University of Groningen

Adaptation and diversification of Escherichia coli K12 MC1000 in a complex environment
Puentes Téllez, Pilar

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2014

Adaptation and diversification of *Escherichia coli K12 MC1000* in a complex environment

Pilar Eliana Puentes Téllez

2014

Cover design: P.E Puentes Téllez
Cover pictures: NIAID (Image of public domain)

rijksuniversiteit
groningen

Adaptation and diversification of *Escherichia coli K12* MC1000 in a complex environment

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 23 juni 2014 om 09.00 uur

door

Pilar Eliana Puentes Téllez

geboren op 23 maart 1980
te Bogota, Colombia

Promotor
Prof. dr. ir. J.D. van Elsas

Beoordelingscommissie
Prof. dr. A. Eliasson Lantz
Prof. dr. J. Kok
Prof. dr. L.W. Beukeboom

To Hans and Anna Lucía

Contents

Chapter 1

General Introduction and

Scope of the Thesis

A heterogeneous environment offers many niches that may vary in either space or time. Heterogeneity and complexity are essential characteristics of many environments in nature (soil, plants, water, air), as well as of environments in the industrial and scientific world. It is known that environment and its complexity act upon the genetic and phenotypic characteristics of living organisms, shaping evolutionary outcomes in a selective manner. Specifically for bacteria, (stress) factors such as differential nutrient and oxygen levels, osmolarity, temperature, acidity and other chemical conditions of the habitat can vary in time and space. Thus differential selective forces are exerted that together have an impact on the local forms. The individual or simultaneous presence of these factors defines an environment as a more or less complex one.

A changing environment thus creates conditions that can be stressful for bacteria, which in turn depend on physiological and genetic acclimation mechanisms to be able to survive and remain active in the face of the stress. For example, organisms may survive immediate stress by altering their allocation of resources from growth to survival pathways. On the other hand, too extreme stresses can force bacteria into dormancy (Farrar and Reboli 1999, Suzina *et al*., 2004) or even kill them. The survival strategies impose physiological costs onto the organism, and thus the composition of an active microbial community can be altered, creating shifts in ecosystem-level, energy and nutrient flows (Schimel *et al*., 2007). Evolution has equipped bacteria with an inherent and sophisticated physiological complexity allowing them to respond flexibly to stresses and also to adapt. The study of adaptation has thus become crucial in understanding the outcomes of evolution.

The rapid emergence of novel molecular technologies in the last decades has brought vast possibilities to gather information about bacteria adapting to diverse environments. Thus, evolutionary and ecological questions that could previously only be approached on the basis of comparative or theoretical methods are increasingly amenable to direct study using the power of the molecular methods (Barrick *et al*., 2013).

However, most of the information about bacterial adaptation collected up to now (Lenski *et al*., 1991; Bennett *et al*., 1993; Barrett *et al*., 2005; Hughes *et al*., 2007; Cooper *et al*., 2010) has been obtained from "simple" environments and does not reflect the heterogeneity and complexity that are normally found in natural and even artificial (industrial) systems. Thus, the forces that drive selection as well as the scope of adaptation to such environments still need to be assessed using experimental setups that reflect these, including their dynamism and stresses.

Experimental evolution as a tool to study diversification in bacterial populations.

Experimental evolution is the study of evolutionary changes occurring in experimental populations as a consequence of the conditions (environmental, genetic, social and so forth) imposed (or not) by the experimenter (Kawecki *et al*., 2012). Experimental evolution represents an alternative research framework that provides the opportunity to study evolutionary processes like adaptation in real time. Thus, patterns of phylogeny, divergence, genome structure and DNA expression level can be tested in experimental evolution setups. The past decade has seen the fast growth of studies that tap into the potential of this tool (Kawecki *et al*., 2012). Such studies are greatly aided by the power of the technological advances that currently speed up and facilitate the analysis of the genetic and molecular evolutionary outcomes.

Most experiments on microbial evolution are conceptually simple (Elena *et al*, 2003). Perhaps the simplest evolution experiments maintain growing bacterial populations that are derived from a single ancestral genotype in a uniform environment, such that selective forces either remain constant or fluctuate in a controlled way (Helling *et al*., 1987; Rosenzweig *et al*. 1994). In natural systems, it is usually difficult to dissect the adaptive changes that occurred in detail because relevant events happened in the proximal or distant past and may have involved unique circumstances that are currently unknown. Thus, bringing evolution into the laboratory has clear advantages, including both the ability to generate a 'fossil' record for later study and the ability to test the predictability and repeatability of evolution across replicate populations (Barrick *et al*.,

2013). In the laboratory, experimental evolution can be based on the use of continuous culture, in which both the replenishment of resources and the removal of individuals occur at a constant rate. Alternatively, periodic serial transfer of a proportion of the population to a new microcosm with fresh resources can be used, in a setup coined sequential (serial)-batch culture. Samples of the ancestral population can be stored (for example, frozen at –80 °C), as can samples from various time points in the experiments. After a population has been propagated for some time, the ancestral and derived (evolved) genotypes can be compared with respect to genetic or phenotypic properties of interest, providing information on the dynamics and extent of the adaptive process (Elena *et al.*, 2003). The so-called "microbial advantage" in experimental evolution refers to the fact that the inherent characteristics of microorganisms (short generation times and large population sizes that are achieved) allow experiments to run for up to thousands of generations (Bennett *et al.*, 2009). Other characteristics are the easy creation of replicate linages and environmental controls, the relatively "easy to handle" genome sizes and the fairly uncomplicated preservation of mixed populations or clones.

Many evolution experiments have sought to understand how populations of organisms such as *Escherichia coli*, *Pseudomonas fluorescens* and *Saccharomyces cerevisiae* adapt to specific environmental conditions, usually defined in terms of a particular factor, such as temperature (Bennett *et al.*, 1993), pH (Hughes *et al.*, 2007), nutrition (Barrett *et al.*, 2005, Cooper *et al.*, 2010) or other defined stressors (Dhar, *et al.* 2011). The majority of these studies initially addressed adaptation as the ultimate consequence of evolution. However, recently the new technologies have allowed to dig into the genetic bases of such adaptation and to observe the deeper outcomes of their trajectories, especially in the presumed genetic targets of selection.

The longest-running and most famous bacterial evolution experiment in sequential batch was started in 1988, when twelve populations were founded by the same strain of *Escherichia coli*. Since then, the populations have been propagated by daily serial transfer, using a 1:100 dilution, in the same defined environment, yielding almost 60,000 generations (Lenski *et al.*, 1991). The growth medium used was "Davis" minimal medium supplemented with glucose at 25 mg/mL (DM25). An arabinose-utilization

phenotype served as a marker in the competition experiments that were performed to measure the relative fitness of the evolved forms. During each daily cycle, the evolving cells experienced the transition from starvation to growth upon transfer into fresh medium, and then back to starvation once the glucose had been exhausted by the growing population. Therefore, selection can act on mutations that affect physiology and performance during the lag, exponential and stationary phases of growth. All replicate populations achieved substantial fitness gains, averaging about 70% after 20,000 generations (Cooper *et al.*, 2000). The rate of the fitness increase was high during the first 2,000 generations, and became progressively slower as the generations elapsed. With time, the populations showed numerous parallel genetic and phenotypic changes, and such targeted parallel changes were defined as the hallmark of adaptive evolution. Of particular interest was the finding of two highly interconnected physiological networks governing DNA superhelicity and the stringent response. These were demonstrated to be involved in the phenotypic and genetic adaptation of the experimental populations (Philippe *et al.*, 2007). In addition, another striking finding was the observation of a mutant strain after 31,000 generations that could metabolize citrate in the presence of oxygen, which the ancestor *E. coli* could not do (Blount *et al.* 2008). This allowed the mutant to outcompete its siblings and dominate the population after 33,000 generations. This experimental system, being the longest and most studied system in experimental evolution, has yielded ground-breaking results on the effects of environmental selection.

Fitness - the measure of adaptation.

Adaptation in microorganisms can be quantified by measuring the changes in fitness (as compared to the ancestor or parental form) that these incurred in the experimental environment. Fitness encompasses the ability of organisms to survive and reproduce in the environment in which they evolved (Orr *et al.*, 2009). In an experimental evolution setup, fitness can be measured using head-to-head (1:1) competition between the evolved form and its ancestor that is genetically or phenotypically marked. Thus, the fitness of an evolved type is generally expressed relative to its ancestor. Unless otherwise specified, the competition environment uses the same environment as that used for the experimental evolution. However, fitness can also be measured in alternative

environments to assess so-called "correlated" fitness, which addresses the presence of physiological *trade-offs* in the population. Evolutionary dynamics is often visualized in terms of the successive steps up on the fitness "ladder" as populations 'climb' ridges and peaks in a fitness landscape (Orr, 2009). In this scenario, the landscape is depicted as a topographical map showing the individual genotypes and their corresponding fitness values. Thus, an evolutionary trajectory of genetic changes can be visualized as a 'walk' and adaptation as a 'climb' in such a fitness landscape (Barrick *et al*., 2013).

Before the fitness assessment, the competing forms are normally acclimatized separately to the environment in which they will eventually compete. This procedure (phenotypic acclimation) ensures that the populations have comparable initial physiological states and cell densities (Hughes *et al*., 2007). Thus, any significant difference between both ancestral and derived genotypes will have just a genetic basis (Lenski *et al*., 1998) and will not be influenced by the possible "shock" faced by either of the two forms when entering the so-called "competing" environment. After competition, the relative fitness (W) of the derived genotype is calculated as the ratio of the Malthusian parameters (m) of both the evolved forms and the ancestor as $W= m$ Evolved form $/m$ Ancestor. Being $m = \ln[$(density at end of competition)/(density at time zero of competition)$]$ (Lenski *et al*., 1991). Density is commonly obtained as CFU/ml.

Up to now, most experimental evolution studies have assessed the fitness of randomly selected colonies (selected individually or as a pool) or of aliquots of mixed end-populations (Leroi *et al*., 1994, Lenski *et al*., 1998, Pepin *et al*., 2006, Enne *et al*., 2004, Hughes *et al*., 2007, MacLean *et al*., 2004) ("Direct assessment"). In these cases, fitness is reported as the average value of relative fitness values against the ancestor. In such cases, homogeneity of the evolved forms is assumed. However, it has been demonstrated that, even in simple environments, diversification can occur after adaptation (Helling *et al*., 1987; Rosenzweig *et al*. 1994). From this, it is clearly possible that complex environments produce rather diverged populations that are in fact a collection of sub-populations. Thus, targeting the population average rather than assuming the presence of sub-populations can overlook the presence of diversification. High statistical variation may be present in the populations. A better approach (proposed here) to the

sampling and analyses might be based on preliminary phenotypic or genomic sorting, reflecting the fact that diversification may have occurred as an outcome of the experiment (figure 1.1).
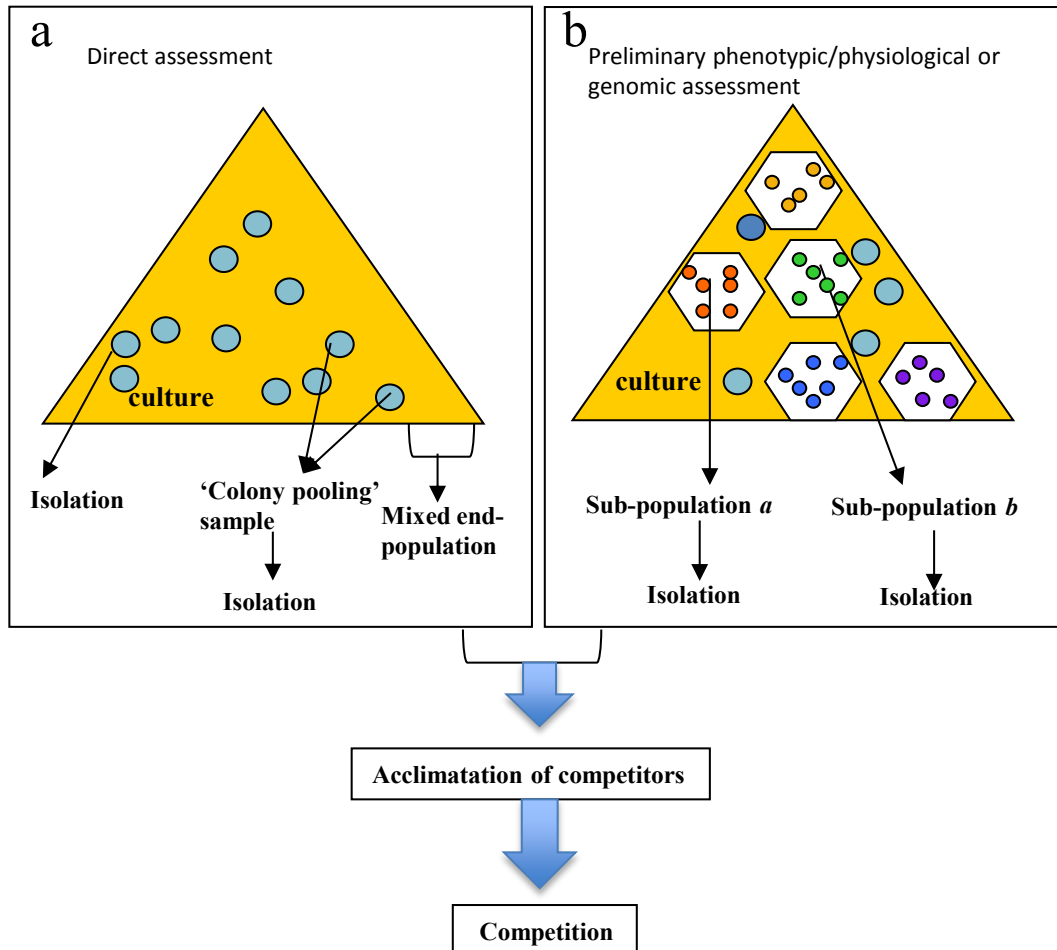
Figure 1.1. a) Direct assessment. Homogeneity in the populations has been widely assumed by the experimenters b). Preliminary phenotypic/physiological or genomic assessment. A proposed model

Zooming in on the essentials: Finding the genetic basis of adaptation.

