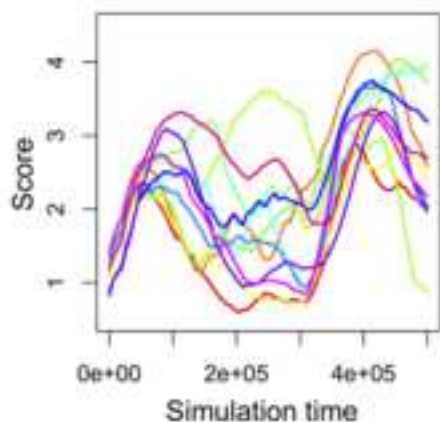
5. Rainey PB, Travisano M. Adaptive radiation in a heterogeneous environment. *Nature*. 1998;394(6688):69–72.
6. Rainey PB, Rainey K. Evolution of cooperation and conflict in experimental bacterial populations. *Nature*. 2003;425(6953):72–74.
7. Helling RB, Vargas CN, Adams J. Evolution of *Escherichia coli* during growth in a constant environment. *Genetics*. 1987;116(3):349–358.
8. Rosenzweig RF, Sharp R, Treves DS, Adams J. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics*. 1994;137(4):903–917.
9. Turner PE, Souza V, Lenski RE. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology*. 1996;77(7):2119–2129.
10. Treves DS, Manning S, Adams J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular biology and evolution*. 1998;15(7):789–797.
11. Rozen DE, Lenski RE. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *The American Naturalist*. 2000;155(1):24–35.
12. Rozen DE, Schneider D, Lenski RE. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *Journal of Molecular Evolution*. 2005;61(2):171–180.
13. Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, Schneider D. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology letters*. 2009;12(1):34–44.
14. Cohan FM. What are bacterial species? *Annual Reviews in Microbiology*. 2002;56(1):457–487.
15. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*. 2003;4(6):457–469.
16. Großkopf T, Consuegra J, Gaffé J, Willison J, Lenski RE, Soyer OS, Schneider D. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC evolutionary biology*. 2016;In press.
17. Spencer CC, Saxer G, Travisano M, Doebeli M. Seasonal resource oscillations maintain diversity in bacterial microcosms. *Evolutionary Ecology Research*. 2007;9(5):775–787.
18. Pfeiffer T, Bonhoeffer S. Evolution of cross-feeding in microbial populations. *The American naturalist*. 2004;163(6):E126–E135.
19. Gerlee P, Lundh T. Rock-Paper-Scissors Dynamics in a Digital Ecology. In: *ALIFE*; 2010. p. 285–294.
20. Hindré T, Knibbe C, Beslon G, Schneider D. New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*. 2012;10(5):352–365.
21. O’Neill B. Digital evolution. *PLoS Biol*. 2003;1(1):e18.

22. Johnson TJ, Wilke CO. Evolution of resource competition between mutually dependent digital organisms. *Artificial Life*. 2004;10(2):145–156.
23. Williams HT, Lenton TM. Evolutionary regime shifts in simulated ecosystems. *Oikos*. 2010;119(12):1887–1899.
24. Crombach A, Hogeweg P. Evolution of resource cycling in ecosystems and individuals. *BMC evolutionary biology*. 2009;9(1):1.
25. Boyle RA, Williams HT, Lenton TM. Natural selection for costly nutrient recycling in simulated microbial metacommunities. *Journal of theoretical biology*. 2012;312:1–12.
26. Chow SS, Wilke CO, Ofria C, Lenski RE, Adami C. Adaptive radiation from resource competition in digital organisms. *Science*. 2004;305(5680):84–86.
27. Gerlee P, Lundh T. Productivity and diversity in a cross-feeding population of artificial organisms. *Evolution*. 2010;64(9):2716–2730.
28. Crombach A, Hogeweg P. Evolution of evolvability in gene regulatory networks. *PLoS Comput Biol*. 2008;4(7):e1000112.
29. Beslon G, Parsons DP, Sanchez-Dehesa Y, Pena JM, Knibbe C. Scaling laws in bacterial genomes: A side-effect of selection of mutational robustness? *Biosystems*. 2010;102(1):32–40.
30. Fischer S, Bernard S, Beslon G, Knibbe C. A model for genome size evolution. *Bulletin of mathematical biology*. 2014;76(9):2249–2291.
31. Weiße AY, Oyarzún DA, Danos V, Swain PS. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*. 2015;112(9):E1038–E1047.
32. Chevin LM. On measuring selection in experimental evolution. *Biology letters*. 2011;7(2):210–213.
33. Ribeck N, Lenski RE. Modeling and quantifying frequency-dependent fitness in microbial populations with cross-feeding interactions. *Evolution*. 2015;69(5):1313–1320.
34. Doebeli M. A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms. *Population Ecology*. 2002;44(2):59–70.
35. Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, et al. Defining seasonal marine microbial community dynamics. *The ISME journal*. 2012;6(2):298–308.
36. Grover JP. Dynamics of competition in a variable environment: experiments with two diatom species. *Ecology*. 1988;69(2):408–417.

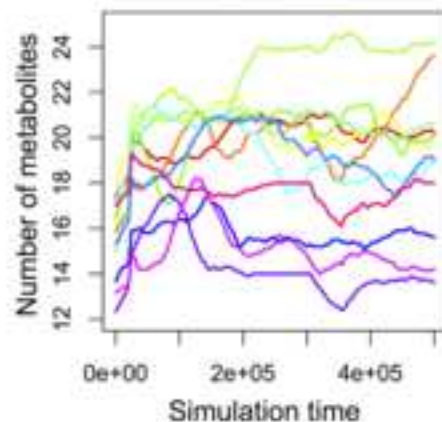
(A) Genotype-to-phenotype mapping**(B) Population-environment level**

(A) Periodic environment

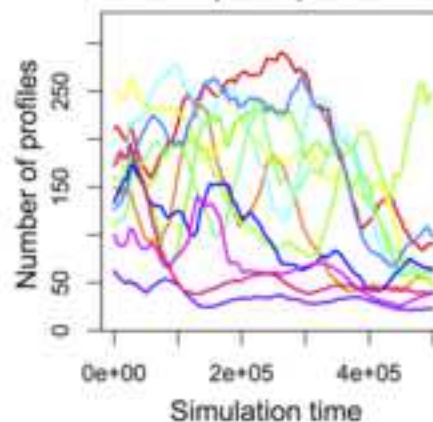
(A.1) Score



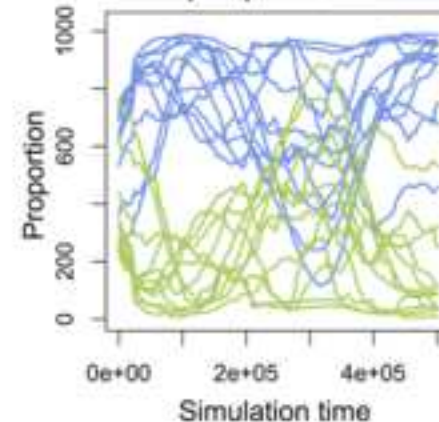
(A.2) Environmental richness



(A.3) Number of trophic profiles

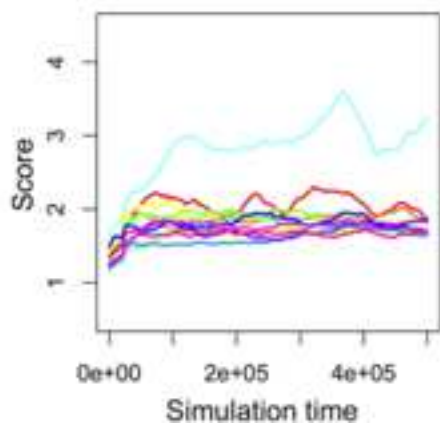


(A.4) Groups A and B proportion

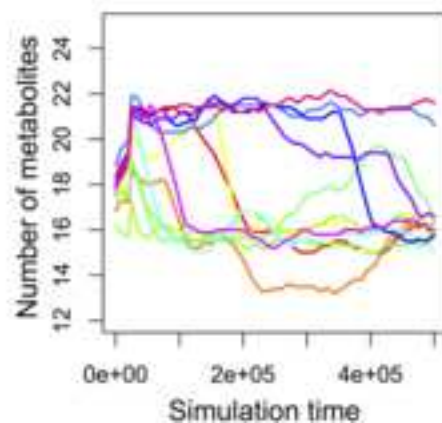


(B) Continuous environment

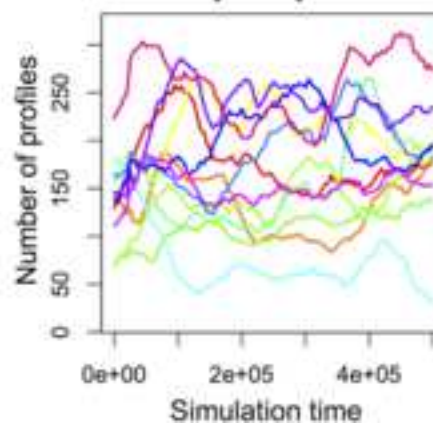
(B.1) Score



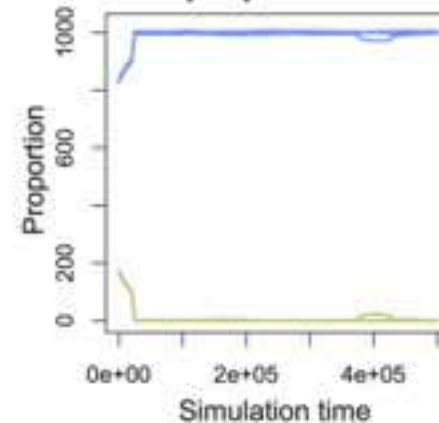
(B.2) Environmental richness

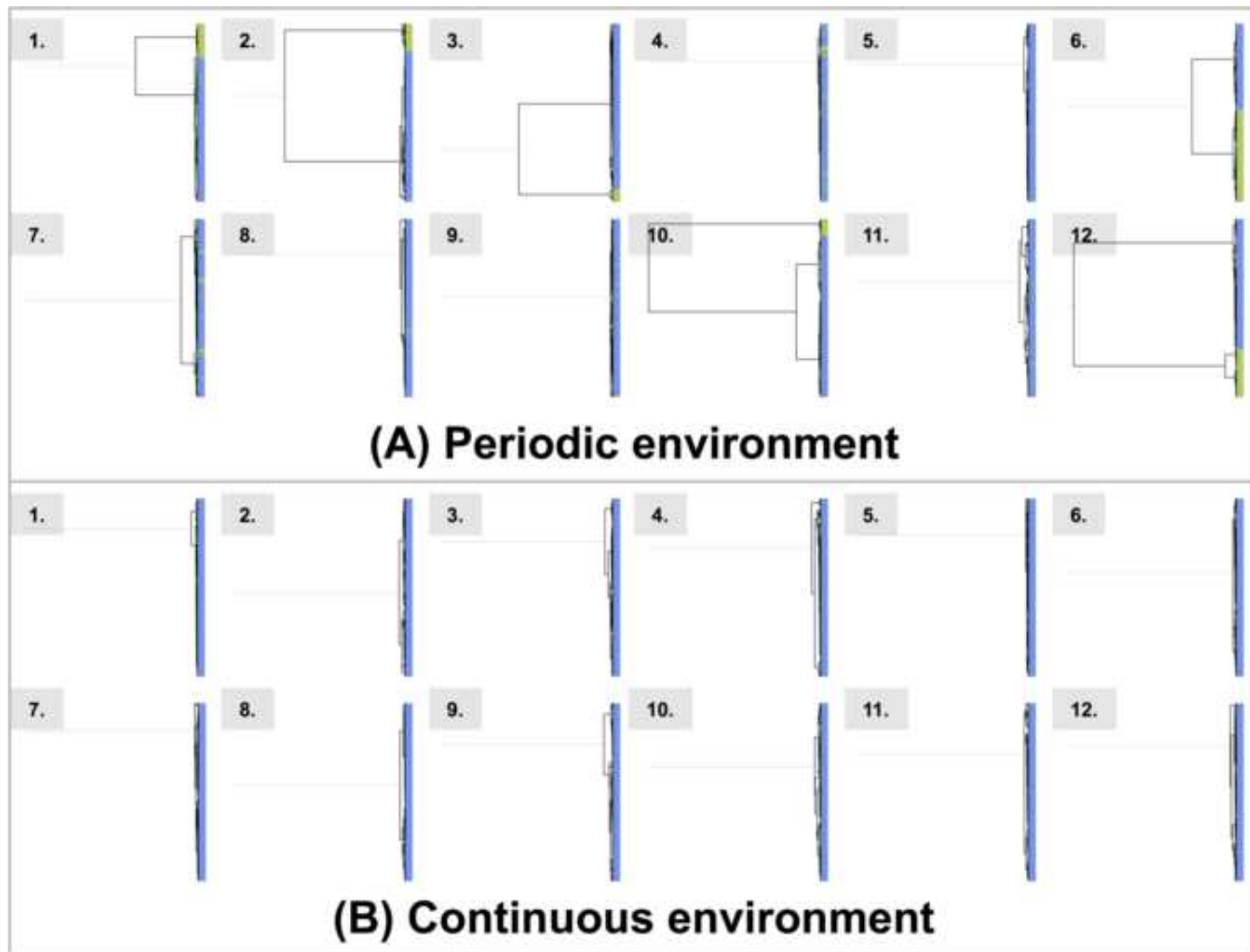


(B.3) Number of trophic profiles



(B.4) Groups A and B proportion

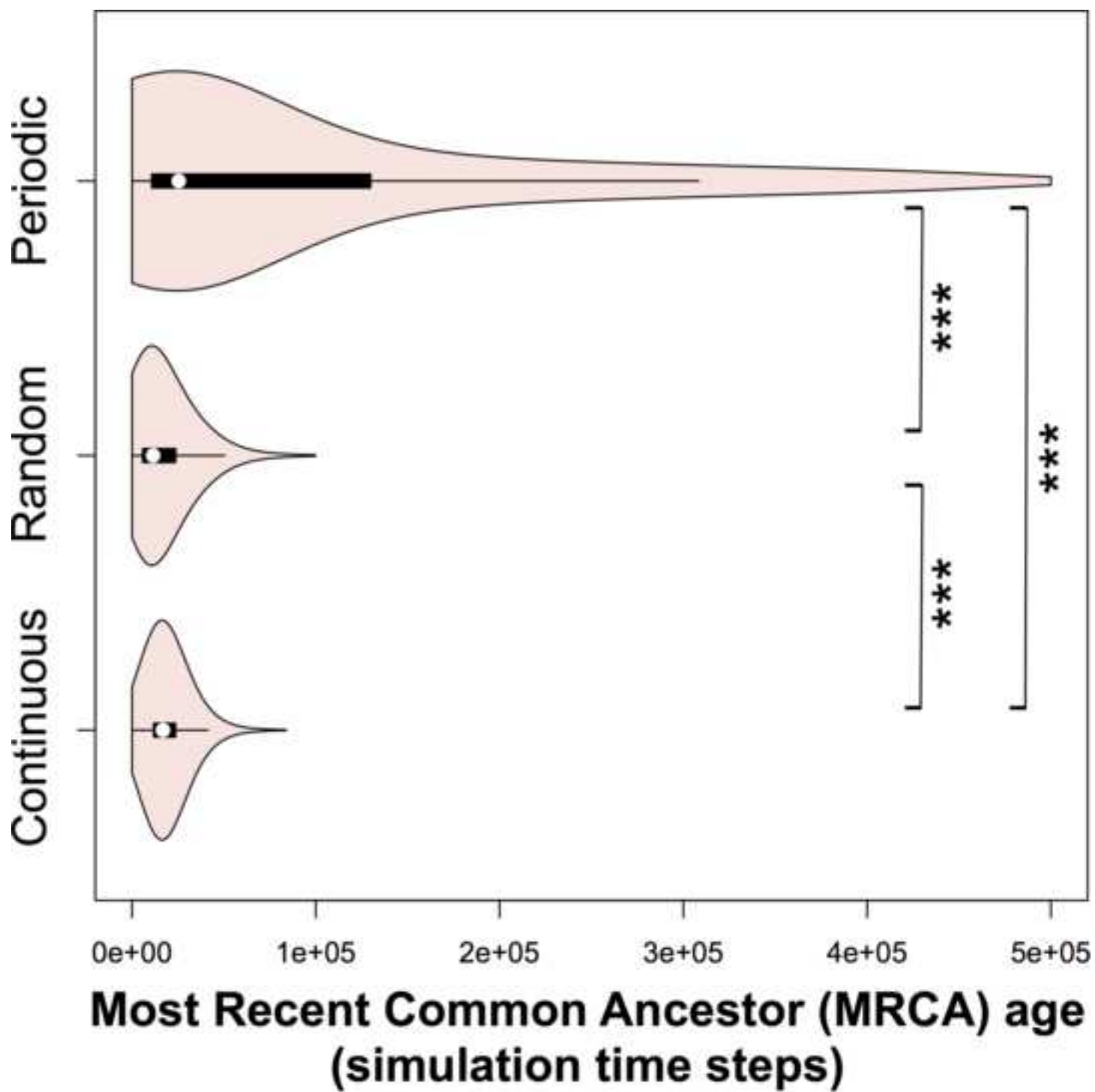


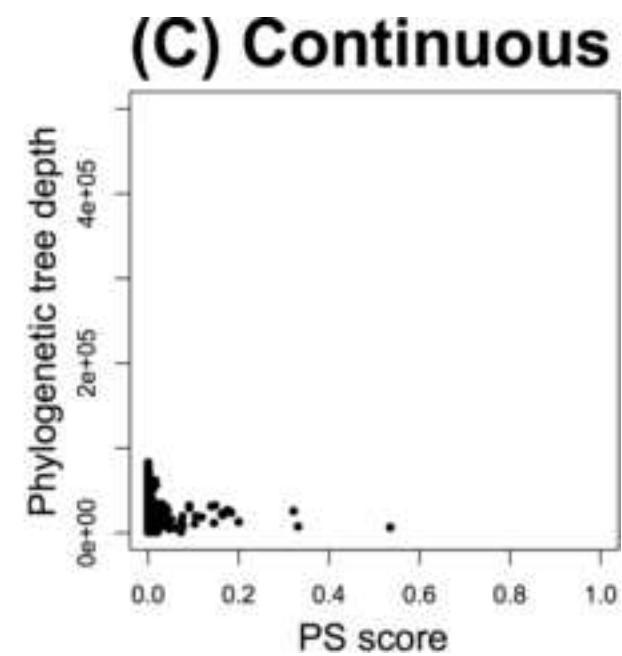
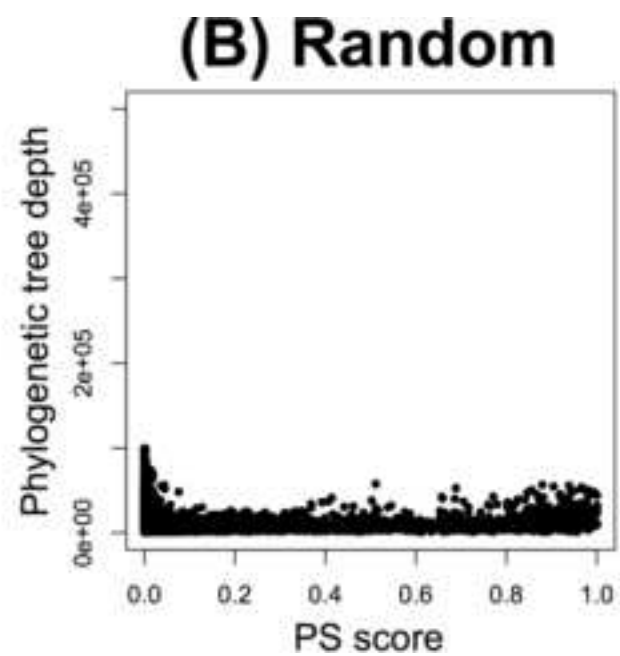
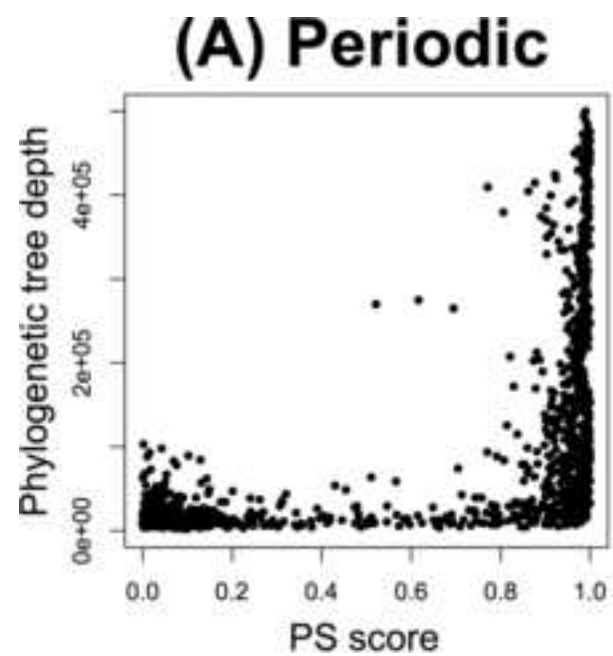


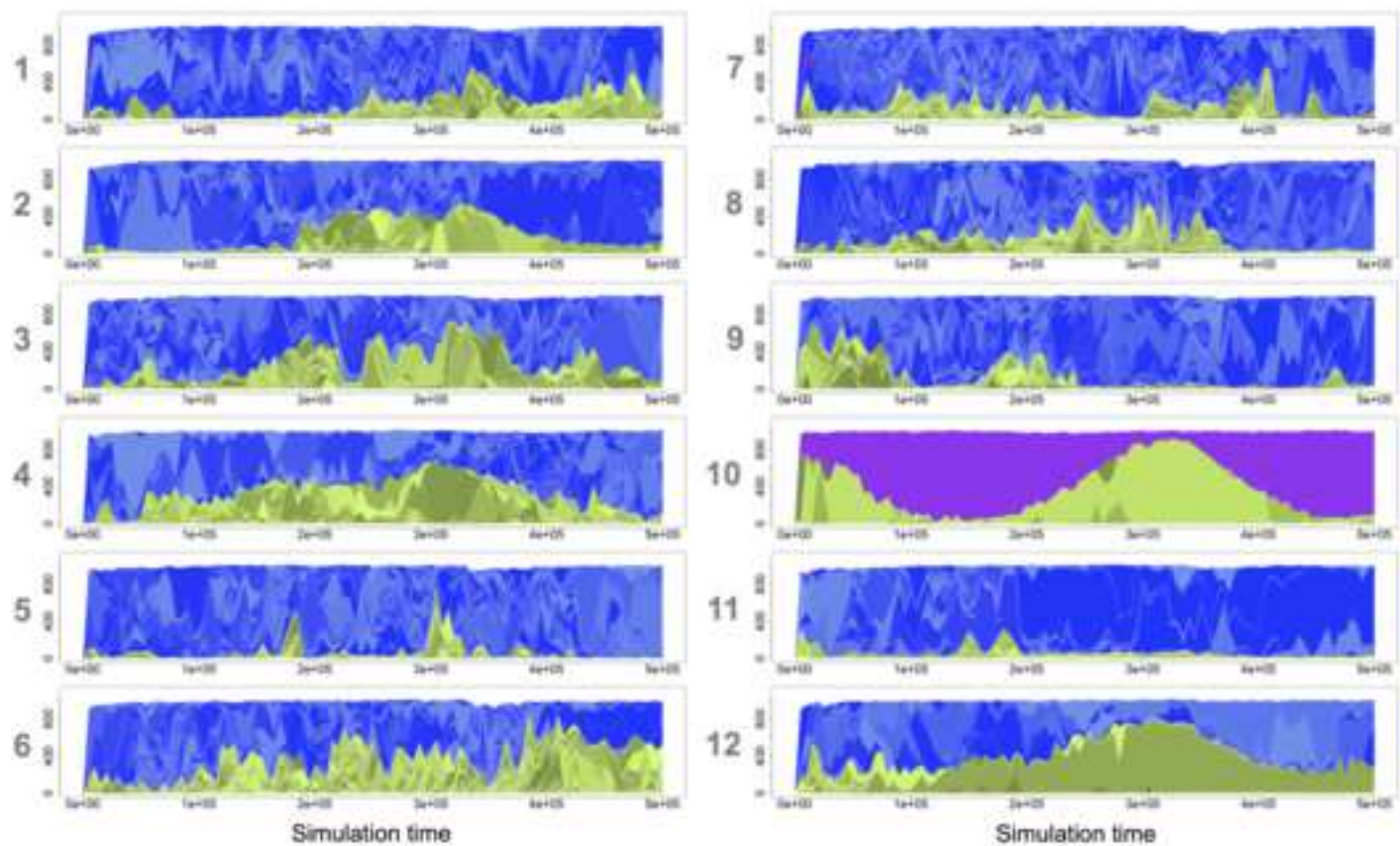
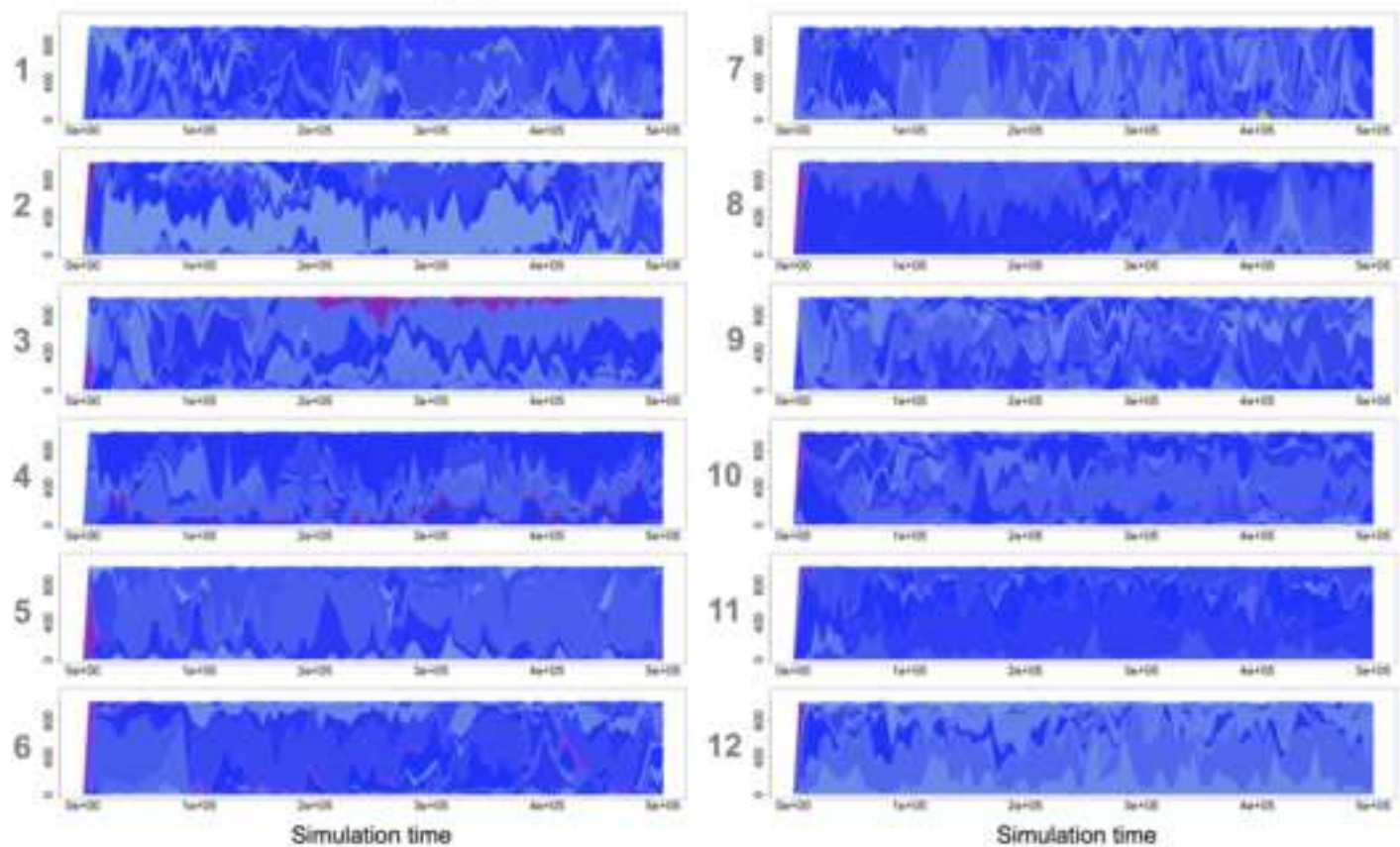
(A) Periodic environment