Initially, the study of bacterial adaptation has been restricted to characterizing evolution at the phenotypic level. However, over the last years it has been possible to come close to the identification of the mutation(s) underpinning the key phenotypes that emerged. Thus, the study of bacterial adaptation to their environment has been narrowed down to the essentials, *i.e.* the observation of the genetic changes (due to mutations) that

occurred and their impact on physiology and ultimately selectability in the environment. A wide variety of evolutionary questions has been addressed, including the dynamics of adaptation and the evolution of diversity and sociality (Brockhurst *et al.*, 2011). The field has acquired the possibility to go from targeted observations (specific genes) to wider perspectives, *i.e.* whole genome observations (Albert *et al*, 2005; Herring *et al*, 2006; Barrick *et al*, 2009; Conrad *et al*, 2009; Atsumi *et al*,2010; Lee *et al*, 2010). Recently, several studies have utilized genome sequencing technologies across evolved clones as well as entire populations (Herron *et al.*, 2013). Thus researchers uncovered single-nucleotide substitutions, insertions/deletions (Barrick *et al*, 2009), large genomic duplications (Herring *et al*, 2006; Conrad *et al*, 2009) and transposable element insertions (Atsumi *et al*, 2010), all events that underpinned the observed adaptations.

Application of comparative genomics between replicate populations can pinpoint those mutations that confer the largest fitness gains, as these are repeatedly selected for in parallel cultures (Elena *et al.*, 2003; Herring *et al*, 2006; Conrad *et al*, 2009). Thus, in experimental evolution most mutations have been fixed because of their fitness benefits and natural selection (Brockhurst, 2011). The genomic studies in this field have also assisted us in the quest to link genetic changes with phenotypic outcome (Knight *et al*, 2006). For example, Beaumont *et al.* (2009) used sequencing to identify mutations conferring bet-hedging behavior in *Pseudomonas fluorescens*. This was attributed to an enzyme which makes part of the pyrimidine and arginine biosynthetic pathways, allowing cells to epigenetically switch on and off the expression of capsule biosynthesis genes (Rainey *et al*. 2011). In spite of this progress, functional explanation of other mutations is often surprisingly poor (Brockhurst, 2011).

What are the targets of selection?

Generally, environmental stress has been perceived to exert negative effects on microbial populations in natural and man-made environments. In the face of stress, bacteria tend to invest in systems that maintain the integrity of the cell membrane, folding of proteins and integrity of their DNA (Booth, 2002). Most bacteria thus sense local

conditions, determine when these become deleterious and then induce such protective systems.

An "adaptive" mutation can occur in bacteria that are exposed to growth-limiting environments, offering either a growth advantage and/or stress relief to the cell (Hersh *et al*., 2004, Foster, 1999). Fitness-enhancing mutations affect either structural genes (or operons) encoding specific metabolic and/or structural proteins, or genes with regulatory functions. The latter range from local regulatory genes resulting in constitutive synthesis of the required enzymes (Cooper *et al*. 2001) to genes encoding global regulators that are involved in cellular networks (Cooper *et al*., 2003).

In experiments with *Escherichia coli* with carbon sources such as glucose, glycerol or 1,2-propanediol in batch or continuous culture conditions, mutations were found in genes for the transport and/or consumption of these substrates as well as in those (structural and regulatory) encoding functions that turned out to be unnecessary in the environment (Ferenci *et al*., 2008; Lee *et al*., 2010; Cooper *et al*., 2001; Herring *et al*., 2006; Pelosi *et al*., 2006). However, these evolutionary pathways could not be regarded as simple local changes. This is due to the fact that specific mutations (for example in a regulatory gene) can have pleiotropic effects that lead to other phenotypes. Most experiments thus do not precisely match the mutation rates found under complex conditions, in which organisms are often nutritionally deprived or otherwise stressed (Barrick *et al*., 2013). In complex environments, mutation rates might increase genome-wide, such that both non-adaptive and adaptive mutations are stimulated ("hypermutation"). The elevated mutation rates are helpful to organisms, allowing them to adapt to sudden and unforeseen threats to survival (Jayaraman, 2011). Such mutagenesis mechanisms appear to be widespread and important in nature (Gonzalez, 2008). At large population sizes and high mutation rates, many cells in the population concurrently acquire mutations of varying adaptive (fitness) values. These forms compete with the ancestral cells and also among themselves for fixation. The one with the mutation of highest benefit will become fixed, while the others may be lost or remain unselected (Jayaraman, 2011).

In experimental evolution setups performed in the laboratory, often low numbers of base substitutions are found (Barrick *et al*., 2009). Because so few mutations apparently accumulated, it is unlikely that the same gene would change in several independently evolving genomes, unless these forms were enriched by selection. Therefore, such genetic parallelism provides a strong signal that the mutations were beneficial.

There are several ways to test if a specific mutation or a set of mutations is a beneficial-adaptive mutation. One strategy is based on the ratio of synonymous to non-synonymous base substitutions (dN/dS) incurred in the mutated form (Barrick *et al* 2013). Another strategy involves the use of genome-editing or genetic exchange to make a basically isogenic construct that differs from another strain by only the single mutation of interest. One can then either test for a change in a trait that is suspected to be related to fitness or compete the two organisms in the environment to determine whether the mutation is beneficial, neutral or deleterious (Barrick *et al* 2013).

Overall, many studies demonstrated that evolutionary innovation often relies on the reorganization of molecular components. Transcriptomics studies applied to Lenski's experiment identified early mutations in genes encoding global regulators that control (1) the stringent response and (2) DNA supercoiling. Thus changes in *spoT, topA* and *fis* (encoding bifunctional (p)ppGpp synthase–hydrolase, DNA topoisomerase 1 and DNA-binding protein Fis, respectively), were discovered in several populations (Cooper *et al*., 2003; Pelosi *et al*., 2006; Crozat *et al*., 2010). Other studies performed with *Escherichia coli* growing in minimal medium have found similar regulatory changes causing the reprogramming of gene expression. Thus *spoT, rpoS* (encoding the stress-specific alternative RNA polymerase σ-factor, σS) and *hfq* (encoding an RNA chaperone that is essential for the regulatory effects of small RNAs (Notkey-McRobb *et al*. 2002; Wang *et al*., 2010) were found to be specific to adaptation. One may conclude that, in general, organisms often evolve by fine-tuning global gene expression rather than by locally restructuring pathways that are involved in specific traits (Hindré *et al*., 2012). The identification of these adaptive mutations enables the investigation of the architecture and chain of networks (systems biology) sustaining fitness improvement and being targets of selection.

What to expect with respect to the evolution of *Escherichia coli* in a complex growth environment?

Adaptive (survival) strategies in bacteria may operate at the individual cell or at the population level, providing an interesting example of complex multicellular behaviour in bacteria. Thus, while coping with hostile and complex environments, bacteria possess mechanisms that generate genetic and/or phenotypic diversity, enhancing the survival chances of the population (Aertsen *et al*., 2004, 2005). This occurs as a consequence of the genetic diversity triggered by the emergence of constitutive or transient mutators. The genetic diversity that is randomly generated by these mutations is subsequently selected for or against by the environment.

Given the fact that spontaneous genomic changes occurring within a population provide raw material for evolutionary processes to take place, ecological opportunities offered by multiple vacant niches commonly found in complex environments are crucial in the evolution of diversity (Rainey and Travisano, 1998). Thus, diversification may occur in an environment with multiple resources as a consequence of intense competition for commonly-used energy sources, yielding specialized forms that shift their nutritional preference and are better able to use as-yet-unexploited resources (MacLean *et al*., 2005).

What could cause such diversification?

Figure 1.2 summarizes the adaptive responses that occur in a population subjected to selective forces that lead to diversification, *i.e*. *(i)* responses at the genetic level (genetic changes), *(ii)* responses at the phenotypic level. The latter can be subdivided in responses at the physiological level and at the gene expression level. This, because physiological changes under stress conditions are mainly and specifically directed towards keeping the energy inside the cell (Chung *et al*., 2006); these changes may include byproduct accumulation, changes in substrate uptake capacity, reduction of growth and increased requirements for energy maintenance. For example, under carbon starvation *Escherichia coli* cells become smaller and rounded, accumulating and using glycogen and polyphosphate (Chung *et al*., 2006). The DNA is condensed and rapid adjustments in metabolism are made. Moreover, the transport of many macromolecular

precursors into the cell is shut down, and ribosome synthesis is blocked. A clear example of observable physiological responses is found in biofilms, and it appears that in such biofilms heterogeneity is prominent (Stewart *et al*., 2008). There are, for instance, regions in the interior of a biofilm in which bacterial growth is arrested due to substrate limitation. In addition, the accumulation of (acidic) waste products in the biofilm interior lowers the local pH and affects the physiological state of the bacterial cells (Stewart *et al*., 2008).
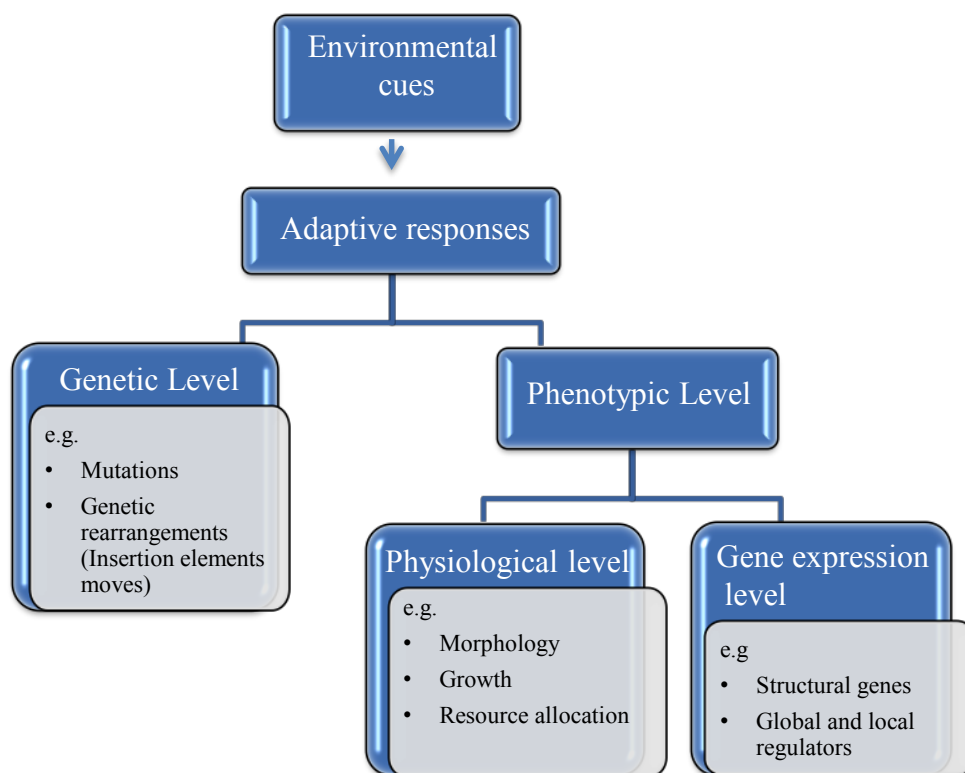
Figure 1.2. Adaptive responses causing microbial diversification

What are the mechanisms underlying and maintaining diversification?

Which factors maintain a stable diversification? This is a crucial question, which is of importance for general as well as industrial microbiology. Until now, several forces have been found to affect the maintenance of diversification. Studies with *E. coli* in spatially-homogeneous glucose-limited continuous (Helling *et al*., 1987; Rosenzweig *et*

al. 1994) and in serial-batch cultures (Rozen eat al., 2000) detected polymorphisms maintained by the ability of particular forms to use the metabolites that were excreted by other types (cross-feeding). A second factor maintaining diversity was found to be the presence of spatial structure. A hallmark experiment (Rainey and Travisano, 1998) found diversification triggered by spatial structure in a *Pseudomonas fluorescens* population. The key role of such spatial structure in maintaining diversity was even more evident when the *P. fluorescens* populations (initially propagated in the spatially heterogeneous environment) were less diverse when switched to an homogeneous environment.