(B) Continuous environment

■ Group A ■ Group B

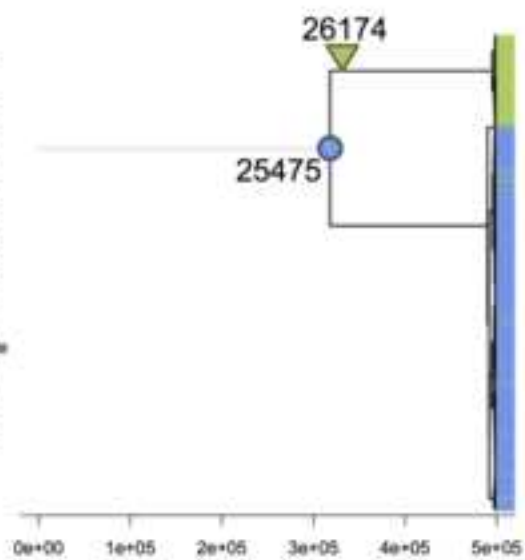




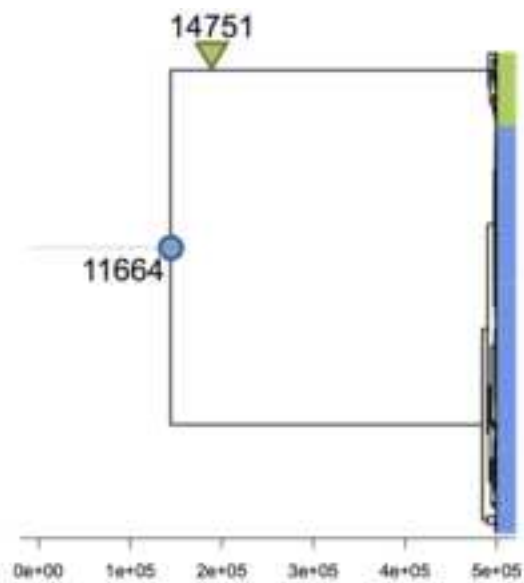
(A) Periodic environment**(B) Continuous environment**