Frequency-dependent selection often lies at the basis of diversification (Friesen *et al*., 2004, Tyerman *et al*., 2005). In frequency-dependent selection, an environment that has favoured the evolution of niche specialists, the maintenance of coexisting genotypes occurs through density-dependent processes. That is, selection operates in a frequency-dependent manner, favouring genotypes when they are rare (because the resources available for them will be abundant) but deselecting these when they reach high frequency (because the per-cell resources are then limiting).

Thus strong forces such as ***frequency-dependent selection*** and other ***interactions*** between different coexisting forms, in combination with the already postulated ***ecological specialization*** under ***spatial heterogeneity*** are the drivers of diversity which otherwise could be lost (Kassen, 2004).

Heterogeneous environment and its potential as a driver of diversification

A heterogeneous environment is a complex environment that holds many niches which may vary in either space or time, and at different scales relative to the generation time of the organism concerned (Kassen 2004). In industrial (as well as natural) environments, bacterial populations are often dispersed and organized by alterations in the factors space (spatial structure) and time (temporal). The effect of spatial structure on adaptation has been addressed by modelling (Rousset 2004). However, only few studies have experimentally shown the adaptation of organisms in structured environments (Chao & Levin 1981; Korona *et al*. 1994; Rainey & Travisano 1998; Dionisio *et al*. 2005;

Habets *et al*. 2006, Saxer *et al*, 2009). Heterogeneity given by spatial structure is a prime cause of diversity and is necessary for its maintenance (Habets *et al*., 2006, Perfeito *et al*., 2008, Ponciano *et al*., 2009, Beaumont *et al*., 2009). None of these studies addressed environment at a smaller spatial scale, that is, the local environment surrounding the cells. In this respect, Kassen (2004) defined heterogeneity by the niches offered when genotype x environment (GxE) as well as genotype x genotype (GxG) interactions act upon the cells, creating subpopulations. Thus, the physiology and response of local cells is determined by the conditions in their "microenvironment" and by the ability of the cells to modulate their metabolic capacities.

In the light of the fact that even subtle differences in the environment can crucially affect bacterial adaption (Tyerman *et al*., 2005) and that spatial structure is important, we endeavor to posit that conditions in bioreactors (being not homogeneous and having small-scale spatial structure) drive diversification and coexistence of divergent types. In the industrial fermentation setups, spatial structure has been little studied, as the environment has been long assumed to be homogeneous. However, substantial heterogeneity may surround cells in different positions in the reactor, where individuals behave differently according to the conditions experienced in their surroundings. This may differentially trigger the responses in the individual cells (Aertsen and Michiels, 2004). Consequently, the type of (micro)environment is a key factor for the development of individuals in a population and so for the population as-a-whole. This can be considered as a driver of its performance. The degree of spatial structure in such an environment is linked to the scale of cultivation (differences between large- and small-scale fermentations). In general, processes at larger scale can show inferior performance, such as reduced yields and productivity accompanied by cell lysis and high byproduct formation (Bylund *et al*., 1998). A main source of spatial structure is the stirring process, which can create zones in which substrate concentrations are low, resulting in nutrient deprivation for the microorganisms. On the other hand, substrate concentrations may be locally quite high, driving local bacterial responses. For example, many bacteria respond by the so-called "overflow" metabolism with high production of acetate when exposed to a glucose concentration above a critical level (Enfors *et al*., 2001).

Cellular metabolism is also intimately linked to the formation of gradients of spatially-explicit factors like oxygen and nutrient availability, level of waste products, variations of pH, redox potential and electron acceptors. Such gradients thus create niches at microscale and select for different bacterial variants. Possibly, forms competing for common resources are differentially selected, resulting in establishment of the best-adapted forms to the local conditions and the niches created (Rayne and Travisano, 1998, Ponciano *et al.*, 2009, Beaumont *et al.*, 2009).

In the light of the foregoing, the physical parameters, together with the metabolic performance of microorganisms in the mix of substrates normally found in bioreactors, drive selection and diversification. Thus, a bioreactor can be considered to encompass an unknown number of spatially-structured environments, where local conditions become heterogeneous with great influence on the local populations (Aertsen *et al.*, 2004).

Oxygen and substrate gradients in a bioreactor

Larsson *et al.* (1996) demonstrated that oxygen gradients in bioreactors will vary with stirrer speed, the time of cultivation and the organism type and process. Cells passing through oxygen-limited zones respond to oxygen depletion, leading to the activation of anaerobic metabolism. Thus, oxygen gradients affect bacteria by exposing them to transient anaerobic conditions, if even of a few seconds (Schweder *et al.*, 1999). Such oxygen gradients may establish temporally different zones in the bioreactor, inciting cell specialization and diversification and consequent fragmentation of the population.

On the other hand, Enfors *et al.* (2001) studied substrate levels at different sites within a large bioreactor. They found that bacterial cells near the feeding point became different (physiologically and morphologically), as carbon source levels were highest at this location. Individual cells that repeatedly passed close to the feeding point were exposed, for a few seconds, to substrate levels higher than the mean level in the reactor. Larsson *et al.* (1996) as well studied the gradients of glucose in time and space in a culture of *Saccharomyces cerevisiae* growing in minimal medium. In this study, a fed-batch concept was used, with glucose as the limiting component. The mean glucose

concentration declined throughout the process and its level was always different at the three sampling ports (bottom/middle/top). Larsson *et al*. (1996) concluded that, if the feed was supplied to the relatively stagnant mixing zone above the top impeller, gradients were more pronounced than when feeding in the well-mixed bottom impeller zone. Moreover, continuous stirring can also create nutrient starvation zones in a bioreactor. Microbial growth rates vary according to the available nutrients and so responses from unrestricted growth to starvation may occur (La Para *et al*., 2002).

The media employed in industrial settings that support high productivities are commonly formulated with inexpensive complex carbon and nitrogen sources (Miller and Churchill, 1986). Given the presence of a vast mixture of substrates in combination with spatial structure and the action of the population upon growth, the conditions inside a bioreactor easily become heterogeneous. Such considerations indicate that the availability and distribution of oxygen and substrates are very important in the general performance of these systems.

Methods for characterizing and describing microbial diversification

Before addressing the tools available to assess the diversification in a microbial population, it is important to understand that adaptation as a response to environmental pressure implies an impact on the different physiological states of the cell. The different growth stages bacteria undergo in culture can be characterized by physiological and metabolical differences. These activities both guide and shape the environment bacteria deal with during growth, and a distinction of the phases can be useful when asssing potential sub-populations. In the exponential growth phase, cells take advantage of abundant nutrients and direct their resources towards replication. Bacteria approaching the stationary phase undergo severe changes in cellular metabolism/physiology as well as morphology, becoming more resistant to several stress conditions. The transition into and survival during stationary phase is under complex physiological control in *Escherichia coli* (Kolter *et al*., 1993). Major regulators (like *rpoS*) are transcribed in late exponential phase and prepare the cell for the challenges it is about to face in the following stage and for long-term survival (*e.g.* the cell morphology regulator *bolA*) (Chang *et al*., 2002).

During stationary phase, the levels of environmental stress increase and many changes at physiological and genetic level can occur. Zambrano *et al*. (1993) found that cells from older cultures had evolved to increased fitness, coined the growth advantage in stationary phase (GASP) phenotype. Cells expressing the GASP phenotype competed better for nutrients released by dying cells in the stationary-phase population. The GASP phenomenon has been observed widely and it may be a common strategy for bacterial survival during prolonged starvation. Thus, stationary phase also involves (genetic) changes that may lead to heterogeneity of populations.

The distinct physiological characteristics of cells in an evolving population (or sub-population) can nowadays be well assessed using advanced tools. Such methodologies may be of physical, chemical and/or molecular nature, together giving information about the response of populations to the environmental cues. Although most of the tools are still not suitable for the *on-line* monitoring of bioprocesses (Lancaster *et al*., 2011), it is undeniable that the current methodologies can thoroughly yield in-depth information from experimental setups in the lab. What are the key methodologies and what can they offer?

First, genome sequencing, although quite laborious, is the current gold standard in mutation detection and single nucleotide polymorphism (SNP) discovery (Kwok *et al*., 2003). Thus high-throughput parallel sequencing can be used to detect genetic changes incurred in bacterial genomes from individuals in the population based on sequence alignments (Beaumont *et al*., 2009, Barrick *et al*., 2013). On the other hand, DNA microarray hybridization has emerged as a very effective methodology to characterize complex microbial communities (Wu, *et al*, 2008). Thus microarrays are widely used in the exploration of organismal transcriptional profiles, addressing the similarities and/or differences in genetic content among experimentally-evolved populations (Ye *et al*., 2001, Riehle *et al*., 2003). Other methods including the study of phenotypic characteristics involve the culturing or growth on specific substrates, which can potentially differentiate metabolic profiles between forms (*e.g* Biolog phenotypic arrays).

Other molecular methods involving genomic fingerprinting like amplified fragment length polymorphism (AFLP) analysis, pulsed-field gel electrophoresis (PFGE) of macro-restriction fragments and repetitive DNA element (rep)-PCR are frequently used methods for genotypic analysis, specially in the experimental settings in the lab. These methods are based on the use of DNA primers corresponding to naturally occurring interspersed repetitive elements in the genome. Examples of rep-PCR are REP, ERIC, BOX and GTG5 PCR methods (Versalovic *et al*., 1994). Generally, rep-PCR fingerprinting is highly reproducible and simple for distinguishing closely related strains, to deduce phylogenetic relationships between strains and to study their diversity in a variety of ecosystems (Rademaker *et al*., 2000). Moreover, more specific techniques with good resolution of large fragments are feasible, like PFGE, which also gives clear data with respect to differences between strains (Bourke, 1996).Suppressive subtractive hybridization (SSH) is another technique allowing the assessment of genomic differences between individual cells which can identify sequences that are present in one bacterial genome, but absent in another one.. Thus, differences between two bacterial genomes can be clearly identified (Bourke, 1996).

On a different notice, to achieve quicker data on processes, the bioprocess industry has increasingly focused on a variety of *on-* and *at-line* monitoring methods (Dabros *et al*., 2008). First, reporter systems (like the green fluorescent protein (GFP)) are used to achieve data on the physiological status of cells and to elucidate diversification. In cultures with strains that are equipped with such reporter systems, the heterogeneity of a population (*i.e.* distribution of the fluorescent compound) is typically monitored by either microscopy (image analysis) or flow cytometry. Both methods are nowadays essential tools for monitoring the physiological heterogeneity of microbial populations, even at single cell level. In addition, these methods allow for monitoring other intrinsic cell properties (*e.g.* cell size) and/or structural/functional parameters (*e.g.* membrane integrity, DNA content), by applying different staining procedures.

Study of diversification and niche partitioning/differentiation in sympatric conditions

The theory of evolutionary branching and adaptive diversification has provided new perspectives for understanding the evolution of diversity caused by ecological interactions. However, we have generally ignored ecological interactions beyond 'scramble' competition for limiting resources. Even in simple laboratory environments, bacterial evolution can lead to niche construction that enables diverged lineages of organisms to coexist for long periods of time (Barrick *et al.*, 2013). However, identifying ecologically-differentiated populations within complex microbial communities remains a true challenge, yet it is critical for interpreting the evolution and ecology of microbes both in industrial settings and in the wild. Direct experimental evidence for adaptive diversification under sympatric complex conditions is still lacking.

In unstructured environments, competitive exclusion was initially thought to control the outcome of different populations, as competition for one resource selects for the fittest variant (Hardin, 1960). However, it has been demonstrated that bacteria in sympatric conditions with one limiting resource can diversify (Rozenzweig, 2009). Key experiments performed by Adams *et al.* (1987) and Rosenzweig *et al*, (1994) demonstrated the presence of stable polymorphisms in *E. coli* in sympatry. This involved the emergence of an acetate utilizer as a consequence of evolution, originated from a unique ancestor. In this system, the use of one unique primary resource (glucose) was effectively increased and in consequence overproduction of acetate occurred, triggering emergence of the type able to utilize this metabolite, next to another compound, glycerol. Three metabolically different strains (glucose, acetate and glycerol consumers) were thus successfully maintained in the population. In another experiment, replicate *E. coli* populations diversified into two coexisting metabolic types due to a diauxic shift in serial batch culture grown under a mix of two substrates (glucose-acetate) (Friesen *et al.*, 2004). Later on, different experimental setups confirmed diversification of *Escherichia coli* growing on mixtures of few substrates (Friesen *et al.*, 2004, Hall *et al.*, 2007, le Gac *et al.*, 2008). As glucose runs out, acetate is being used by cells. This opportunity for acetate 'specialists' suggests that ample chances for niche differentiation exist in

seemingly homogeneous cultures in simple media. Thus other populations might also be on the cusp of evolving complex ecologies.

These studies are consistent with the notion that nutrient mixtures (either from the beginning or created by metabolic action of a member of the population) confer a type of "spatial structure" to a culture, with very fine-scale variation (Kassen 2002, Kassen *et al*., 2004). In an industrial setting, the response of organisms to substrate mixes probably depends on the culture conditions. Sequential utilization of substrates may be more likely to occur in batch culture, whereas simultaneous utilization is likely to occur in continuous culture (Harder and Dijkhuizen 1976).

Thus, bacterial diversification in an environment with multiple resources can and will occur as a consequence of the intense competition between individual cells for the common energy sources. The emergence of specialized forms that are able to use unexplored or novel resources generates divergent populations (Mac Lean *et al*., 2005). Also, in sympatric conditions new niches may form through the action of organisms themselves.

The need for more complexity

The components of polymorphic cultures were initially characterized on the basis of growth parameters and the kinetics of uptake of specific substrates (Claasen *et al*. 1986; Helling *et al*., 1987), patterns of gene expression (Kurlandzka *et al*. 1991, Rosenzweig *et al*. 1994, Kinnersley M., 2009) and genetic changes (Adams *et al*. 1992, Herron *et al*., 2013). Thus, such studies provided insights into evolution in real time. It turns out that simplification is essential to understand the fundamental features that underlie complex processes such as adaptation. But can this strategy be applied to solve problems in the complex man-made industrial next to natural environments?

A meta-analysis of studies on diversification in sympatric conditions teaches us that one quickly reaches a limit of adaptive responses to only few energy sources. Thus elucidation of ecological and evolutionary mechanisms in complex systems is difficult

(Rainey *et al.*, 2000) and thus far experimental evolution has not addressed the whole complexity of systems. Instead, researchers have attempted to capture the essence of evolving systems in general. Thus, experimental evolution has been poor in explaining the patterns in complex systems in which many selective forces act simultaneously (Buckling *et al.*, 2009). Hence, higher environmental and ecological complexity should be incorporated into laboratory models to get a closer simulation of the complexity of nature as well as industrial bioreactors (Dettman *et al.*, 2012).

The competition for nutrients confers an overriding evolutionary pressure that is thought to shape the long-term evolution of biological networks (Gresham *et al.*, 2008).

Studies on adaptation in complex environments will undoubtedly allow an explanation of how microorganisms respond to selective pressures in this environment. The basis of stable co-existence can thus be explained. It would be also very interesting to see what similarities and differences occur between the dynamics of genome evolution in the laboratory and in industrial or natural settings. For example, genomes of bacteria that are sampled at multiple points over time in any environment can be sequenced and compared, and this work can extend to natural or even medical settings, such as the course of chronic infections (Barrick *et al.*, 2013).

Scope of this thesis

In previous work, substantial information has been obtained on the evolutionary trajectories that *Escherichia coli* may take in sequential-batch as well as continuous culture. However, most of this work was done in simple systems, and adaptation to complex systems has been largely ignored. Moreover, we now understand that large-scale industrial fermentations in complex media often work poorly and that a lack of adaptation to the established conditions may be at the basis of this poor performance. The two "worlds" of science, although both belonging to microbial ecology and physiology, have traditionally been widely apart, and the possibility that a deliberately heterogeneous culture works superior to an initially homogeneous one has been largely overlooked. To go on this path, one needs to understand the basis of adaptation of *Escherichia coli* to "conditions" (*i.e.* the medium and the oxygen supply) that reign in industrial settings, allowing the identification of different evolved types that may come about and coexist. A

close examination of such co-existence as well as the types of interactions between them that may play a role in the system will assist us to better understand the drivers of diversity within species communities within and even beyond industrial settings. Some of this information can shed light on general processes that are expected to occur in living organisms on earth, providing insight in fundamental evolutionary phenomena.

In other experimental evolution setups, selection for changes in either specific phenotypes (including growth in the presence of inducible or non-native substrates, and resistance to stresses such as antibiotics, atypical pH or temperature) or "broad" phenotypes (*e.g.* growth on preferred carbon sources or fluctuating levels of resources) have been used. However, the understanding and observation of the processes in complex environments has been very limited. Experiments can be made more complex, with multiple interacting selective forces included to provide a clearer picture of evolution. Adding complexity to the systems will reflect what bacteria really face in environments like those of industrial settings (that are characterized by complexity).

This study thus addressed the adaptation and diversification of *Escherichia coli K12*, an organism that is extensively used in large-scale fermentation processes, in a complex environment involving sequential (serial)-batch cultures. The experimental setup encompassed all the complexity of this medium, including multiple-energy sources without establishing one clear limiting resource. It involves seven bacterial populations, founded from the same ancestral genotype, propagated for ~1,000 generations in three different environments, *i.e.* (1) constant availability of oxygen, (2) lack of oxygen and (3) cycling regime of these two conditions. Serial transfers after each 24h subjected the cultures to alternating periods of "feast" (fresh substrate) and "famine" (substrate depletion), exposing them to an exponential growth phase followed by a prolonged stationary (stressful) phase. Figure S1.1 (supplementary material) resumes the experimental setup. The conditions used are thought to mimic the high degree of environmental pressure and gradients that organisms face in complex industrial settings. Thus, the importance of this work from an industrial point of view lies in assessing if maintenance of optimal performance (high metabolism of the population as-a-whole and

resistance to the global and local environmental cues) is feasible, taking into account that industrial populations are influenced by vast numbers of biotic and abiotic factors.

The scope of this work is restricted, as it "zooms in" (as a common landmark) into the end-result of selection after ~1,000 generations in the three environments applied (Figure S1.1). The fitness of (morphologically) selected forms was assessed in 1:1 competition assays against a well-conserved ancestor. Fitness values were obtained from morphologically "pre-classified" colonies as a proposed approach assuming the presence of sub-populations in the population. The outcomes and scope of selection and evolution were assessed in an integrative way in which an array of advanced molecular and phenotypic determination tools was applied (genome sequencing and transcriptomic and phenotypic studies). We aimed to address the potential for co-existence of diverse *E. coli* forms in a highly complex medium as well as their differential genomic and phenotypic characteristics, also to identify the driver(s) of such co-existence.

The strain and the growth medium

Many microorganisms have been successfully "domesticated" for use in liquid culture in laboratory cultures or in bioreactors. *Escherichia coli* is a key organism for both scientific research and industrial fermentations. There is an exceptionally large amount of information about the genetics, molecular biology, biochemistry and physiology of this organism. Owing to this availability, evolution experiments with this model species as the ancestor are rewarding in an integrative way. The ancestral strain used in the present study is *Escherichia coli K-12* strain MC1000. This strain has been genetically characterized as *araD139 Δ (araA-leu)7697, Δ (codB- LacI), galE15 galU, strA* (Casadaban *et al.*, 1980). It does not carry any transmissible plasmids and therefore its evolution can be thought of as clonal (asexual).

Furthermore, in industrial reactors, gradients in oxygen and a diverse availability of nutrients are key facets of many cultures. This work thus focused on mimicking the presence of these two factors as drivers of genomic and phenotypic responses of the populations and general performance of the cultures. In industrial systems, complex media continue to be key as these directly affect productivity as well as process

economics (Dahod, 1999). Fermentation broths normally are complex mixtures of nutrients, waste products, cells, cell debris and desired products. The ingredients of most fermentation broths including carbon sources and metabolic by-products have been found to impact yields of desired products (Parekh *et al.*, 2000). Moreover, there is a remarkable impact of the medium on the response of the populations. However, because of their complexity many substances in these media remain uncharacterized. In our study we opted for the use of Luria-Bertani (LB) broth as the complex-medium model broth, as it is widely in use in laboratory studies and provides a broad base of nutrients. LB represents a good model simulating the complexity of the environment to which bacteria are exposed in industrial settings. Furthermore, to our knowledge, this is the first study analysing the outcome of evolution of *Escherichia coli* growing in this medium. The composition of LB (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl per liter) comprises two of the most used ingredients of fermentation media (tryptone and yeast extract) (Zhang, *et al.*, 2003). These are both derived from complex biological sources. Tryptone is an enzymatic digest of casein, being rich in tryptophan, large and small peptides and amino acids, whereas yeast extract is derived from baker's/brewer's yeast through autolysis at ~50°C. It has a high content of amino acids, peptides, water-soluble vitamins, growth factors, trace elements and carbohydrates (Crueger and Crueger, 1989). Chromatographic studies have elucidated some of the components of LB broth (Supplementary material, Figure S1.2). Importantly, LB medium provides only a scant amount of carbohydrates, some of them have been detected after the metabolic action of microorganisms on the initial substrates (Hanko *et al.*, 2004, 2013). In addition, small amounts of other utilizable carbon sources (amino acids) are present in LB. Surprisingly, these characteristics make this medium quickly stressful for growing cells.

Aim of this thesis and research questions

The general aim of this thesis is to establish the scope of adaptation and the level of diversification of *Escherichia coli* developing in a heterogeneous (complex) environment. The emphasis was on the unravelling of the genetic and physiological changes that the organism undergoes, allowing it to adapt, develop diverse sub-populations and occupy different niches.

The general hypotheses underlying this work are:

(i) Well-adapted diverged forms of *Escherichia coli K12* MC1000 are formed upon long-term sequential-batch culturing in a complex environment under different oxygen regimes and genetic changes resulting in altered phenotypes are the underlying mechanisms of this diversification.

(ii) The drivers of diversification of *Escherichia coli K12* MC1000 include not only the main abiotic factor (oxygen availability) but mainly the complexity of the medium.

(iii) The number of adaptive genetic and/or phenotypic changes correlates with the level of complexity and heterogeneity of the environment.

(iv) The presence of sub-populations resulting from the selective pressures is reflected in the resulting fitness measurements.

(v) The diversification of *Escherichia coli K12* MC1000 upon long-term sequential-batch culturing in LB gives rise to co-existence of forms with enhanced capacity to utilize galactose and with enhanced tolerance to stresses.

General research questions:

From the foregoing hypotheses, the following research questions are launched:

*Is there an increase of fitness in the populations of *Escherichia coli K12* MC1000 emerging upon long-term culturing in sequential-batch LB? What is causing this enhanced fitness?

* How heterogeneous (genetically and phenotypically) will the emerged *Escherichia coli K12* MC1000 populations be?

* Is there intra-populational diversification? If so, what makes the emerged forms different from each other and from the ancestor?

* Were new niches created within the cultures, and can these novel niches be characterized? If so, is there stable coexistence between the forms that occupy these niches?

These hypotheses and general research questions are addressed in more detail in the following sections:

Chapter 2 deals with the assessment of adaption of *Escherichia coli K12* MC1000 after evolution under complex heterogeneous conditions. The emerged (evolved) forms are tested in competition with the ancestral strain. The chapter also includes a comparative and functional analysis of the types of genomic changes incurred, using the data from whole-genome sequencing of 19 selected evolved forms (from all populations) versus the ancestor. This study also includes metabolic phenotyping of all selected forms using phenotypic microarrays. The results indeed report adaptation of the selected forms to the imposed conditions and a striking level of diversification in all populations analysed. A clear adaptive response related to the consumption of galactose was found, and its relevance for growth in LB medium highlighted. **Chapter 3** describes the assessment of the transcriptome of all selected forms (versus the ancestral form) based upon the use of whole-genome microarrays. A comparative approach using the genomic information of chapter 2 was followed, in an attempt to correlate the parallel and non-parallel active (upregulated) phenotypic changes with the genetic innovations found. These transcriptomic data indicated a remarkable within-population phenotypic diversity in one of the populations. In **chapter 4**, a deeper metabolic characterization was performed in the selected population in order to confirm the diversification that had taken place under the sympatric conditions in LB. Key metabolic preferences and the kinetics of two morphologically-distinct types are unrevealed, which is suggestive of an interactive coexistence of such types. In **chapter 5** further phenotypic characteristics of the forms are studied in order to define potentially distinct ecological roles within the population. This evaluation was performed with two different physiological states using different approaches (survival under stress and quantitative (real-time) PCR). The occurrence of niche partitioning and the presence of physiological *trade-offs* were assessed and niche partitioning was confirmed as a phenomenon that had taken place in the population. The sequential occurrence of each type's favorite conditions was consistent with the affirmed stability of the coexistent pair. Finally, **chapter 6** summarizes the findings and discusses the overall results obtained in the experimental

approaches used in this study. Concluding remarks are made along with considerations on future perspectives in this challenging and interesting area.

References

Adams J, Puskas-Rozsa S., Simlar J., Wilke CM. 1992 Adaptation and major chromosomal changes in populations of *Saccharomyces cerevisiae.* Curr Genet (1992)22:13-19.

Aertsen A, Michiels CW. (2005) Diversify or die: generation of diversity in response to stress. Crit Rev Microbiol. 31(2):69-78.

Aertsen, A. and Michiels, C. (2004). Stress and how bacteria cope with death and survival. Critical Reviews in Microbiology. 30 (4):263-273.

Albert TJ, Dailidiene D, Dailide G, Norton JE, Kalia A, Richmond TA, Molla M, Singh J, Green RD, Berg DE. (2005). Mutation discovery in bacterial genomes: metronidazole resistance in *Helicobacter pylori*. Nat Methods 2: 951-953

Atsumi S,Wu TY, Machado IM, HuangWC, Chen PY, Pellegrini M, Liao JC. (2010) Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. Mol Syst Biol 6: 449

Barrett RDH, MacLean RC, Bell G. (2005). Experimental evolution of Pseudomonas fluorescens in simple and complex environments. Am. Nat. 166:470-480.

Barrick JE, Lenski R. (2013). Genome dynamics during experimental evolution. Nature Reviews Genetics 14, 827-839.