Pure group A profile
 Group A profile
 Group B profile

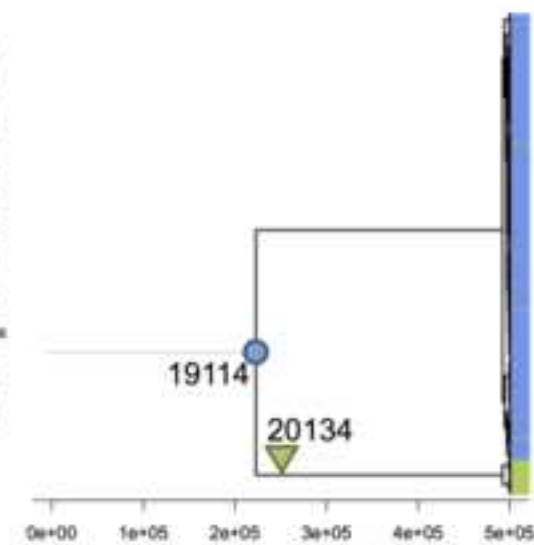
Repetition 1



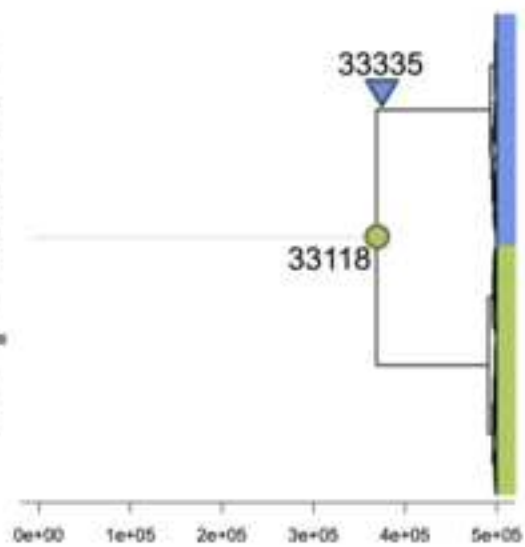
Repetition 2



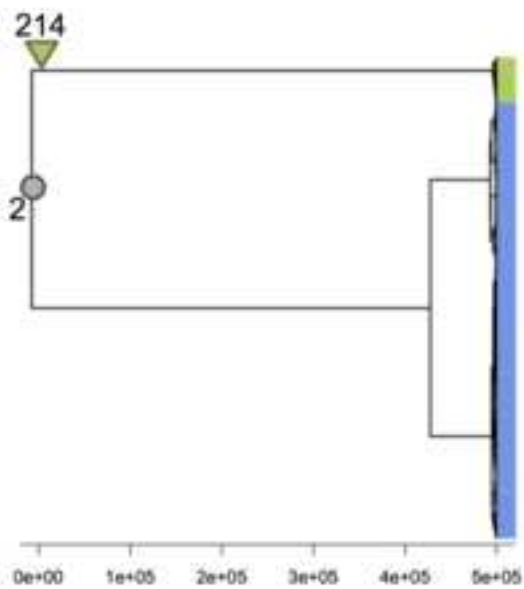
Repetition 3



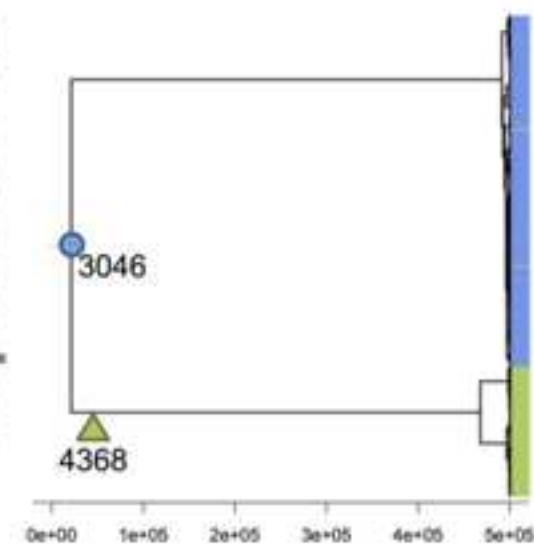
Repetition 6



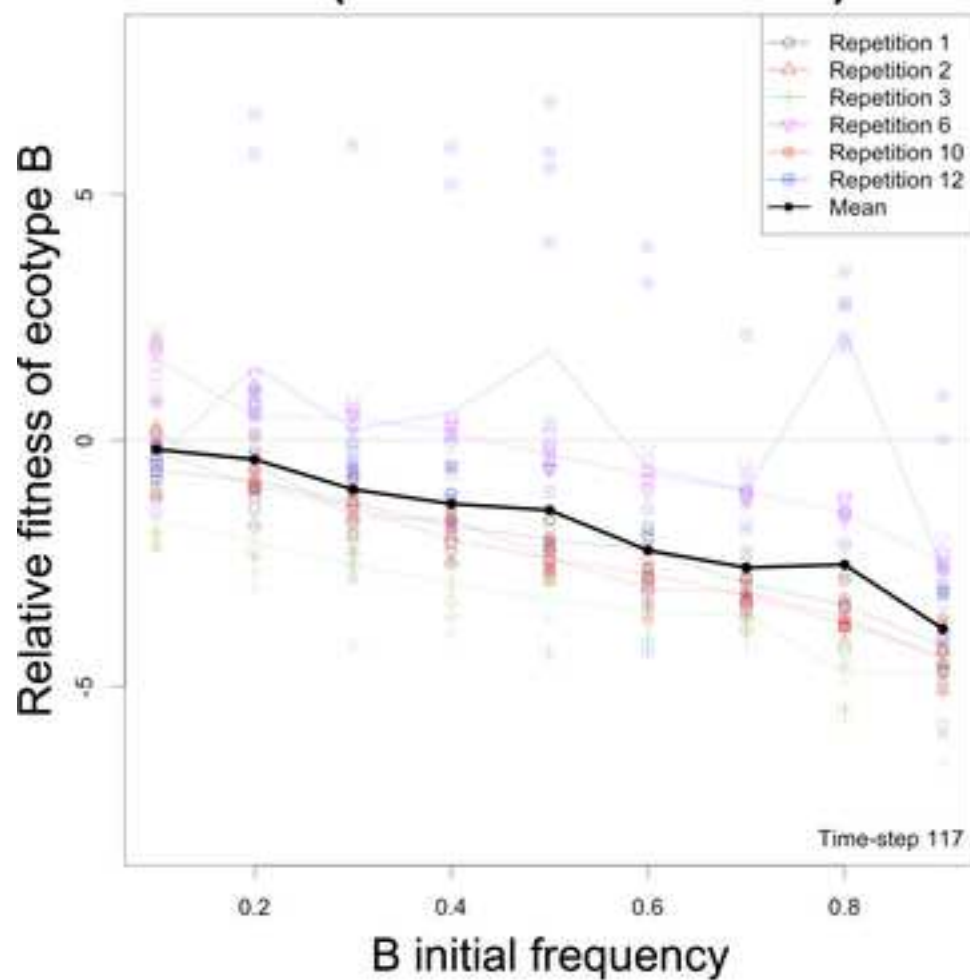
Repetition 10



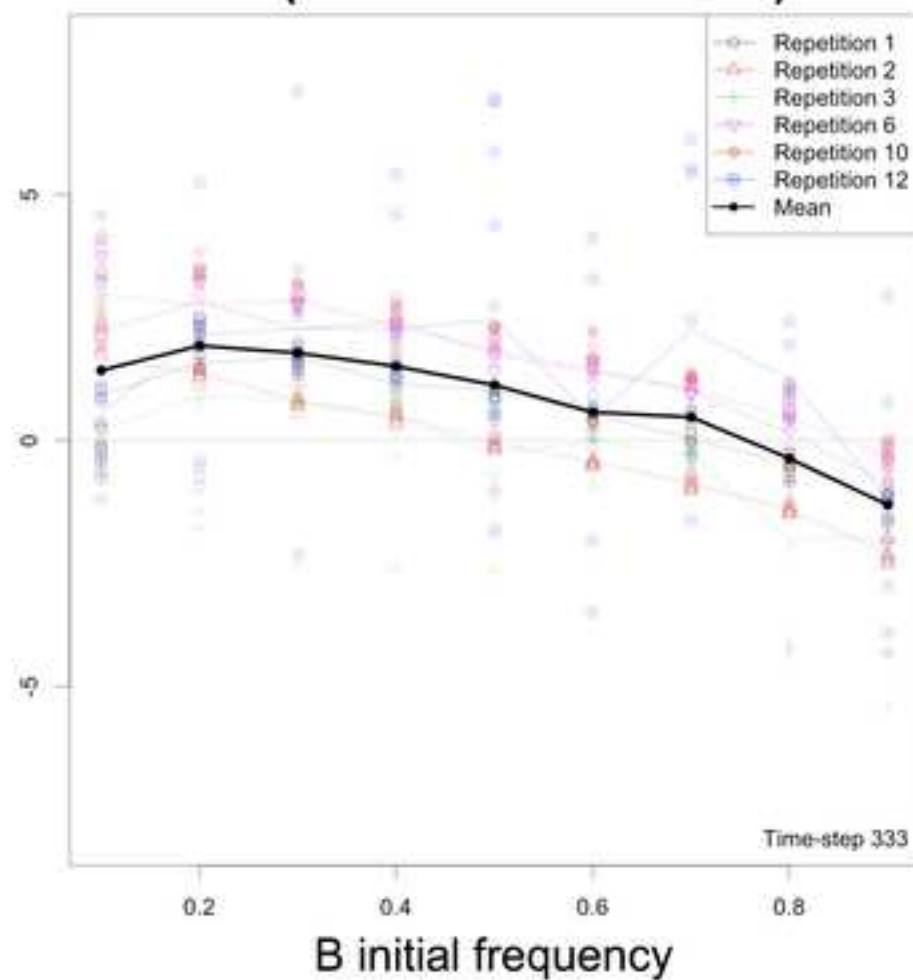
Repetition 12

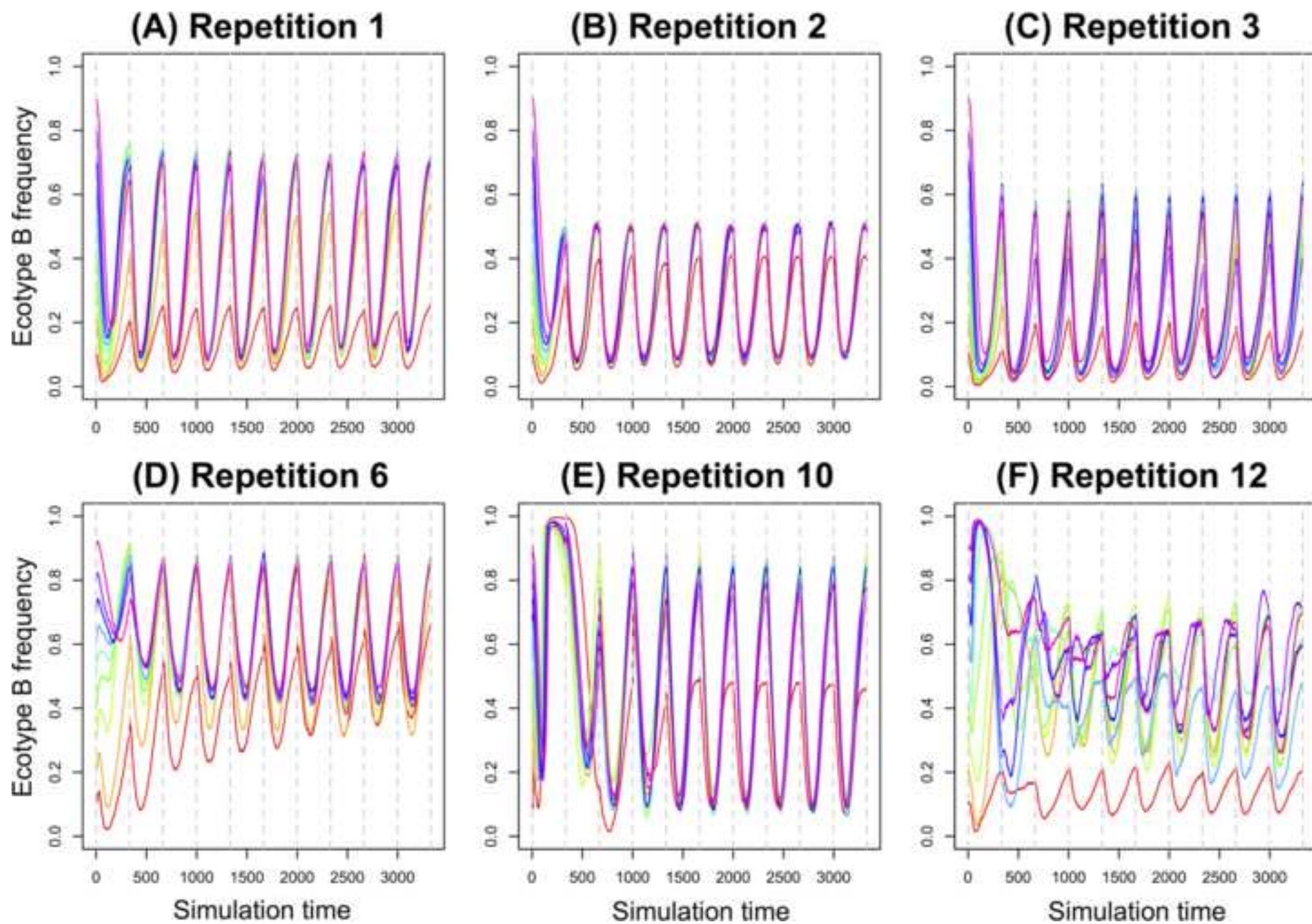


**(A) Worst B relative fitness
(35% of the last season)**

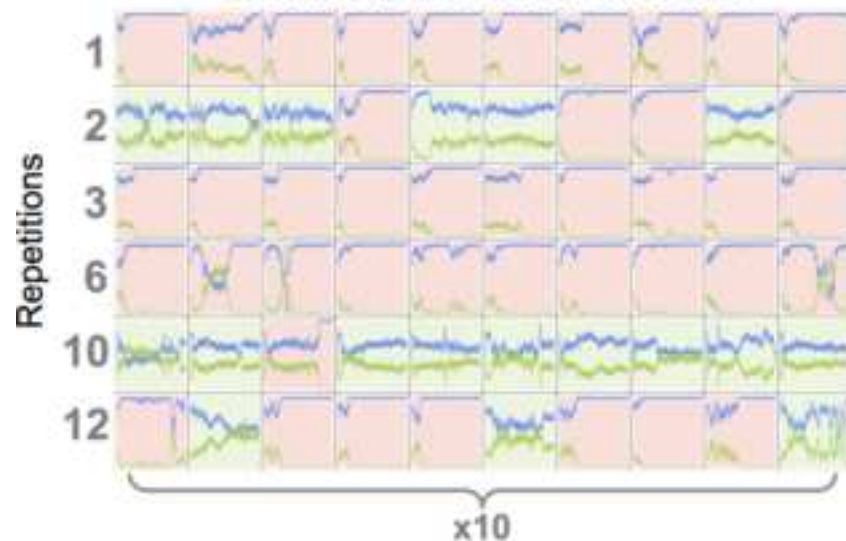


**(B) Best B relative fitness
(end of the last season)**

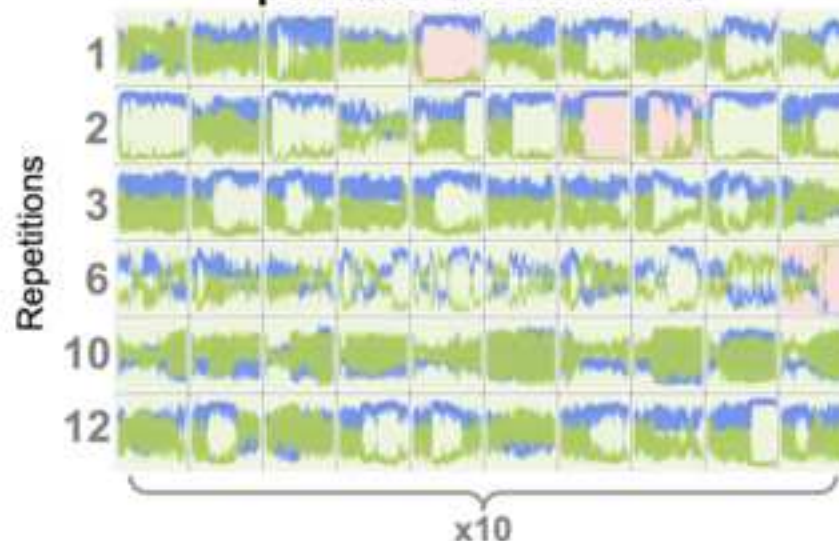




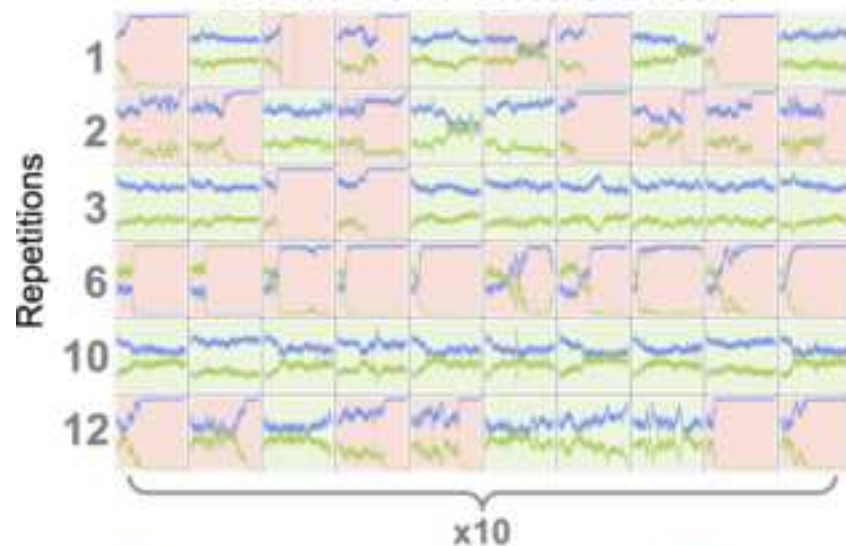
(A) Early populations in the continuous environment



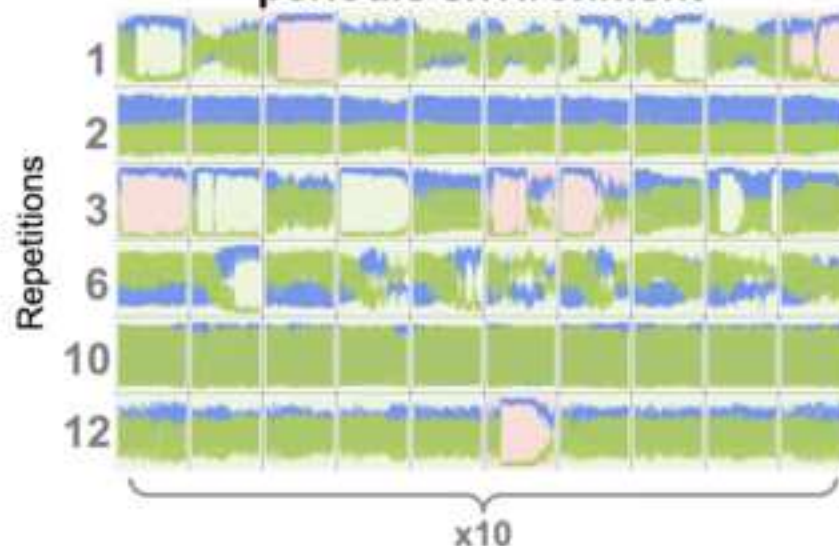
(B) Early populations in the periodic environment

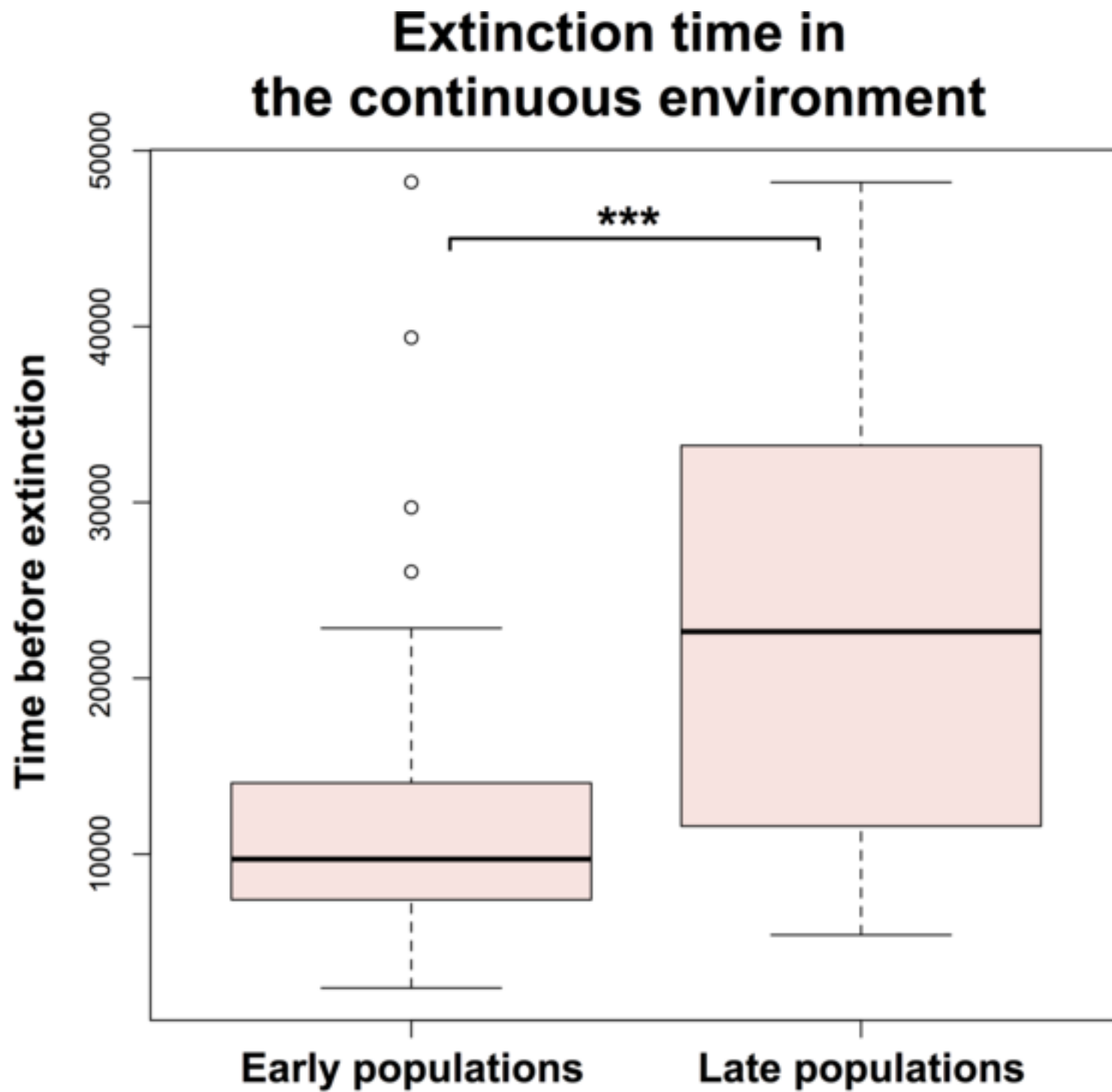


(C) Late populations in the continuous environment

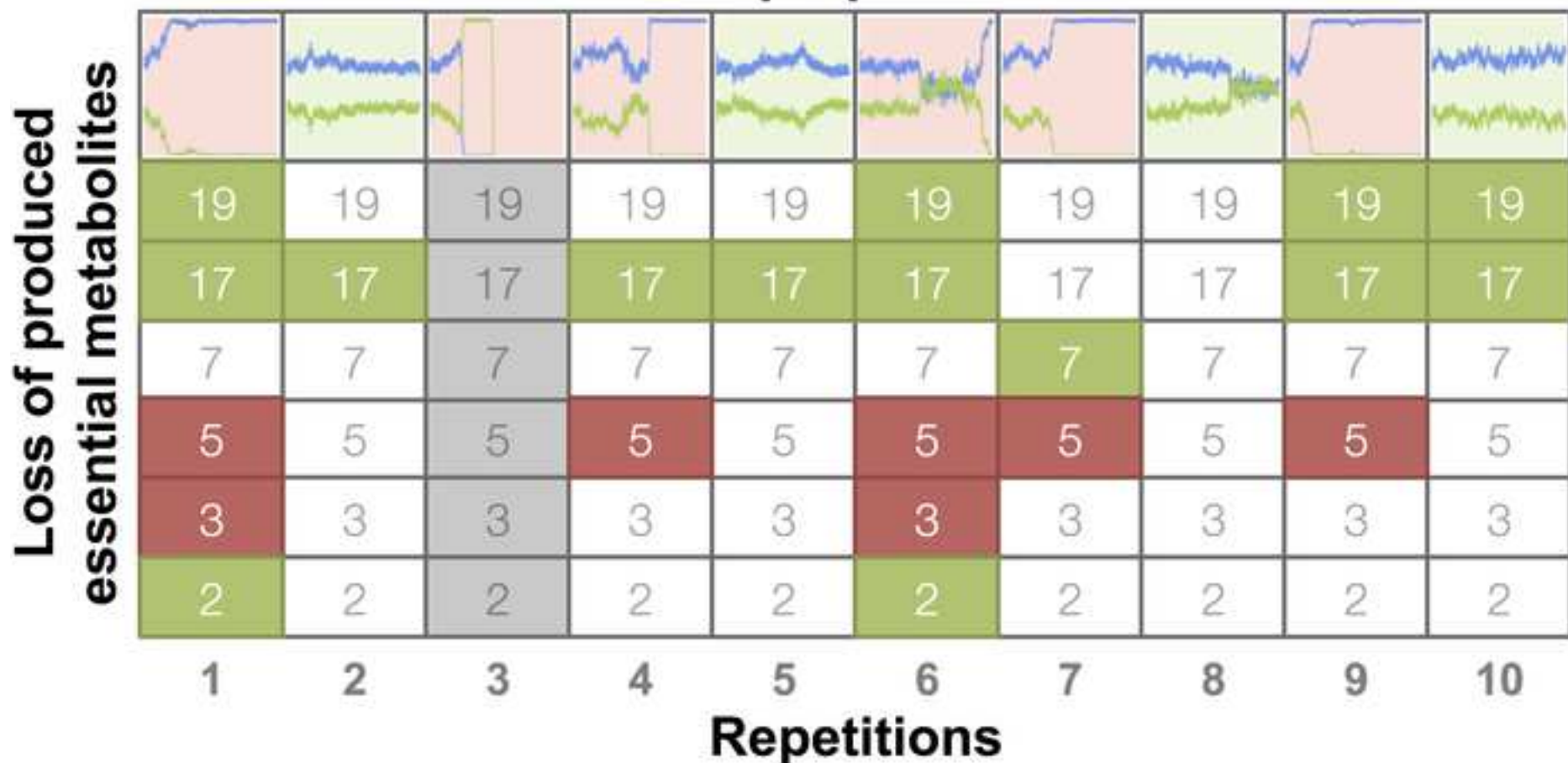


(D) Late populations in the periodic environment





Evolution of A/B proportions



M Essential metabolites lost by ecotype A and not consumed by ecotype B

M Essential metabolites lost by ecotype A and consumed by ecotype B

2 Essential metabolites not lost by ecotype A

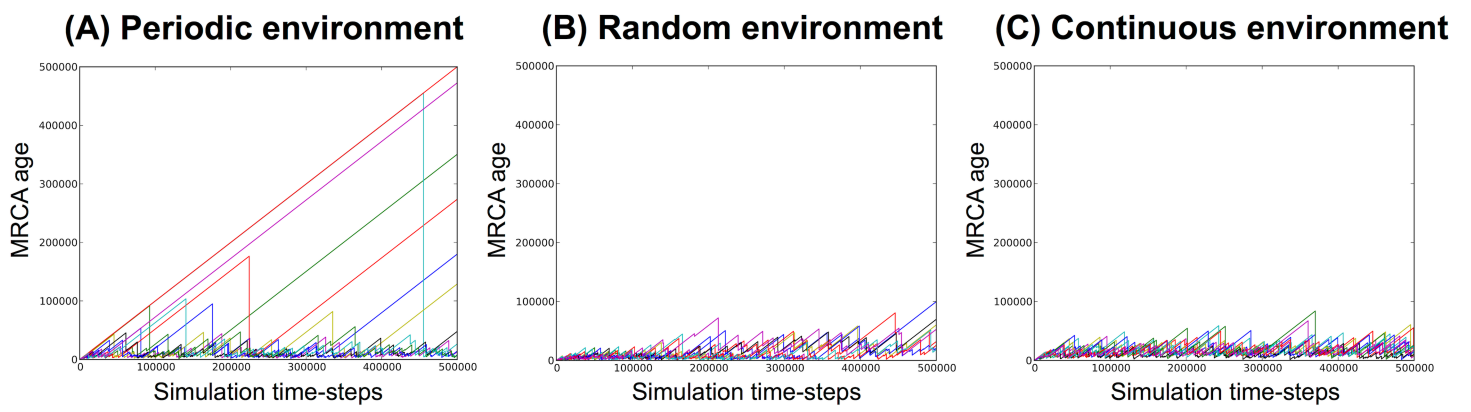
M Whole population extinction

Table S1

Parameter	Symbol	Value	Parameter	Symbol	Value
Simulation time		5e+05	Substrate tag mutation size		1
Grid width	W	32	Product tag mutation size		1
Grid height	H	32	k_m tag mutation size		0.01
Initial number of non-coding units (NC)		10	k_{cat} tag mutation size		0.01
Initial number of promoter units (P)		10	β tag mutation size		0.1
Initial number of enzyme units (E)		10	Maximum genome size		10000
Point mutation rate		1e-03	Non essential metabolites toxicity threshold		1
Duplication rate		1e-03	Essential metabolites toxicity threshold		1
Deletion rate		1e-03	Death probability	p_{death}	0.02
Translocation rate		1e-03	Primary food	m_{exo}	10
Inversion rate		1e-03	Diffusion rate	D	0.1
Transition rate		1e-03			

Those parameters are common to all the simulations of the experimental protocol.

Fig. S1



**PART III: GENERAL
DISCUSSION**

Understanding the ecological, physiological and molecular bases of emergence and long-term maintenance of bacterial polymorphisms, especially in sympatric conditions, is an important challenge. From a fundamental point of view, this may give us valuable information about the origins of life diversity. From a practical point of view, it may be useful to implement new strategies to limit the emergence of virulence, antibiotic resistance for public health or inefficient strains in biotechnology.

Ecological theories of adaptive diversification (see Introduction, *section I.i*), have been developed to explain how environmental conditions affect organismal evolution and production of diversity. Most of these theories are built on the knowledge acquired from comparisons of contemporary organisms (as the example of adaptive radiation of Darwin finches, (Grant and Grant 2006) or between already adapted organisms and their ancestral fossil records. However, these theories are not amenable to direct experiments owing to the lack of both “living” ancestors and complete fossil records that would allow for direct fitness comparisons. Moreover, adaptive events occurred in a distant past, and we therefore see their consequences but not their causes. Experimental evolution strategies (*in vivo*, *in vitro*, *in silico*) overcome these limitations and are therefore essential tools to test the classical theories of adaptation.

Studies of evolutionary processes have been improved during the last decades owing to the implementation of experimental evolution strategies. They allow both the control of the selective pressures acting on natural selection and the accessibility of the ancestors and all evolved living forms, providing with the opportunity to rigorously analyze all intermediate steps from selective pressures and evolutionary outcomes.

With the development of new technologies in genetics and molecular biology, many studies have screened large microbial communities to establish their phylogenetic relationships.

Combined analyses of the structure (whole-genome sequencing, optical maps) and expression (global transcription profiles, proteomics, ChIP-Seq,...) provide detailed insights about the evolutionary changes and origins of adaptation at the regulatory level. Finally, development of computer models (Flux Balance Analyses, *in silico* evolution models,...) allow predictions to be made and testable by experiments. All these approaches can be combined with experimental evolution allowing for the study of the dynamics of evolutionary processes.