Barrick JE, Lenski RE (2009) Genome-wide mutational diversity in an evolving population of *Escherichia coli*. Cold Spring Harb Symp Quant Biol 74: 119-129

Beaumont, H. J. E., Gallie, J., Kost, K., Ferguson, G. C. & Rainey, P. B. (2009). Experimental evolution of bet hedging. Nature 462, 90-93.

Bennett A., Hughes B. (2009). Microbial experimental evolution. Am J Physiol Regulatory Integrative Comp Physiol 297:17-25.

Bennett, A.F., Lenski, R.E. (1993) Evolutionary adaptation to temperature. II. Thermal niches of experimental lines of *Escherichia coli*. Evolution 47, 1-12

Blount, Z.D., Borland, C.Z. and Lenski, R.E. (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl Acad. Sci. USA 105:7899-7906

Booth, P.J., Templer, R.H., Meijberg, W., Allen, S.J., Curran, A.R., and Lorch, M. (2002). In vitro studies of membrane protein folding. Crit. Rev. Biochem. Mol. Biol. 36 501-603

Bourke B, Sherman PM, Woodward D, Lior H, Chan VL. (1996). Pulsed-field gel electrophoresis indicates genotypic heterogeneity among *Campylobacter upsaliensis* strains. FEMS Microbiol Lett. 143(1):57-61.

Brockhurst, M. a, Colegrave, N., & Rozen, D. E. (2011). Next-generation sequencing as a tool to study microbial evolution. Molecular ecology, 20(5), 972-80. doi:10.1111/j.1365-294X.2010.04835.x

Buckling, A., Craig Maclean, R., Brockhurst, M. a, & Colegrave, N. (2009). The Beagle in a bottle. Nature, 457(7231), 824-9. doi:10.1038/nature07892

Bylund, F., Collet, E., Enfors, S-O. and Larsson, G. (1998). Substrate gradient formation in the large -scale bioreactor lowers cell yield and increases by product formation. Bioprocess engineering. 18: 171-180.

Casadaban MJ, Cohen SN.(1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138:179-207

Chang DE., Smalley DJ., Conway T. (2002). Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model Molecular Microbiology (2002) 45(2), 289-306

Chao L, Levin BR (1981) Structured habitats and the evolution of anticompetitor toxins in bacteria. Proc Natl Acad Sci USA 78: 6324Ð8.

Chung, H.J., Bang, W and Drake M.A. (2006). Comprehensive Reviews In Food Science and Food Safety Stress Response of *Escherichia coli*. Comprehensive reviews in food science and food safety. 5 : 52-56.

Claassen, P.A.M.; Kortstee, G.J.J.; Oosterveld-van Vliet, W.M.; van Neerven, A.R.W. (1986). Colonial heterogeneity of *Thiobacillus versutus*. J of Bacteriol 168 p. 791 - 794.

Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO. (2009). Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol 10: R118

Cooper TF, Lenski RE. (2010). Experimental evolution with *E. coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMCEvol. Biol. 10:11. doi:10.1186/1471-2148-10-11.

Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc. Natl Acad. Sci. USA, 100, 1072-1077.

Cooper VS, Lenski RE. (2000). The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407: 736-739. 20.

Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol, 183, 2834-2841.

Crozat, E., Winkworth, C., Gaffé, J., Hallin, P. F., Riley, M. a, Lenski, R. E., & Schneider, D. (2010). Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. Mol Biol Evol, 27(9), 2113–28. doi:10.1093/molbev/msq099

Crueger W, Crueger A. (1989). Substrates for industrial fermentation. In: Biotechnology, a textbook of industrial microbiology. Sunderland, MA: Sinauer. p 59–63

Dabros M, Schenk J, Marison I, von Stockar U. (2008). The ongoing quest for truly on-line bioprocess monitoring using spectroscopy. In: Flynne WG, editor. Biotechnol. Bioeng. New York: Nova Science Publishers, Inc.99-119.

Dahod SK. (1999). Raw material selection and medium development for industrial fermentation processes. In: Demain AL, Davies JE, editors. Manual of industrial microbiology and biotechnology, 2nd ed. Washington, DC: ASM Press. p 213–220.

Dettman JR, Rodrigue N, Melnyk H, Wong A, Bailey SF, Kassen R. 2012. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol. Ecol. 21:2058–2077.

Dhar R, Sägesser R, Weikert C, Yuan J, Wagner A. (2011) Adaptation of Saccharomyces cerevisiae to saline stress through laboratory evolution. J. Evol. Biol. 24, 1135-1153

Dionisio F, Conceicao I, Marques A, Fernandes L, Gordo I.(2005). The evolution of a conjugative plasmid and its ability to increase bacterial fitness Biol. Lett 1 250-252.

Elena SF, Lenski RE. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev 4: 457-469

Enfors S.O., Jahic, M., Rozkov, A., B, Xu., Hecker, M., JŸrgen, B., KrŸger, E., Schweder, T., Hamer, G., O'Beirne,D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., TrŠgŒrdh, C., Fuchs, L., Revstedt, J., Friberg, P.C., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., 28. Hoeks, F., Lin, HY., Neubauer, P., van der Lans, R., Luyben, K., Vrabel, P., Manelius, A. (2001). Physiological responses to mixing in large scale bioreactors. J Biotechnol. 85(2):175-85.

Enne, V. I., Delsol, a a, Roe, J. M., & Bennett, P. M. (2004). Rifampicin resistance and its fitness cost in *Enterococcus faecium*. The Journal of antimicrobial chemotherapy, 53(2), 203-7. doi:10.1093/jac/dkh044

Farrar, W., and A. Reboli. (1999). The genus Bacillus: medical. In M. Dworkin, editor. The prokaryotes: an evolving electronic resource for the microbiological community. Third edition, release 3.0. Springer-Verlag, New York, New York, USA.

Ferenci, T. (2008). Bacterial physiology, regulation and mutational adaptation in a chemostat environment. Adv. Microb. Physiol. 53, 169-229.

Foster PL. (1999) Mechanism of stationary phase mutation: a decade of adaptive mutation. Ann. Rev. Genet.;33:57-88.

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(2): 245-260.

Gonzalez CP, Hadany L, Ponder RG, Price M, Hastings PJ, Rosenberg SM (2008) Mutability and importance of a hypermutable cell subpopulation that produces stress-induced mutants in *Escherichia coli*. PLoS Genet 4:e10000208

Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. a, Ward, A., É Dunham, M. J. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS genetics, 4(12), e1000303. doi:10.1371/journal.pgen.1000303

Habets, M. G. J. L., Rozen, D. E., Hoekstra, R. F., & de Visser, J. A. G. M. (2006). The effect of population structure on the adaptive radiation of microbial populations evolving in spatially structured environments. Ecology letters, 9(9), 1041-8. doi:10.1111/j.1461-0248.2006.00955.

Hall AR, Colegrave N. (2007). How does resource supply affect evolutionary diversification. Proc R Soc B. 274:73–78.

Hanko V, Rohrer J. (2013). The Determination of Carbohydrates, alcohols and glycols in fermentation broths. Thermo Fisher Scientific, Sunnyvale, CA, USA.

Hanko, V. P., & Rohrer, J. S. (2004). Determination of amino acids in cell culture and fermentation broth media using anion-exchange chromatography with integrated pulsed amperometric detection. Anal Biochem, 324(1), 29–38. doi:10.1016/j.ab.2003.09.028

Harder, W., Dijkhuizen, U. (1976). Mixed substrate utilization in microorganisms. In: Continuous culture 6: Applications and new fields pp. 297 - 314. Chichester, Oxford: Ellis Harwood

Hardin G. (1960). The competitive exclusion principle. Science. 131(3409): 1292-7

Helling R., Vargas, CN., Adams, J. (1987). Evolution of *Escherichia coli* During Growth in a Constant Environment. Genetics, 116, 349-358.

Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO. (2006). Comparative genome sequencing of *Escherichia coli*allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38: 1406-1412

Herron, M. D. and M. Doebeli. (2013). Parallel Evolutionary Dynamics of Adaptive Diversication in *Escherichia coli*. PLoS Biol 11 (2), e1001490

Hersh, M. N., Ponder R, Hastings PJ and Rosenberg SM. (2004). Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. Res. Microbiol. 155: 352–359.

Hindre T, Knibbe C, Beslon G, Schneider D.(2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat Rev Microbiol. 2012 Mar 27;10(5):352-65.

Hughes BS, Cullum AJ, Bennett A. (2007). Evolutionary adaptation to environmental pH in experimental lineages of *Escherichia coli*. Evolution 61:1725-1734.

Jayaraman, R. (2011). Hypermutation and stress adaptation in bacteria. J. genet. 90 (2):383-91.

Kassen R, Buckling A, Bell G, Rainey PB. (2000). Diversity peaks at intermediate productivity in a laboratory microcosm. Nature 406:508-12

Kassen R, Rainey P. (2004). The ecology and genetics of microbial diversity. Annu. Rev. Microbiol. 58:207-31

Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. (2012). Experimental evolution. Trends Ecol Evol.27:547-560.

Kinnersley MA, Holben WE, Rosenzweig F. (2009).E Unibus Plurum: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli.* PLoS Genet. 5(11):e1000713. doi: 10.1371/journal.pgen.1000713.

Knight CG, Zitzmann N, Prabhakar S, Antrobus R, Dwek R, Hebestreit H, Rainey PB (2006) Unraveling adaptive evolution: how a single Molecular point mutation affects the protein coregulation network. Nat Genet 38: 1015-1022

Kolter, R., Siegele, D. A. & Tormo, A.(1993). The stationary phase of bacterial life cycle. Annu Rev Microbiol 47, 855-874.

Korona R., Nakatsu C.H., Forney L.J., Lenski R.E. (1994) Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. Proc Natl Acad. Sci. USA. 91, 9037-9041

Kurlandzka A, Rosenzweig RF, Adams J. 1991. Identification of adaptive changes in an evolving population of *Escherichia coli*: the role of changes with regulatory and highly pleiotrophic effects. Mol. Biol. Evol. 8:261–281.

Kwok PY, Chen X. (2003). Detection of single nucleotide polymorphisms. Curr. Issues Mol. Biol. 5:43-60

La Para, T.M. Zakharova,T., Nakatsu,C.H and Konopk, A. (2002). Functional and Structural Adaptations of Bacterial Communities Growing on Particulate Substrates under Stringent Nutrient Limitation. Microbial Ecology. 44(4): 317-326.

Larsson, G., Tornkvist,M., Wernersson, E.S., Tragardh, C., Noorman, H, Enfors S-O. (1996). Substrate gradients in bioreactors: origin and consequences. Bioprocess engineering. 14: 281-289

Le Gac, M., Brazas, M. D., Bertrand, M., Tyerman, J. G., Spencer, C. C., Hancock, R. E. W., & Doebeli, M. (2008). Metabolic changes associated with adaptive diversification in *Escherichia coli*. Genetics, 178(2), 1049–60. doi:10.1534/genetics.107.082040

Lee DH, Palsson BO. (2010). Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a Nonnative carbon source, L-1,2- propanediol. Appl Environ Microbiol 76: 4158-4168.

Lencastre Fernandes, R., Nierychlo, M., Lundin, L., Pedersen, A.E., Puentes Tellez, P.E., Dutta, A., Carlquist, M., Bolic, A., SchŠpper, D., Brunetti, A.C., Helmark, S., Heins, A.-L., Jensen, A.D., Nopens, I., Rottwitt, K., Szita, N., van Elsas, J.D., Nielsen, P.H., Martinussen, J., Sorensen, S.J., Lantz, A.E., Gernaey, K.V. (2011). Experimental methods and modeling techniques for description of cell population heterogeneity. Biotechnology Advances 29(6):575 - 599

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

Lenski, R. E., Mongold, J. a, Sniegowski, P. D., Travisano, M., Vasi, F., Gerrish, P. J., & Schmidt, T. M. (1998). Evolution of competitive fitness in experimental populations of E. coli: what makes one genotype a better competitor than another? Antonie van Leeuwenhoek, 73(1), 35-47.

Leroi, a M., Bennett, a F., & Lenski, R. E. (1994). Temperature acclimation and competitive fitness: an experimental test of the beneficial acclimation assumption. Proc. Natl Acad. Sci. USA, 91(5), 1917-21.

MacLean RC.(2005). Adaptive radiation in microbial microcosms. J Evol Biol. 18(6):1376-86

MacLean, R. C., Bell, G., & Rainey, P. B. (2004). The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. Proc. Natl Acad. Sci. USA, 101(21), 8072-7. doi:10.1073/pnas.0307195101

Marles-wright, J., Lewis, R. Stress responses of bacteria. (2007). Curr Opin Struct Biol. 17(6): 755-760

Miller TL, Churchill BW (1986). Substrate for large scale fermentation : In Demain AL, Davies JE (eds). Manual of Industrial Microbiology and Biotechnology ASM Press, Washington, pp. 123-136.

Moat, A., Foster JW., Spector M. (2002). Microbial physiology . Fourth edition.

Nadege Philippe,1 Estelle Crozat,1 Richard E. Lenski,2 and Dominique Schneider.(2007). Evolution of global regulatory networks during a long-term experiment with *Escherichia coli.* BioEssays 29:846-860.

Notley-McRobb, L., King, T. & Ferenci, T. (2002). *rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses. J. Bacteriol. 184, 806-811.

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. Nature Rev. Genet. 10, 531-539.

Pelosi L, Kuhn L, Guetta D *et al*. (2006) Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics, 173, 1851-1869.

Parekh S, Vinci VA, Strobel RJ. (2000). Improvement of microbial strains and fermentation processes. Appl Microbiol Biotechnol. 54(3):287-301.

Pepin, K. M., Samuel, M. a, & Wichman, H. a. (2006). Variable pleiotropic effects from mutations at the same locus hamper prediction of fitness from a fitness component. Genetics, 172(4), 2047-56.

Perfeito, L., Pereira, M. I., Campos, P. R. a, & Gordo, I. (2008). The effect of spatial structure on adaptation in *Escherichia coli.* Biology letters, 4(1), 57-9. doi:10.1098/rsbl.2007.0481

Philippe N, Crozat E, Lenski RE, Schneider D.(2007). Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. BioEssays 29:846-860.

Ponciano, J. M., La, H.-J., Joyce, P., & Forney, L. J. (2009). Evolution of diversity in spatially structured *Escherichia coli* populations. Appl Environ Microbiol 75(19), 6047-54. doi:10.1128/AEM.00063-09

Rademaker J.L.W., De Bruijn F.J. (1998). Characterization and classification of microbes by rep-PCR fingerprinting and computer-assisted pattern analysis. Appl Environ Microbiol 64:2096-2104.

Rainey P., Beaumont H., Ferguson GC, Gallie J., Kost C., Libby E., Zhang X. (2011). The evolutionary emergence of stochastic phenotype switching in bacteria. Microb Cell Fact. 10(Suppl 1):S14

Rainey, P., Buckling, A., Kassen, R., & Travisano, M. (2000). The emergence and maintenance of diversity: insights from experimental bacterial populations. Trends Ecol Evol. 15(6), 243–247

Rainey, P.B. and Travisano, M. (1998) Adaptive radiation in a heterogeneous environment. Nature. 394, 69-72.

Riehle M., Bennett, A., Lenski, R Long, A. (2003). Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature. Physiol Genomics. 14:47-58.

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.

Rousset F. (2004). Genetic Structure and Selection in Subdivided Populations. Princenton University Press.

Rozen, D. E., Habets, M. G. J. L., Handel, A., & de Visser, J. A. G. M. (2008). Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. PloS one, 3(3), e1715. doi:10.1371/journal.pone.0001715.

Rozen, D.E. and Lenski, R.E. (2000) Long-term experimental evolution in *Escherichia coli* dynamics of a balanced polymorphism. Am. Nat.155, 24-35.

Saxer G., Doebeli M., Travisano M.(2009). Spatial structure leads to ecological breakdown and loss of diversity. Proc. R. Soc. B 276, 2065-2070.

Schimel, J.,Balser, T. and Wallenstein, M. (2007). Microbial stress-response physiology and its implications for ecosystem function. Ecology. 88(6): 1386-1394.

Schweder T, Kruger E, Xu B, Jurgen B, Blomsten G, Enfors SO, Hecker M. 1999. Monitoring of genes that respond to process related stress in large scale bioprocesses. Biotech Bioeng 65:51-159.

Stewart, P., Franklin, M. (2008). Physiological heterogeneity in biofilms. Nature Reviews Microbiol. 6: 199- 210.

Suzina, N. E., A. L. Mulyukin, A. N. Kozlova, A. P. Shorokhova, V. V. Dmitriev, E. S. Barinova, O. N. Mokhova, G. I. El'-Registan, and V. I. Duda. (2004). Ultrastructure of resting cells of some non-spore-forming bacteria. Microbiology 73:435-447.

Tyerman J, Havard N, Saxer G, Travisano M, Doebeli M. (2005). Unparallel diversification in bacterial microcosms. Proc Biol Sci.272(1570):1393-8.

Versalovic J, Schneider M, de Bruijn FJ and Lupski, JR (1994) Meth. Cell. Mol. Biol. 5: 25-40.

Wang, L., Spira, B., Zhou, Z., Feng, L., Maharjan, R. P., Li, X., Ferenci, T. (2010). Divergence Involving Global Regulatory Gene Mutations in an *Escherichia coli* Population Evolving under Phosphate Limitation. Genome biology and evolution, 2, 478-87. doi:10.1093/gbe/evq035

Wu, L., Kellogg, L., Devol, A.H., Tiedje, J.M and Zhou, J. (2008). Microarray-Based Characterization of Microbial Community Functional Structure and Heterogeneity in Marine Sediments from the Gulf of Mexico. Appl Environ Microbiol. 74 (14):4516-4529.

Ye, R.W. Wao, T., Bedzik, .L, Croker, K. (2001), Applications of DNA microarrays in microbial systems, J Microbiol Methods 47 : pp. 257-272

Zambrano MM, Siegele DA, Almiron M, Tormo A & Kolter R (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. Science 259: 1757-1760.

Zhang, J., Reddy, J., Buckland, B., & Greasham, R. (2003). Toward consistent and productive complex media for industrial fermentations: studies on yeast extract for a recombinant yeast fermentation process. Biotechnol Bioeng, 82(6), 640-52. doi:10.1002/bit.10608.

Chapter 2

Adaptation and Heterogeneity of *Escherichia coli K12* MC1000 Growing in Complex Environments

Pilar Eliana Puentes-Téllez, Martin Asser Hansen,

Søren Johannes Sørensen, Jan Dirk van Elsas

Abstract

In a study aiming to assess bacterial evolution in complex growth media, we evaluated the long-term adaptive response of *Escherichia coli K12* MC1000 in Luria-Bertani (LB) medium. Seven parallel populations were founded and followed over 150 days in sequential batch cultures under three different oxygen conditions (defined environments), and 19 evolved forms were isolated. The emergence of forms with enhanced fitness was evident in competition experiments of all evolved forms versus the ancestral strain. The evolved forms were then subjected to phenotypic and genomic analyses relative to the ancestor. Profound changes were found in their phenotypes as well as whole-genome sequences. Interestingly, considerable heterogeneity was found at the intrapopulational level. However, consistently occurring parallel adaptive responses were found across all populations. The evolved forms all contained a mutation in *galR*, a repressor of the galactose operon. Concomitantly, the new forms revealed enhanced growth on galactose as well as galactose-containing disaccharides. This response was likely driven by the LB medium.

Introduction

In natural habitats, bacteria are exposed to highly dynamic and complex conditions. In contrast, in artificial environments, such as those found in fermentation systems, this dynamism is often more controlled. However, variations (in space or time) of carbon and energy resources can incite dynamism or heterogeneity, even in such artificial systems.

The complex growth media used in large-scale industrial fermentations may constitute typical examples of such heterogeneous and dynamic environments. Such growth media usually contain inexpensive complex carbon and nitrogen sources that allow good performance (Baev *et al*., 2006; Silva *et al*., 2009; Tripathi *et al*., 2009; Pasupuleti *et al*., 2010; Wang *et al*., 2011). Besides the complexity of growth media, other parameters, such as mixing efficiency and perturbation, are also understood to allow variability of local substrate concentrations, pH, temperature, and oxygen levels (Enfors *et al*., 2001). Such fluctuations may incite dynamism in the perception by the bacteria of their local environment, potentially leading to differential behavior and diversification of the organism that is being grown.

In previous work, locally divergent conditions were suggested to constitute a negative factor for process yields, as bacterial growth and effectiveness were found to be lower in heterogeneous than in homogeneous systems (Lara *et al*., 2006). However, in such studies the ability of the bacteria to adapt to environmental conditions (even when they are complex) has been disregarded. Thus, a better understanding of the adaptive response and the mechanistic causes of adaptation and diversification in complex media will contribute to an improvement of our capacities to direct population behavior. This may ultimately improve our abilities to avoid or overcome loss of effectiveness in fermentations.

The fundamental processes involved in the adaptation of bacteria to their environment have been traditionally studied using experimental evolution setups

(Nakatsu *et al*., 1998), as such experiments allow assessment of the dynamics of adaptation by natural selection within populations. Thus, the growth of a bacterial strain in a defined medium can be maintained for hundreds or thousands of generations. Adaptation can be assessed, since the fitness of evolved forms can be compared, at any time, to that of the ancestral strain (Bennett *et al*., 2009). Experimental evolution has thus enabled the study of the adaptation of bacteria to the conditions posed by structured versus unstructured habitats (Bennett *et al*., 1992; Finkel *et al*., 1999; Lenski *et al*., 1991; Rainey *et al*., 1998). The majority of the experiments have used simple limiting/selective conditions (limiting substrates), and only a few studies have included more complex conditions (Barrett *et al*., 2005), for instance, growth under fluctuations in temperature (Leroi *et al*., 1994), pH (Hughes *et al*., 2007), or with combined limited resources (Cooper *et al*., 2010). As we currently ignore how bacteria may adapt to prolonged growth in complex media, there is a need for studies that address this question (Dettman *et al*., 2012), which is of major importance to address the effects of the putative coexistence of differentially evolved forms.

Traditionally, fitness of evolved forms has been taken as the average value of a pool of randomly selected evolved forms in competitions separately against their ancestor. However, the variations among these values have often been disregarded. If such variance were considerable, this might indicate the occurrence of heterogeneous responses to the conditions imposed on the evolved populations. This might especially occur in an environment in which the complexity of resources could give space for the occupation of several niches (Kassen *et al*., 2004b). Given the elusive nature of current data on evolution in complex media, this study aims to assess the adaptation of an *Escherichia coli* strain to growth conditions offered by a complex medium in a long-term sequential batch culture setup. We used Luria-Bertani (LB) medium, in which a multiplicity of energy/carbon sources was present. We thus simulated both spatial and temporal dynamic conditions by including daily propagation into fresh medium and allowing differential oxygen availabilities, including low-oxygen (static oxygen) conditions. The use of LB medium may place limits on the study of physiological characteristics (Sezonov *et al*., 2007); however, we sought to determine if in this complex setup observable fitness diversification would occur at the level of treatment, as well as

between or within populations. We thus assessed the evolved forms—versus the ancestor strain— both phenotypically and genetically. The findings point to the occurrence of both common and divergent responses. Substantial evidence was obtained for the occurrence of a major selective genetic event likely having major consequences for the fitness of all evolved forms.

Materials and Methods

Bacterial strain and growth medium.

Escherichia coli K-12 strain MC1000 (Casadaban *et al*., 1980) is a facultatively anaerobic organism that was used in this study as the ancestral strain. This strain contains several auxotrophies, being genetically characterized as *araD*139 Δ (*araA-leu*)7697 Δ (*codB- LacI*) *galE*15 *galU strA*. Colonies of this strain on LB agar (after 19 h of incubation at 37°C) are morphologically regular, with smooth surfaces and with sizes of approximately 1.5 mm. We used the complex growth medium LB, which consisted of tryptone (10 g/liter), yeast extract (5 g/ liter), and NaCl (5 g/liter). Tryptone and yeast extract are both derived from complex sources and have been commonly used as ingredients of fermentation media (Pasupuleti *et al*., 2010). Tryptone is an enzymatic digest of casein and is rich in tryptophan, large and small peptides, and amino acids. Yeast extract is derived from baker's/brewer's yeast by autolysis at ~50°C. It has a rich content of amino acids, peptides, water-soluble vitamins, growth factors, trace elements, and diverse carbohydrates. A major carbohydrate of LB medium is the disaccharide trehalose (Hanko *et al*., 2000; Pasupuleti 2010).

The long-term evolution experiment.

Seven populations derived from one overnight culture of *E. coli* K-12 strain MC1000 were propagated under three different oxygen conditions. For this, 0.6-ml aliquots of the washed overnight cultures were introduced into a suite of (i) replicate 100-ml Erlenmeyer flasks containing 60 ml of sterile LB medium (final optical density at 600 nm [OD_{600}], 0.5 $\pm$ 0.1) and (ii) 60-ml screw-cap bottles containing 60 ml of sterile LB medium. Two populations (populations 1 and 2) were grown in Erlenmeyer flasks under shaking conditions (200 rpm, oxygen rich; treatment A), three populations (3, 4, and 5)

were grown under a cycling environment (oxygen rich, shaking [200 rpm]) in Erlenmeyer flasks, then oxygen-poor static growth in screw-cap bottles, and then returned to shaking (200 rpm) (treatment B); and two populations (6 and 7) were grown in static flasks (oxygen poor; treatment C). All treatment groups were incubated at 37°C during 24 h. Each day, after 24 h ($\pm$30 min) of growth in each environment, propagation of the then-stationary-phase cells was performed by applying a 100-fold dilution into fresh medium (final concentration of ~1x 10^7 CFU/ml). In this way, populations went through exponential and stationary phases differentially by treatment, reaching a maximal density of ~1 x10^9 CFU/ml and yielding approximately 7 generations per day. After 150 days, a total of ~1,000 generations were obtained. Flasks that had been kept under static conditions were mixed before each subsequent transfer, so that the sample would be representative for the diversity in this spatially structured environment. In order to confirm identity and possible visible genomic changes, genomic profiles (based on BOX-A1R-based repetitive extragenic palindromic-PCR [BOX-PCR] and enterobacterial repetitive intergenic consensus-PCR [ERIC-PCR]) were determined every 10 days on several randomly picked clones of each population (Versalovic *et al* 1991, 1994). Throughout all analyses, colony and cell morphologies were also analyzed visually and using microscopy, respectively, supported or not by Gram staining.