My PhD work consisted in the study of the physiological, genetic and ecological parameters driving adaptive diversification. The initial studies of the stable and dynamic polymorphism that occurs in the Ara-2 population were mostly based on the phenotypic characterization of the two S and L lineages and their interactions (Rozen and Lenski 2000), their phylogenetic origin (Rozen, Schneider, and Lenski 2005), and their ecological drivers (Rozen and Lenski 2000, Rozen et al. 2009). Later studies focused on the ecology of the long-term maintenance of the polymorphism (Le Gac et al. 2012), and the genetic bases of its emergence (Plucain et al. 2014). Despite the large amount of information obtained in the last years, several aspects remain unclear, including the physiological mechanisms of the emergence, and the role of several ecological parameters that have been postulated to be important in the emergence and long-term maintenance of diversity such as: tradeoffs, character displacement and seasonality. Moreover, the impact of these physiological and ecological parameters on shaping bacterial cells at the genetic and molecular levels is mostly unknown.

Mechanistic details of niche construction in Ara-2

Initial studies about the Ara-2 population demonstrated that the emergence of the polymorphism involves niche construction interactions (Rozen and Lenski 2000). However, the physiological drivers were poorly understood and the metabolite secreted by evolved clones from the L lineage that was exploited by the S lineage was unknown. During this work, we identified acetate as the main component of the new L-constructed niche. Moreover, we developed a modeling framework, called *evoFBA* that enabled us to hypothesize specific ecological and physiological mechanisms that produce and sustain this polymorphism (See Results, Chapter 1; (Großkopf et al. 2016)). These model predictions were confirmed experimentally allowing us to draw the general picture of the physiological/metabolic and ecological bases of S and L emergence and coexistence.

The S and L lineages have distinct metabolic flux patterns that promote a cross-feeding interaction. We showed that acetate secretion was the primary metabolic driver leading to the emergence of the polymorphism. Moreover, during the 50,000 generations of their coexistence, the S and L lineages evolved metabolic tradeoffs between glucose and acetate consumption. These tradeoffs prevent the simultaneous optimization of growth on both carbon sources and thus the two lineages became niche specialists with the L lineage specializing on glucose and the S lineage on acetate (See Results, Chapter 1; (Großkopf et al. 2016)).

The emergence and maintenance of the Ara-2 polymorphism were also predicted by the implementation of a second *in silico* evolutionary model (See Results, Chapter 5). In this theoretical work, a multiscale framework using digital organisms, mimicking bacterial cells, with their own genome encoding a regulatory and a metabolic network was designed by the team of Guillaume Beslon (INRIA, Beagle team, Lyon, (Rocabert et al. 2016)). The simulations predicted the emergence and long-term coexistence of two ecotypes based on niche

construction and negative frequency-dependent interactions, the same kind of polymorphism that characterizes the S and L lineages. This diversification event was only possible in seasonal conditions, provided in batch culture experiments as in the LTEE. These results support the essential role of seasonality in temporal niche partitioning and are in agreement with the experimental results obtained in the Ara-2 population (Rozen et al. 2009).

Seasonality, and thus changes in the relative abundances of carbon sources, stimulates competition between nascent ecotypes producing new selective pressures, and therefore new ecological niches, for divergence. These processes are associated with shifts in characters, including tradeoffs in traits as metabolite uptake rates and use. This phenomenon, also called ecological character displacement is postulated to be a key driver of adaptive diversification and specialization (Stuart and Losos 2013). We studied the effect of competition on character displacement and the emergence of the Ara-2 polymorphism (See Results, Chapter 2). We eliminated competition by means of a short 300-generation allopatric evolution which showed that, without competition, the two sympatric divergent lineages evolve in a convergent way. This resulted in the loss of frequency-dependent interactions between the two S and L lineages. This loss was likely caused by a phenotypic convergence in resource use with simultaneous invasion of both ecological niches by either lineage, thereby eliminating the ecological character displacement. These results support the essential role of competition in both character displacement and tradeoffs of the L and S lineages, thereby highlighting its role as a key driver of adaptive diversification and specialization.

Genetics of the polymorphism in Ara-2

Specialization of the L lineage on glucose allowed the evolution of higher glucose uptake rates, thereby resulting in acetate overflow and construction of a new niche exploited by the S lineage. The S improvement on acetate is related to increased promoter activities of *acnB* and *aceB* (both encoding key enzymes of the TCA cycle, responsible for acetate consumption), which is consistent with higher fluxes through the TCA cycle and glyoxylate shunt in this lineage (See Results, Chapter 1). Moreover, we showed that mutations affecting both *acs* and *arcA* were essential for the S lineage to evolve a higher capacity to consume acetate (See Results, Chapter 3). The *acs* mutation resulted in an upregulation of *acs* that encodes acetyl-CoA synthase, which is involved in acetate consumption by producing acetyl-CoA from acetate. The *arcA* mutation increased *acnB* and *aceB* transcription in the S lineage, which we had shown previously to be related with greater fluxes through the TCA cycle and glyoxylate shunt in this lineage (See Results, Chapter 1).

Ecological and physiological drivers of adaptive diversification shape bacteria at the genetic and molecular levels. In Ara-2, the physiological changes of adaptation to glucose consumption in the L lineage produced a new ecological niche in the sympatric environment of the LTEE. This new ecological opportunity promoted the emergence of the S lineage, which was previously demonstrated to be associated with three specific mutations that affected the global regulatory genes *spoT*, *arcA* and *gntR* (Plucaín et al. 2014). Here, we showed how the *arcA* mutation shapes bacteria at the molecular level and produces regulatory changes that allow for the exploitation of the new ecological opportunity (See Results, Chapter 3). The *arcA* mutation resulted in the L93F change in the evolved ArcA protein, and we showed that this change altered its DNA-binding activity. This resulted in higher transcription of genes involved in acetate consumption including *acs*, *acnB* and *aceB*. The evolved ArcA protein thereby

promotes a restructuring of the metabolic network that resulted in the later emergence of the S lineage.

The emergence of long-term polymorphism by niche construction in the LTEE was unexpected (See Introduction, Section II.iii.b). Consistently, adaptive diversification by niche construction has been described only in the Ara-2 population of the LTEE. Even if other LTEE populations show evidence of negative frequency-dependent fitness, this diversity is however not maintained in the long-term owing to clonal interference (Elena and Lenski 1997, Maddamsetti, Lenski, and Barrick 2015). One possible explanation is that the establishment of the S lineage relies on epistatic interactions between multiple mutations, which may have limited its evolutionary accessibility (Plucain et al. 2014). We confirmed and identified even more epistatic interactions during this work. Indeed, epistatic interactions resulted in temporal successive mutational events that were important at certain timepoints during evolution for the establishment of the S lineage. Moreover, allele specificity has been shown to be important in S-specific phenotypic traits. Indeed, only the S-evolved *arcA* allele, and not evolved *arcA* alleles from other populations of the LTEE, confers the S phenotype (Plucain et al. 2014). This may also explain why other populations from the LTEE have been reported to improve growth rates in acetate relative to the ancestor at 20,000 and 50,000 generations (Leiby and Marx 2014) without any adaptive diversification and long-term coexistence of different lineages.

The evolutionary accessibility of the mutations responsible for the S lineage emergence and specific phenotypic traits may have been facilitated by the mutator state of this population (See Results, Chapter 4). Indeed, bacterial evolvability in response of ecological challenges requires mutations of DNA sequences as source of long-term evolutionary adaptation. The Ara-2 population evolved an early mutator phenotype that emerged at ~2000 generations and was completely fixed by 3500 generations (Shaver et al. 2002). The adaptive diversification event in this population occurred after the appearance of the mutator genotype, and thus the two L

and S lineages share the mutator allele and the 10-fold increased mutation rate compared to the ancestor (See Results, Chapter 4). This raises the question about the role of the mutator state in the emergence of the polymorphism.

The high mutation rates were reverted late during evolution (See Results, Chapter 4). Indeed, the S lineage reverted first between 30,000 and 40,000 generations and then the L lineage between 40,000 and 50,000 generations. Theoretical studies predict that fitness of populations with high mutation rates increases faster during the process of adaptation and that the mutator phenotype will decline once the adaptive peak is reached (Taddei et al. 1997). Adaptation in Ara-2 is much more intricate than in other mutator populations owing to the selective pressures exerted by the two lineages. Indeed, adaptation to their preferred niches, *i.e.* glucose for the L lineage and acetate for the S lineage, may take much longer than the adaptation to glucose in the other populations. This may explain the late reversion of the mutator phenotype. We compared the dynamics of the growth rates of evolved L and S clones in glucose or acetate media (see Results, Chapter 1; (Großkopf et al. 2016)) to the dynamics of their mutation rates. We observed that growth rates of S clones in glucose stabilize over time and in acetate is very high at 40,000 generations when the S mutator state reverted. Moreover, at the time of mutator reversion of the L lineage, its growth rates in glucose and acetate are at their maximum and minimal values, respectively.

Comparison with other evolution experiments

The emergence of adaptive diversification is expected, and has occurred, in other evolution experiments (See Introduction, section I.iv), for instance when two exogenous carbon sources are provided (Friesen et al. 2004, Tyerman et al. 2008), and in chemostats in media with high amounts of glucose in which substantial doses of acetate are produced (Vemuri et al. 2006, Rosenzweig et al. 1994). The adaptive diversification observed in the LTEE was however unexpected owing to the presence of glucose as a single carbon source which was moreover supplied at a low concentration. This was thought to avoid acetate overflow (See Introduction, Section II.iii.b). Indeed, the acetate concentrations differ significantly in the three evolutionary systems from 50 μM in our case (see Results, Chapter 1; (Großkopf et al. 2016)), to 3.4 mM (Friesen et al. 2004) and ~6 mM in the chemostat study (Rosenzweig et al. 1994). The much lower acetate concentrations detected in the LTEE might explain that the acetate ecological niche was more difficult and longer to access. Thereby, the S phenotype needed more time to emerge and invade this niche, thereby reaching the highest adaptive peak in acetate (See Results, Chapter 1). This may also explain the “need” for multiple mutations (See Results, Chapter 3) which obviously were more accessible owing to the higher supply of mutations provided by the high mutation rates (See Results, Chapter 4). It is important to emphasize here that the number of mutations that resulted in the emergence and establishment of the S lineage was high. Indeed, it was previously shown that 3 mutations in *spoT*, *arcA* and *gntR* were involved (Plucain et al. 2014). We showed here that a fourth mutation, in *acs*, was also needed and that additional mutations interacted epistatically with these mutations, indicating that the establishment of such polymorphism is a complex evolutionary outcome.

Among all evolution experiments, only the LTEE has been carried out for long enough to investigate the long-term coexistence of polymorphisms. Other evolution experiments (Rosenzweig et al. 1994, Friesen et al. 2004) were not carried out for more than ~1500

generations so the long-term fate of polymorphisms is therefore unknown. Moreover, our *in silico* evolution experiments predict that in continuous environments, where the primary food is constantly provided like in a chemostat, cross-feeding interactions emerge but are unstable because of competitive exclusion (See Results, Chapter 5). In this case, organisms increase the diversity of their environment through their metabolic activity. New ecological opportunities emerge and mutants are temporarily able to feed on by-products. However, the absence of seasonality prevents any possibility for the stabilization of cross-feeding interactions at the long-term.

Due to their asexual nature, bacterial species are defined as ecotypes according to Cohan's "Ecological Species Concept" (See Introduction, section I.i). Cohan defined an ecotype as an independent monophyletic cluster occupying a specific ecological niche (Cohan 2001). The S/L cross-feeding polymorphism can be analyzed in the light of this definition. In the seasonal environment, S and L lineages undergo periodic and independent selection events. This has been proven by phylogenetic analyses that showed the monophyletic origins of both lineages (Rozen, Schneider, and Lenski 2005). Moreover, S and L coexist owing to negative frequency-dependent interactions for tens of thousands of generations. This coexistence is related to the evolution of metabolic tradeoffs where each lineage specializes in a specific ecological niche. Taking all these results together, we hypothesize that the S and L polymorphism is an ongoing speciation event.

Future work

At the molecular level, the *gntR* mutation may be analyzed. First, as for ArcA, purification of the ancestral and evolved GntR proteins and assays of their DNA-binding affinities to known target promoters may be performed. Second, specificity of the Ara-2 allele in the S lineage emergence can be analyzed. Hence, *gntR* alleles from other populations may be tested to know whether they can also be involved in the S emergence. At a more global level, we determined the global transcription profiles of several strains: the ancestor, the 6.5KS1 clone and derived constructed strains with and without the evolved S alleles of the *spoT*, *arcA* and *gntR* genes, and also with and without alleles from other populations. Analyses of these profiles will help us to understand the effects of these mutations on the emergence of the S lineage, for example by identifying the new regulons of these altered global regulators.