Samples taken after every ~70 generations were kept in glycerol at -80°C as backups in the event of a laboratory mishap. After ~1,000 generations (over 150 days, excluding mishaps), tubes containing each population (containing the so-called evolved forms) were placed in 20% glycerol at -80°C for further analyses.

Relative/correlated fitness responses and analysis of culture heterogeneity.
The fitness of the evolved forms was measured by competition of each selected form against the ancestor in a crossed experiment. To achieve this, selectable markers were introduced by selecting spontaneous rifampin-resistant (Rp^r) mutants of the ancestor as well as each of the evolved forms on 30 μg/ml of rifampin (Rp). Such mutants were then streaked to purity and analyzed by measuring their fitness against the unmutated isogenic form. Only equally fit forms were used in subsequent experiments. Thus, the competitiveness of each of the (marked and unmarked) evolved forms was tested against

the (unmarked and marked) ancestor (Bennett, 1996, Gagneux *et al.*, 2006). These marker-crossed experiments thus served to assess the level of fitness change versus the ancestor as a presumed result of adaptive evolution.

In detail, prior to the competition experiment, each strain was grown separately in an overnight culture under the condition (environment) for which competition would take place (acclimation phase) (Leroi *et al.*, 1994). After acclimation, the competitors were mixed in 1:1 ratios and placed in the competition environment (competition phase). Each experiment was replicated 6-fold. Over time, samples were taken and plated in duplicate onto LB agar with or without Rp supplementation to estimate the population size of each competitor. After overnight incubation of the plates, the mean numbers of colonies on the Rp-containing plates were subtracted from those on plates without rifampin. Both data sets were then used to obtain estimates of the numbers of both competitors (Enne *et al.*, 2004).

Malthusian parameters (m) for each competitor were calculated as follows: $m = \ln[(\text{density at end of competition})/(\text{density at time zero of competition})]$ (Lenski *et al.*, 1991). Relative fitness values were obtained for the evolved forms in competition with the ancestral strain under the specific environment. The relative fitness, W, of an evolved form (E) relative to the ancestor (A) was calculated as the ratio of both Malthusian parameters (Lenski *et al.*, 1991): $W = m_E/m_A$. The evolved forms from treatment B were competed, including daily transfers from shaking to static conditions, for 3 days. This way, competitors were exposed to both transitions of this treatment. Acclimation took place under aerobic conditions (Bennett, 1996).

Correlated fitness analysis (growth in alternative environments) was performed in order to examine the specificity of adaptation with respect to each treatment (presence of specialists or generalists). Thus, following the method described above, clones from each experimental line were subjected to the other two experimental treatments in competition with the ancestor (Bennett *et al.*, 1992; Hughes *et al.*, 2007; Cooper *et al.*, 2010).

Phenotypic diversity

The metabolic potentials of each selected evolved form as well as the ancestor were analyzed in duplicate on the 95 substrates of GN2 plates (Biolog Inc., Hayward, CA). The absorbance values at 600 nm were measured using a microplate reader (VersaMax microplate reader; Molecular Devices Corp.). For each strain, cells were pulled into phosphate-buffered saline (PBS; pH 6.5), and the suspension was set at an OD_{600} of 0.04. Then, 100 μl was transferred to each well of the GN2 plates, including one blank. The plates were incubated at 37°C and read after 24, 48, and 72 h. Average absorbance values were used for cluster analysis after scoring, as follows: 0, no catabolism; 1, catabolism (weak responses included). The cutoff point between 0 and 1 was an OD of 0.2 (Maharjan *et al*., 2007).

Specific results were independently confirmed in test tubes with M9 minimal medium (1x M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.01% yeast extract) supplemented or not with 0.4% of specific carbon sources. Bacterial growth was followed during 24 h, and OD_{600} values were recorded. Moreover, time zero and late-exponential-phase samples were plated on LB agar for CFU counting.

We included cell length as a phenotypic property to observe diversification. Cells of each treatment, next to the ancestor, were taken directly from the endpoint populations (60 cells per population) and observed under a microscope (Axiostar Plus HBO 50Ac; Zeiss). Cell length was further calculated by using AxionVision software v.4.8 2.0 (Carl Zeiss MicroImaging GmBH).

Comparative genome sequencing and screening for genetic diversity. The endpoint populations of each evolved strain, next to the ancestor, were streaked onto LB agar plates. The selected colonies were used to found cultures for DNA isolation using the Gnome DNA isolation kit (MP Biomedical). Following extraction, the genomic DNA was fragmented using a Bioruptor (Diagenode) for 10 cycles of 30 s on/30 s off. Then, they were converted into paired-end Illumina sequencing libraries by using the NEBNext Quick DNA sample prep master mix set 2 (NEB, Ipswich, MA), according to the manufacturer's protocol. Multiplex adaptors (MA; multiplexing sample preparation

oligonucleotide kit; Illumina, San Diego, CA) were used in the adaptor ligation. Library enrichment was carried out in two amplification steps using primers and indices described earlier (Kampmann *et al.*, 2011). Genome libraries with individual indices were pooled, to contain 10 or 11 genomes per pool, and sequenced, with one lane per pool. Sequencing was performed by 107 cycles on a HiSeq2000 instrument (Illumina), yielding an average of >20 million reads of, on average, 91 bases each (subtracting the index).

To identify all mutations in the evolved strains versus the ancestor, we used Breseq v.014 (Barrick *et al.*, 2009). This is a computational pipeline for the analysis of short-read sequencing data. Thus, we used the sequenced and annotated GenBank file of the genome of *E. coli* K-12 MG1655 (GenBank U00096.2 (Blattner, 1997)) as a reference and mapped all high-quality reads of the sequenced genomes. The quality of the reads was determined using Biopieces (http: //www.biopieces.org) and BWA (Li *et al.*, 2009). Mutations were confirmed by identification of single-nucleotide polymorphisms (SNPs) and deletion-insertion polymorphisms (DIPs) by using SAMtools (Li *et al.*, 2009b). Differences found in our ancestral strain from the reference *E. coli* K-12 MG1655 genome were considered the "genetic property" of our ancestor. In this manner, unique differences between the evolved and ancestor strains could be visualized. After obtaining the positions and descriptions of all mutations, the putative functions of the gene hits were assessed using DAVID v.6.7 (functional annotation tool) at default settings (Huang *et al.*, 2009, 2009b). The resulting function terms were further grouped by similarities to determine the enrichment (geometric mean of all modified Fisher's exact test P values) of biological processes. High enrichment scores were taken into account, but genes with a low enrichment score (<1.3) were also analyzed, as recommended by the developer.

Results and Discussion

Adaptation of evolved forms

After 150 days of daily propagation under three different oxygen conditions, we obtained cultures of evolved *E. coli K12* MC1000 that had undergone approximately 1,000 generations. The three different environments used involved cultures growing in

flasks with LB medium under constant shaking conditions (treatment A, oxygen rich), cycling between shaking and static growth (treatment B, cyclic oxygen rich and oxygen poor), and static growth (treatment C, oxygen poor).

The final populations at day 150 were diluted and plated on LB agar in order to produce single colonies representing evolved forms. Colony polymorphism was clearly observed across all evolved populations. At this point, forms were classified as (i) small colonies ("a" type), with diameters of <1 mm, (ii) large/ rough/irregular colonies ("b" type), with diameters of >1 mm, and (iii) large/smooth/regular colonies ("c" type), with diameters of >1 mm. Treatments A and B showed colony types a, b, and c in all populations. In contrast, treatment C showed only types b and c colonies. A total of 19 colonies were selected across all populations. These were subjected to all subsequent analyses of phenotype and genotype. Throughout the experiment, genomic fingerprinting analysis did confirm the identity of all colonies as *E. coli K12* MC1000, as evidenced from their characteristic patterns, i.e., similar to the ancestor (data not shown).

To assess the overall fitness of each of the evolved forms, we evaluated their fitness by competing them against the ancestor. These marker-crossed experiments revealed consistencies in the fitness results between the use of marked and unmarked evolved forms, as there were no significant differences between any set of results ($P > 0.05$). Also, marked and unmarked ancestors did not show any significantly different fitness response when competed against each other. Thus, the marker-crossed experiments allowed us to assess any fitness changes in each of the evolved forms versus the ancestor.

Table 2.1. Direct and correlated fitness responses after 150 days of culture[a].

Response type and treatment group	Condition[b]	W [c] (mean+SEM)	P value for fitness comparison between:		
			Evolved vs ancestor[d]	*Direct vs correlated*[e]	Different populations[f]
Direct fitness responses					
A		1.162 (+0.03690)	**<0.05**		**0.002**
B		1.114 (+0.00991)	**<0.001**		0.058
C		1.351(+0.03153)	**<0.001**		**0.029**
Correlated fitness responses					
A	*Cl*	1.059 (+0.02983)		**<0.05**	
A	*St*	1.073 (+0.01337)		**<0.05**	
B	*Sh*	1.129 (+0.02232)		0.696	
B	*St*	1.279 (+0.02432)		**<0.001**	
C	*Sh*	1.103 (+0.00919)		**<0.001**	
C	*Cl*	0.978 (+0.01332)		**<0.001**	

[a] Fitness results obtained after competition between the selected isolates and the ancestor. Significant differences (P<0.05) are shown in boldface.
[b] Correlated fitness competition occurred under the following conditions: Sh, Shaking; Cl, cycling,; St, static.
[c] *W* represents the mean of all the competitions for each treatment,
[d] Based on two tailed *t*-test. Ancestor with a null hypothetical value = 1.0.
[e] Comparison between correlated and direct fitness responses.
[f] Based on one-tailed *t* test performed on replica populations belonging to the same treatment group (for treatment B, a one-way ANOVA was employed).

The ancestral *E. coli K12* MC1000 strain grew and survived well in both the oxygen-rich and oxygen-limited LB cultures. However, it was outcompeted by all endpoint evolved forms in all environments (significant at $P < 0.05$) (Table 2.1). The forms that had evolved under treatments A and C revealed large fitness increases in these environments compared to the ancestor, whereas the forms from treatment B showed smaller increases. The raised fitness values observed in the more stable environments suggested that such noncycling conditions may have given better niche space or time for fitness-enhancing changes to be fixed.

Competition of evolved forms with the ancestor in the alternative environments (correlated fitness) was considered to reveal the specificity of adaptation for either the medium or oxygen conditions. The results showed that populations from both treatments A and C grown in the two alternative environments revealed relatively lower fitness increases than those in their own environment ($P < 0.05$) (Table 2.1). This suggested the occurrence of possible *trade-offs* in these forms, decreasing their relative competitiveness

enhancement in the alternative environments and increasing it under their "own" conditions. On the other hand, the treatment B forms outcompeted the ancestor under static conditions in a stronger manner than in their own environment, showing no significant difference when grown under shaking conditions ($P < 0.05$) (Table 2.1). This suggested the presence of more generalist forms.

To observe the fitness potential of the total endpoint populations versus the ancestor (and a possible existence of interactions), samples of each endpoint population (all clones together) were competed 1:1 against the ancestor under their "own" conditions. Results showed the fitness increase values to be similar to the averaged values of the individually competed forms for treatments A and C ($P > 0.05$) (data not shown). A striking difference was observed in the populations of treatment B, which revealed average fitness increases of 18% compared to the average fitness of the individual clones. This increase suggested that genotype-genotype interactions (e.g., yielding cross-feeding) may have played a role (Kassen *et al.*, 2004b). Adaptive changes in the forms evolving under treatment B might have resulted in the emergence of generalists that are able to cope with the oxygen-rich or oxygen-limited conditions and/or the transitions between these two conditions. When living together, the presence of forms that are fit under one of the conditions might diminish the impact of the change to conditions of another form (which may thus be coselected) (Wimpenny *et al.*, 1991). This contention was consistent with the positive response of forms from treatment B to the correlated fitness analysis.

Heterogeneity of fitness values within and between populations

A closer observation of the relative fitness values of the evolved forms highlighted a remarkable fitness diversity between replicate populations as well as within populations. The former comparison (Table 2.1) indicated that the average fitness gains versus the ancestor were different between the replicate populations of treatments A and C ($P < 0.05$; standard error of the mean [SEM] > 0.03), whereas they were similar for those of treatment B ($P = 0.06$). This indicated that, in each replicate population of treatments A and C, forms may have coemerged that evolved along a different path, modulating their cellular machineries in response to the environment in different ways.

For treatment B, a more coherent evolutionary path may have been followed. We then assessed the within-population variations in the fitness gains in treatment groups A through C. Nested analysis of variance (ANOVA) revealed a wide dispersion of values in all populations of treatment groups A and C (PASW-SPSS, 18.0.3; $P < 0.05$), with variance values of 0.0081 and 0.0039, respectively. In contrast, clones from treatment B showed more similar relative fitness values between each other (variance, 0.0008).