In order to understand the complete landscape of the diversification event in the Ara-2 population, future work may also focus on the metabolic restructuring underlying the tradeoffs that produced the S and L polymorphism. Since ArcA and GntR are regulators of genes encoding enzymes of central metabolism as the Krebs cycle and the Entner-Doudoroff pathway, the measurement of the activities of relevant metabolic enzymes in the ancestor and S and L clones together with metabolic flux analysis from ¹³C-glucose experiments will allow to identify how central metabolism is rewired. All these data will then be analyzed to understand how the *arcA* and *gntR* mutations conferred the phenotypes that are specific to the S lineage. These approaches on the L lineage will help to to understand the physiologic/metabolic changes in glucose metabolism that allowed acetate overflow and then niche construction.

Finally, the role of mutation rates in the emergence of adaptive diversification events can be studied both *in silico* and *in vivo*. *In silico* approaches may include computational models of evolution (See Results, Chapter 1 (Großkopf et al. 2016)). Implementations using varying

mutations rates or varying times of appearance of the mutator phenotype may give us valuable information about the impact of high mutation rates in the evolvability of sympatric diversification in an unstructured environment. Similar approaches may be used *in vivo*, for example by evolution replay experiments by using ancestral strains lacking the mutator allele at the time of fixation and observing the speed of adaptive diversification with ancestral mutation rates.

All these studies will provide a complete picture of the mechanisms of the emergence of bacterial diversity. New insights will arise on the most fundamental life process, evolution, and for public health including the evolutionary dynamics of emergent and nosocomial diseases and new potential therapies integrating the evolutionary dynamics of pathogens.

REFERENCES

- Arenas, C. D., and T. F. Cooper. 2013. Mechanisms and selection of evolvability: experimental evidence. *FEMS Microbiology Reviews* 37 (4):572-582.
- Banerjee, R., B. Johnston, C. Lohse, S. Chattopadhyay, V. Tchesnokova, E. V. Sokurenko, and J. R. Johnson. 2013. The clonal distribution and diversity of extraintestinal *Escherichia coli* isolates vary according to patient characteristics. *Antimicrobial agents and chemotherapy*. 57 (12):5912-5917.
- Bantinaki, E., R. Kassen, C. G. Knight, Z. Robinson, A. J. Spiers, and P. B. Rainey. 2007. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. mutational origins of wrinkly spreader diversity. *Genetics* 176 (1):441-453.
- Barrick, J.E., D.S. Yu, S.H. Yoon, H. Jeong, T.K. Oh, D. Schneider, R.E. Lenski, and J.F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461 (7268):1243-U74.
- Bertani, G. 1951. Studies on lysogenesis 1. the mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology* 62 (3):293-300.
- Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489 (7417):513.
- Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 105 (23):7899-7906.
- Brown, W. L., and E. O. Wilson. 1956. Character Displacement. *Systematic Zoology* 5 (2):49-64.

- Chao, L., and E. C. Cox. 1983. Competition between high and low mutating strains of *Escherichia coli*. *Evolution* 37 (1):125-134.
- Cohan, F. M. 2001. Bacterial species and speciation. *Systematic Biology* 50 (4):513-24.
- Cohan, F. M. 2002. What are bacterial species?. *Annual Review of Microbiology* 56:457-487.
- Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 100 (3):1072-7.
- Cooper, V. S., and R. E. Lenski. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407 (6805):736-739.
- Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of Bacteriology* 183 (9):2834-2841.
- Crozat, E., N. Philippe, R. E. Lenski, J. Geiselmann, and D. Schneider. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169 (2):523-532.
- Crozat, E., C. Winkworth, J. Gaffé, P. F. Hallin, M. A. Riley, R. E. Lenski, and D. Schneider. 2010. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Molecular Biology and Evolution* 27 (9):2113-28.
- 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.

- Davis, BD, and ES Mingioli. 1950. Mutants of *Escherichia coli* requiring methionine or vitamin B12. *Journal of Bacteriology* 60 (1):17-28.
- Denamur, E., S. Bonacorsi, A. Giraud, P. Duriez, F. Hilali, C. Amorin, E. Bingen, A. Andreumont, B. Picard, F. Taddei, and I. Matic. 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *Journal of Bacteriology* 184 (2):605-609.
- Denamur, E., G. Lecointre, P. Darlu, O. Tenailon, C. Acquaviva, C. Sayada, I. Sunjevaric, R. Rothstein, J. Elion, F. Taddei, M. Radman, and I. Matic. 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 103 (5):711-721.
- Ferenci, T. 1996. Adaptation to life at micromolar nutrient levels: The regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. *Fems Microbiology Reviews* 18 (4):301-317.
- Ferenci, T. 2016. Trade-off mechanisms shaping the diversity of bacteria. *Trends in Microbiology* 24 (3):209-223.
- Fierer, N., and R. B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* 103 (3):626-631.
- Fierer, N., and J. T. Lennon. 2011. The generation and maintenance of diversity in microbial communities. *American Journal of Botany* 98 (3):439-448.
- Friesen, M. L., G. Saxer, M. Travisano, and M. Doebeli. 2004. Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution* 58 (2):245-260.

- Giraud, A., I. Matic, O. Tenaillon, A. Clara, M. Radman, M. Fons, and F. Taddei. 2001. Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut. *Science* 291 (5513):2606-2608.
- Gresham, D., and J. Hong. 2015. The functional basis of adaptive evolution in chemostats. *Fems Microbiology Reviews* 39 (1):2-16.
- Großkopf, T., J. Consuegra, J. Gaffé, J.C. Willison, R.E. Lenski, O. S. Soyer, and D. Schneider. 2016. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC Evolutionary Biology* 16 (1):163.
- Hall, B. G. 1982. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *Journal of Bacteriology* 151 (1):269-273.
- Hall, B. M., C. X. Ma, P. Liang, and K. K. Singh. 2009. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* 25 (12):1564-1565.
- Jayaraman, R. 2009. Mutators and hypermutability in bacteria: the *Escherichia coli* paradigm. *Journal of Genetics* 88 (3):379-391.
- Jeong, H., V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S. H. Choi, A. Couloux, S. W. Lee, S. H. Yoon, L. Cattolico, C. G. Hur, H. S. Park, B. Segurens, S. C. Kim, T. K. Oh, R. E. Lenski, F. W. Studier, P. Daegelen, and J. F. Kim. 2009. Genome Sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *Journal of Molecular Biology* 394 (4):644-652.

- Kassen, R. 2009. Toward a general theory of adaptive radiation insights from microbial experimental evolution. In *Year in Evolutionary Biology 2009*, edited by C. D. Schlichting and T. A. Mousseau, 3-22. Malden: Wiley-Blackwell.
- King, T., A. Ishihama, A. Kori, and T. Ferenci. 2004. A regulatory trade-off as a source of strain variation in the species *Escherichia coli*. *Journal of Bacteriology* 186 (17):5614-5620.
- Koser, S. A. 1923. Utilization of the salts of organic acids by the colon-aerogenes group. *Journal of Bacteriology* 8 (5):493-520.
- Kurabayashi, K., Y. Hirakawa, K. Tanimoto, H. Tomita, and H. Hirakawa. 2014. Role of the CpxAR Two-Component signal transduction system in control of fosfomycin resistance and carbon substrate uptake. *Journal of Bacteriology* 196 (2):248-256.
- Lara, F. J. S., and J. L. Stokes. 1952. Oxidation of citrate by *Escherichia coli*. *Journal of Bacteriology* 63 (3):415-420.
- Le Gac, M., M. D. Brazas, M. Bertrand, J. G. Tyerman, C. C. Spencer, R. E. W. Hancock, and M. Doebeli. 2008. Metabolic changes associated with adaptive diversification in *Escherichia coli*. *Genetics* 178 (2):1049-1060.
- Le Gac, M., and M. Doebeli. 2010. Epistasis and frequency dependence influence the fitness of an adaptive mutation in a diversifying lineage. *Molecular Ecology* 19 (12):2430-2438.
- Le Gac, Mickael, J. Plucain, T. Hindre, R. E. Lenski, and D. Schneider. 2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 109 (24):9487-9492.

- LeClerc, J. E., B. G. Li, W. L. Payne, and T. A. Cebula. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274 (5290):1208-1211.
- Leiby, N., and C. J. Marx. 2014. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *Plos Biology* 12 (2):10.
- Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli* 1. adaptation and divergence during 2,000 generations. *American Naturalist* 138 (6):1315-1341.
- Lenski, R. E., and M. Travisano. 1994. Dynamics of adaptation and diversification of a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences of the United States of America* 91 (15):6808-6814.
- Lenski, R. E., M. J. Wisser, N. Ribeck, Z. D. Blount, J. R. Nahum, J. J. Morris, L. Zaman, C. B. Turner, B. D. Wade, R. Maddamsetti, A. R. Burmeister, E. J. Baird, J. Bundy, N. A. Grant, K. J. Card, M. Rowles, K. Weatherspoon, S. E. Papoulis, R. Sullivan, C. Clark, J. S. Mulka, and N. Hajela. 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proceedings of the Royal Society B-Biological Sciences* 282 (1821):9.
- Levert, M., O. Zamfir, O. Clermont, O. Bouvet, S. Lespinats, M. C. Hipeaux, C. Branger, B. Picard, C. Saint-Ruf, F. Norel, T. Balliau, M. Zivy, H. Le Nagard, S. Cruvellier, B. Chane-Woon-Ming, S. Nilsson, I. Gudelj, K. Phan, T. Ferenci, O. Tenaillon, and E. Denamur. 2010. Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections. *Plos Pathogens* 6 (9):19.

- Levins, Richard. 1968. *Evolution in Changing Environments: Some Theoretical Explorations*: Princeton University Press.
- Lieberman, T. D., J. B. Michel, M. Aingaran, G. Potter-Bynoe, D. Roux, M. R. Davis, D. Skurnik, N. Leiby, J. J. LiPuma, J. B. Goldberg, A. J. McAdam, G. P. Priebe, and R. Kishony. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics* 43 (12):1275-80.
- Lister, Bradford C. 1976. The nature of niche expansion in west indian anolis lizards II: evolutionary components. *Evolution* 30 (4):677-692.
- Lo, Y., L. X. Zhang, B. Foxman, and S. Zollner. 2015. Whole-genome sequencing of uropathogenic *Escherichia coli* reveals long evolutionary history of diversity and virulence. *Infection Genetics and Evolution* 34:244-250.
- Loso, J. B. 2000. Ecological character displacement and the study of adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 97 (11):5693-5695.
- Ma, W. T., G. V. Sandri, and S. Sarkar. 1992. Analysis of the luria-delbruck distribution using discrete convolution powers. *Journal of Applied Probability* 29 (2):255-267.
- Maddamsetti, R., R. E. Lenski, and J. E. Barrick. 2015. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200 (2):619.
- Maharjan, R., S. Seeto, L. Notley-McRobb, and T. Ferenci. 2006. Clonal adaptive radiation in a constant environment. *Science* 313 (5786):514-517.

- Matic, I., M. Radman, F. Taddei, B. Picard, C. Doit, E. Bingen, E. Denamur, and J. Elion. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* 277 (5333):1833-1834.
- Mayr, Ernst. 1963. *Animal Species and Evolution*: Harvard University Press.
- McDonald, M. J., S. M. Gehrig, P. L. Meintjes, X. X. Zhang, and P. B. Rainey. 2009. Adaptive divergence in experimental populations of *Pseudomonas fluorescens* IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics* 183 (3):1041-1053.
- McDonald, M. J., Y. Y. Hsieh, Y. H. Yu, S. L. Chang, and J. Y. Leu. 2012. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Current Biology* 22 (13):1235-1240.
- Oh, M. K., L. Rohlin, K. C. Kao, and J. C. Liao. 2002. Global expression profiling of acetate-grown *Escherichia coli*. *Journal of Biological Chemistry* 277 (15):13175-13183.
- Oliver, A., R. Canton, P. Campo, F. Baquero, and J. Blazquez. 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* 288 (5469):1251-1253.
- Pelosi, L., L. Kuhn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173 (4):1851-1869.
- Philippe, Nadege, E. Crozat, R. E. Lenski, and D. Schneider. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *Bioessays* 29 (9):846-860.

- Plucain, J., T. Hindre, M. Le Gac, O. Tenaillon, S. Cruveiller, C. Medigue, N. Leiby, W. R. Harcombe, C. J. Marx, R. E. Lenski, and D. Schneider. 2014. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* 343 (6177):1366-1369.
- Prosser, J. I., B. J. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, and J. P. Young. 2007. The role of ecological theory in microbial ecology. *Nature Reviews Microbiology* 5 (5):384-92.
- Rainey, P. B., A. Buckling, R. Kassen, and M. Travisano. 2000. The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends in Ecology & Evolution* 15 (6):243-247.
- Rainey, P. B., and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. *Nature* 394 (6688):69-72.
- Rosenzweig, R. F., R. R. Sharp, D. S. Treves, and J. Adams. 1994. "Microbial Evolution In A Simple Unstructured Environment - Genetic Differentiation In *Escherichia-Coli*." *Genetics* 137 (4):903-917.
- Rozen, D. E., and R. E. Lenski. 2000. Long-term experimental evolution in *Escherichia coli* VIII. Dynamics of a balanced polymorphism." *American Naturalist* 155 (1):24-35.
- Rozen, D. E., D. Schneider, and R. E. Lenski. 2005. Long-term experimental evolution in *Escherichia coli* XIII. Phylogenetic history of a balanced polymorphism. *Journal of Molecular Evolution* 61 (2):171-180.

- Rozen, D. E., N. Philippe, J. A. de Visser, R. E. Lenski, and D. Schneider. 2009. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters* 12 (1):34-44.
- Scheutz, F, and NA Strockbine. Genus I. *Escherichia*. In *Bergey's Manual of Systematics Bacteriology*. New York, NY: Springer.
- Schluter, D. 2000a. Ecological character displacement in adaptive radiation." *American Naturalist* 156:S4-S16.
- Schluter, Dolph. 2000b. The ecology of adaptive radiations. . Oxford University Press.
- Shaver, A. C., P. G. Dombrowski, J. Y. Sweeney, T. Treis, R. M. Zappala, and P. D. Sniegowski. 2002. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162 (2):557-566.
- Shaver, A. C., and P. D. Sniegowski. 2003. Spontaneously arising mutL mutators in evolving *Escherichia coli* populations are the result of changes in repeat length. *Journal of Bacteriology* 185 (20):6076-6082.
- Simpson, George Gaylord. 1953. The major features of evolution. New York: Columbia University Press.
- Sniegowski, P. D., P. J. Gerrish, T. Johnson, and A. Shaver. 2000. The evolution of mutation rates: separating causes from consequences. *Bioessays* 22 (12):1057-1066.
- Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* 387 (6634):703-705.
- Spencer, C. C., M. Bertrand, M. Travisano, and M. Doebeli. 2007. Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *Plos Genetics* 3 (1):6.

- Spiers, A. J., J. Bohannon, S. M. Gehrig, and P. B. Rainey. 2003. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Molecular Microbiology* 50 (1):15-27.
- Spiers, A. J., S. G. Kahn, J. Bohannon, M. Travisano, and P. B. Rainey. 2002. Adaptive divergence in experimental Populations of *Pseudomonas fluorescens* I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* 161 (1):33-46.
- Stern, D. L. 2000. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* 54 (4):1079-1091.
- Stuart, Y. E., and J. B. Losos. 2013. Ecological character displacement: glass half full or half empty? *Trends in Ecology & Evolution* 28 (7):402-408.
- Taddei, F., M. Radman, J. MaynardSmith, B. Toupance, P. H. Gouyon, and B. Godelle. 1997. Role of mutator alleles in adaptive evolution. *Nature* 387 (6634):700-702.
- Tamames, J., J. J. Abellan, M. Pignatelli, A. Camacho, and A. Moya. 2010. Environmental distribution of prokaryotic taxa. *BMC Microbiology* 10:14.
- Tawfik, D. S. 2014. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency?. *Current Opinion in Chemical Biology* 21:73-80.
- Tenaillon, O., H. Le Nagard, B. Godelle, and F. Taddei. 2000. Mutators and sex in bacteria: Conflict between adaptive strategies. *Proceedings of the National Academy of Sciences of the United States of America* 97 (19):10465-10470.
- Tenaillon, O., D. Skurnik, B. Picard, and E. Denamur. 2010. The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology* 8 (3):207-217.

- Tenaillon, O., B. Toupance, H. Le Nagard, F. Taddei, and B. Godelle. 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* 152 (2):485-493.
- Tenaillon, O., J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, S. Wielgoss, S. Cruveiller, C. Médigue, D. Schneider, and R. E. Lenski. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* advance online publication.
- Thomas, T., J. Gilbert, and F. Meyer. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2:3.
- Treves, D. S., S. Manning, and J. Adams. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular Biology and Evolution* 15 (7):789-797.
- Trobner, W., and R. Piechocki. 1984. Selection against hypermutability in *Escherichia coli* during long-term evolution." *Molecular & General Genetics* 198 (1):177-178.
- Vasi, F., M. Travisano, and R. E. Lenski. 1994. Long-term experimental evolution in *Escherichia coli* II. Changes in life-history traits during adaptation to a seasonal environment. *American Naturalist* 144 (3):432-456. doi: 10.1086/285685.
- Vaz-Moreira, I., O. C. Nunes, and C. M. Manaia. 2014. Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *Fems Microbiology Reviews* 38 (4):761-778.
- Vemuri, G. N., E. Altman, D. P. Sangurdekar, A. B. Khodursky, and M. A. Eiteman. 2006. Overflow metabolism in *Escherichia coli* during steady-state growth: Transcriptional

- regulation and effect of the redox ratio. *Applied and Environmental Microbiology* 72 (5):3653-3661.
- Viraphong, L. 2015. Dynamique expérimentale et théorique des taux de mutation. Doctor in Biology, EDCSV, Université Joseph Fourier.
- Wielgoss, S., J. E. Barrick, O. Tenaillon, M. J. Wisner, W. J. Dittmar, S. Cruveiller, B. Chanee-Woon-Ming, C. Medigue, R. E. Lenski, and D. Schneider. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proceedings of the National Academy of Sciences of the United States of America* 110 (1):222-227.
- Wisner, M. J., N. Ribeck, and R. E. Lenski. 2013. Long-term dynamics of adaptation in asexual populations. *Science* 342 (6164):1364-1367.
- Yang, L., L. Jelsbak, R. Lykke Marvig, S. Damkiaer, C. T. Workman, M. Holm Rau, S. Kirkelund Hansen, A. Folkesson, H. K. Johansen, O. Ciofu, N. Hoiby, M. O. A. Sommer, and S. Molin. 2011. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences of the United States of America* 108 (18):7481-7486.
- Yoder, J. B., E. Clancey, S. Des Roches, J. M. Eastman, L. Gentry, W. Godsoe, T. J. Hagey, D. Jochimsen, B. P. Oswald, J. Robertson, B. A. J. Sarver, J. J. Schenk, S. F. Spear, and L. J. Harmon. 2010. Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology* 23 (8):1581-1596.

Abstract

Diversification events are central issues in evolution since they generate phenotypic innovation such as colonization of novel ecological niches and, ultimately, speciation. To study the ecological and molecular drivers of adaptive diversification, we used the longest still-running evolution experiment. Twelve independent populations are propagated in a glucose limited minimal medium from a common ancestor of *Escherichia coli* by serial daily transfers since 1988 for more than 60,000 generations. In one of the twelve populations, called Ara-2, a unique diversification event occurred: two phenotypically-differentiated lineages, named S (Small) and L (Large) according to their cell size, emerged from a common ancestor at ~ 6500 generations. The two lineages co-exist ever since, owing to negative frequency-dependent selection whereby each lineage is favored and invades the other when rare, such that no lineage gets extinct. Moreover, and before the split between the two S and L lineages, the population Ara-2 evolved a hypermutator phenotype, owing to a defect in a DNA repair gene. The objective of this thesis is to characterize the ecological, physiological and molecular mechanisms that allowed the emergence and stable co-existence of the S and L lineages.

First, we used a combination of *in vivo* and *in silico* experimental evolution to determine the ecological and physiological drivers of the emergence of the polymorphism. Several ecological mechanisms including tradeoff, seasonality and character displacement are involved in the emergence and long-term persistence of diversity. In particular, we showed that the L lineage secretes acetate which generates a new ecological opportunity that the S lineage exploited. In addition, the S and L lineages became fitter and fitter over time in their respective ecological niches, respectively acetate and glucose. Second, we propagated S and L clones separately to remove competition between the two lineages. In these conditions, frequency-dependent interactions between the S and L clones evolved separately were completely abolished, revealing the importance of competition in the maintenance of the polymorphism. Third, we combined genetic, physiological and biochemical approaches to determine the role of an S-specific mutation that was previously found in *arcA*, encoding a global regulator, in the emergence of the S and L polymorphism. We showed that the evolved *arcA* allele conferred to the S lineage the capacity to grow on acetate by increasing the transcription of target genes involved in acetate consumption. During this study, we found an additional mutation, in the *acs* gene involved in acetate metabolism, that was also involved in the emergence of the S lineage. We further showed that these two mutations were favorable to the S lineage early during its emergence, and that other mutations occurred later that interacted epistatically with the *acs* and *arcA* evolved alleles. Therefore, these data showed that the establishment and further maintenance of the S and L polymorphism was a multi-step process involving epistatic interactions between several mutations. Fourth, we identified the long-term dynamics of mutation rates in this divergent population. A first early rise of a hypermutator was followed by a full reversion of this mutator state twice independently in each of the two S and L lineages.

The emergence of a long-term bacterial polymorphism reflects a complex restructuring of the metabolic and regulatory networks in the co-existing lineages, resulting in the generation and exploitation of a new ecological opportunity. Competition and evolution of divergent resource consumption were the selective forces driving the maintenance of the polymorphism.

Résumé

Les événements de diversification adaptative sont des éléments primordiaux de l'évolution. En effet, ils engendrent des innovations phénotypiques telles que la colonisation de nouvelles niches écologiques et au final, la spéciation. Afin d'étudier les ressorts écologiques et moléculaires de la diversification adaptative, nous utilisons la plus longue des expériences d'évolution en cours. Depuis 1988, soit plus de 60 000 générations, douze populations indépendantes d'un ancêtre commun d'*Escherichia coli* sont propagées quotidiennement dans un milieu minimum comportant une faible quantité de glucose. Un événement unique de diversification s'est produit dans une des 12 populations (Ara-2). Deux lignées de phénotypes différents sont apparues après environ 6500 générations, les S pour «Small» et les L pour «Large», chacune présentant des tailles cellulaires différentes. Les deux lignées coexistent grâce à une sélection négative dépendant de la fréquence qui favorise la lignée la plus rare et permet de supplanter sa concurrente; ainsi, aucune des deux lignées ne s'éteint. Avant l'événement de diversification, la population Ara-2 a développé un phénotype hypermutateur suite à la mutation d'un gène de réparation de l'ADN. L'objectif de cette thèse est de caractériser les mécanismes écologiques, physiologiques et moléculaires sous-tendant l'émergence et la coexistence des lignées S et L.

En premier lieu, nous avons utilisé un ensemble d'expériences d'évolution *in vivo* et *in silico* afin de déterminer les moteurs écologiques et physiologiques de l'émergence de ce polymorphisme. Plusieurs mécanismes écologiques, incluant les compromis (trade-off évolutifs), la saisonnalité et les déplacements de caractères interviennent dans l'émergence et la persistance de la diversité au long terme. Nous avons montré que la lignée L, en produisant de l'acétate, créait une nouvelle opportunité écologique exploitée par les S. De plus, au cours du temps, les S et les L s'adaptent à leur niche écologique, respectivement l'acétate et le glucose. En second lieu, nous avons cultivé les S et les L séparément pour éliminer la compétition entre les deux lignées. Dans ces conditions, il y a perte des interactions dépendantes de la fréquence entre les S et les L. Ceci démontre l'importance de la compétition dans le maintien du polymorphisme. En troisième lieu, nous avons combiné des approches génétiques, physiologiques et biochimiques pour déterminer le rôle, dans l'émergence du polymorphisme, d'une mutation spécifique aux S survenant dans le gène *arcA*, codant un régulateur global. Nous avons montré que l'allèle évolué de *arcA* augmentait la transcription de gènes du métabolisme de l'acétate dans la lignée S. Au cours de cette étude, nous avons identifié une mutation supplémentaire dans le gène *acs*, impliqué dans le métabolisme de l'acétate, intervenant dans l'émergence de la lignée S. Nous avons aussi démontré que ces deux mutations étaient favorables à la lignée S au début de son émergence, puis que des mutations plus tardives agissaient de façon épistatiques avec les allèles évolués de *acs* et de *arcA*. Ainsi, ces résultats démontrent que l'établissement et le maintien du polymorphisme des S et des L est un processus en plusieurs étapes nécessitant des interactions épistatiques entre plusieurs mutations. En quatrième lieu, nous avons identifié la dynamique au long terme des taux de mutations dans cette population. L'apparition et l'invasion rapide du phénotype hypermutateur est suivie d'une réversion complète mais indépendante dans chacune des lignées S et L indépendamment.

L'émergence d'un polymorphisme bactérien durable reflète une restructuration complexe des réseaux métaboliques et de régulation dans ces lignées qui co-existent, ce qui aboutit à l'apparition et à l'exploitation de nouvelles opportunités écologiques. La compétition et l'évolution de l'utilisation de ressources différentes sont des forces sélectives permettant le maintien du polymorphisme.