Phenotypic characteristics confirmed diversification and also paralleled adaptive responses

The cell sizes of the evolved forms versus the ancestor were determined using cell length (in μm) as a proxy. In comparison to the ancestral strain, the evolved forms from treatment groups A and B showed increased cell lengths, whereas those of treatment C revealed decreased cell lengths ($P < 0.01$). Specifically, the ancestor cells had an average length of 1.48 μm (± 0.182), whereas those of treatments A and B had lengths of 1.82 μm (± 0.251) and 1.66 μm (± 0.349), respectively. In contrast, cells of treatment group C showed about 18% decreased cell lengths relative to the ancestor (1.22 ± 0.189 μm).

We surmised that the increased cell sizes observed in treatment groups A and B might be attributable to the oxygen-rich conditions used. Possibly, the harvested evolved forms had become phenotypically well adapted to the utilization of the complex substrates of LB under oxic conditions, resulting in larger cells (Shehata *et al*., 1975). Considering treatment C, reduction of cell lengths of the evolved forms might be related to processes under these conditions, in which σ^S-controlled gene products may generate changes in the cell envelope and overall morphology (Hengge-Aronis, 2002).

For all evolved forms, the Biolog phenotypic arrays revealed weak responses to most substrates after 24 h. After 48 h, avid consumption of most of the utilizable substrates became evident, whereas after 72 h the data were confirmed. Replicate assessments were consistent, with, on average, less than a 6.3% difference between these. Expectedly, the ancestor strain metabolized 44.2% of the substrates, which was consistent with previous work with *E. coli* K-12 (King *et al*., 2004). The evolved forms of treatment

group A metabolized, on average, 63.5%, and those of treatment groups B and C metabolized 54.7% and 56.3% of the substrates, respectively. Multidimensional scaling (MDS) was performed to observe similarities among the responses of the evolved forms and compare them to the responses of the ancestor (Fig. 2.1a). All evolved forms clustered consistently per treatment (except for 2a, i.e., from population 2 of treatment A) and away from the ancestor (stress, 0.08). Interestingly, the responses of forms from treatments B and C resembled each other. These two treatments shared the oxygen limitation condition, suggesting that comparable metabolic pathways could have been affected in the evolved forms.

Then, a matrix with a total of 24 informative substrates (those that had shown differential responses of evolved forms relative to the ancestor) was constructed (Fig. 2.1b). Using this matrix, both differential responses (even within populations) and consistent ones (even across all treatments) were noted. The finding of common responses in evolved forms across all treatments (mostly metabolic upshifts) was striking. A closer examination using the level of KEGG networks revealed that there was consistency in the metabolism of carbohydrates. In particular, several early (at 24 h) upshifts were related to the galactose metabolic pathway. With shunting into this pathway, disaccharides with a galactose moiety, such as α-D-lactose, D-meliobiose, and lactulose, can be quickly consumed. This metabolic upshift was exactly what we found. Moreover, a consistent upshift in L-arabinose utilization (which provides an entrance to the pentose phosphate pathway) was also observed.

Thus, a consistent adaptive response might have been triggered by a factor common to all treatments, most likely (and particularly the sugar components of) the LB medium used. Considering this fact, we decided to assess the growth of the evolved forms and the ancestor in M9 minimal medium by using the disaccharides found as potential selective factors (Biolog phenotypic array data) as sole carbon sources (D-lactose, D-melibiose, and D-galactose). We also used L-arabinose and D-trehalose to confirm positive metabolism of the evolved forms. All results were positive; in other words, all populations inoculated at 10^3 CFU/ml reached concentrations of 10^7 CFU/ml in late exponential phase (after 9 h), whereas the ancestor was not able to grow (except in D-

trehalose, where it reached 10^6 CFU/ml after 9 h). LB medium is known to lack glucose. However, it offers several sugars (next to alcohols and organic acids), i.e., trehalose, arabinose, lactose, and maltose, in significant levels (Hanko *et al*, 2000). In fact, the level of the major sugar component of LB medium (trehalose) has been reported to be as high as 17% (Hanko *et al*., 2000, 2004; Baev *et al*., 2006). Trehalose (with an α, α-1,1-glucoside bond between two α-glucose units) may thus offer the major entrance to the glycolytic pathway for organisms growing in LB medium. Moreover, alternative pathways, such as the galactose and pentose phosphate pathways, may serve as sources of intermediates (Geanacopoulos *et al*., 1997; Kim *et al*., 2008). Our findings suggested an association of the consistent metabolic up-shifts in the evolved forms with the selection of metabolic strategies converging to glycolysis.

a)

b)

Treatment	A						B									C			
Population	1			2			3			4			5			6		7	
Substrate	1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	4c	5a	5b	5c	6b	6c	7b	7c
Dextrin																			
Glycogen																			
Tween 40																			
Tween 80																			
L-Arabinose																			
D-Galactose																			
a-D-Lactose																			
Lactulose																			
D-Melibiose																			
D-Psicose																			
Turanose																			
Acetic Acid																			
Cis-Aconitic Acid																			
Formic Acid																			
a- Hydroxybutyric Acid																			
ß- Hydroxybutyric Acid																			
a-Keto Butyric Acid																			
Propionic Acid																			
D-Saccharic Acid																			
Bromosuccinic Acid																			
Succinamic Acid																			
L-Aspartic Acid																			
L-Glutamic Acid																			
Glycyl-L- Glutamic Acid																			
Hydroxy-L- Proline																			
L-Threonine																			
Y.-Amino Butyric Acid																			

Figure 2.1. Phenotypic microarrays results (with Biolog GN2 plates). (a) Nonmetric multidimensional scaling (MDS) plot displaying similarity between all selected clones from the three treatment groups in response to growth in the 95 substrates of the GN2 Biolog plates. The position of the ancestor is shown. The plot was derived using a Bray Curtis similarity matrix, and the stress level is indicated. Two main clusters were observed (circles). (b) Differential Biolog profile (relative to the ancestor) of the selected clones from treatment groups A, B, and C. Green boxes are novel substrates metabolized after 150 days of growth under each regimen, red boxes indicate lost ability to metabolize the substrate. Boxes that are light green or light red indicate substrates with a weak response (i.e., they either changed to be consumed [green] or did not [red] in comparison with the ancestor). Blank boxes indicate no change occurred.

Additionally, these metabolic tests revealed that forms even from within the same population may have different metabolic profiles. For example, when population 2 (from treatment A) was grown in D-galactose, a remarkable difference in the response of form 2a compared to the other forms could be easily observed. Compared to forms 2b and 2c, form 2a showed delayed metabolism of D-galactose, reaching only 1.0×10^5 CFU/ml after 9 h (see Fig. S2.1,supplementary information). From the differential matrix (Fig. 2.1b), we also observed that the "a" form types from populations 1, 2, and 3 (1a, 2a, and 3a) were metabolically unique compared to the other two types in the same populations. Some heterogeneous responses were seen for substrates belonging to related pathways, e.g., the alanine, aspartate, and glutamate metabolic pathway (L-aspartic acid, L-glutamic acid, and glycyl L-glutamic acid) and pathways for several related carboxylic acids.

Genomic changes and natures of affected genes

Analyses of the BOX-PCR and ERIC-PCR profiles of all evolved forms did not show any differences from those of the ancestor. To detect genomic differences, we then performed Illumina-based sequencing of genomic DNA from each of the selected evolved forms (totaling 20 genomes, including the ancestor). Overall, the SNP analysis methods SAMtools and Breseq revealed the same mutations per evolved form. Assuming all synonymous mutations to be neutral and all nonsynonymous mutations to be deleterious, neutral, or advantageous, we calculated the ratio of the number of nonsynonymous substitutions (dN) to that of synonymous substitutions (dS) (using only coding regions). A dN/dS ratio of >1 implies that the changes incurred are positive and related to adaptation (Yang *et al*., 2000). The dN/dS ratios for treatments A and C were 2.5 and for treatment B it was 2, thus indicating that the changes found were, overall, beneficial for adaptation. Compared to other experimental evolution studies, in which whole-genome sequencing has been applied, the numbers of nonsynonymous mutations were rather high (Dettman *et al*., 2012). However, the former data were based on studies in defined media, and this is the first report in which differences in medium complexity (combined with variations in oxygen availability) were included. Interestingly, a total of 55 nonsynonymous mutations occurred in parallel in all evolved forms in all treatment groups (Fig. 2.2a). In addition, 23 mutations consistently occurred in intergenic

(presumably noncoding) regions (see Table S2.1 in the supplementary information). Parallel mutations found in independent evolving lines have been reported in long- term experiments (Wichman *et al.*, 2000; Elena *et al.*, 2003; Barrick *et al.*, 2009; Pelosi *et al.*, 2006; Woods *et al.*, 2006). These types of parallel changes have been characterized as beneficial targets of selection.

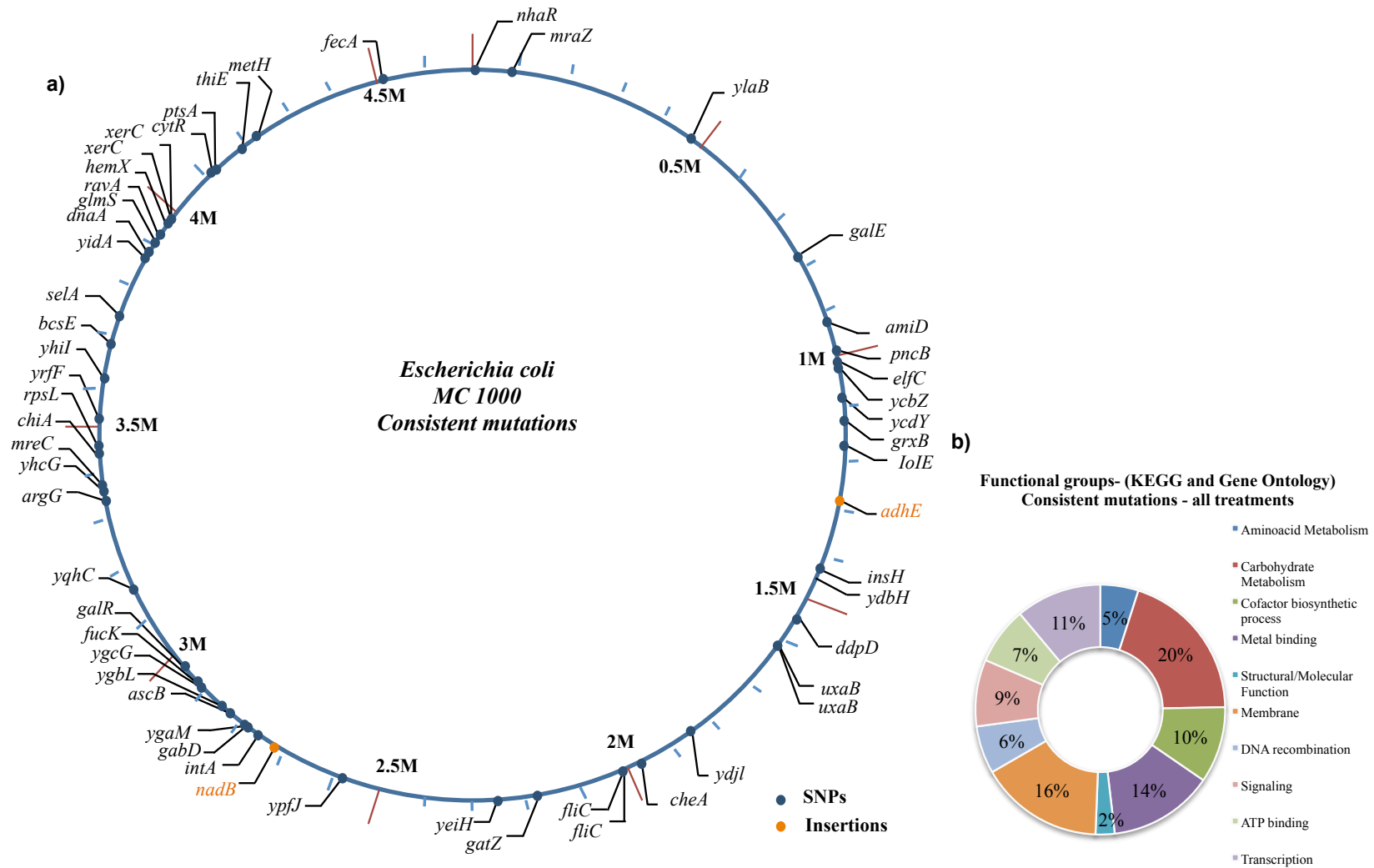**Figure 2.2**. Consistent mutations (single-nucleotide polymorphisms [SNPs] and deletion/insertion polymorphisms [DIPs]) found after the long-term experiment. (a) A total of 55 intragenic nonsynonymous mutations in all treatment groups were consistent and were distributed all around the genome. (b) Functional groups of all consistent mutations found with the DAVID v.6.7 functional annotation tool.

Bioinformatics-based results (using DAVID v.6.7) suggested that the major functional genes (functional groups) that had been consistently affected were genes involved in carbohydrate metabolism (20% of the total). Another 16% of affected genes were related to membrane processes, and 14% were related to processes involving metal binding (based on KEGG and Gene Ontology analysis [DAVID v.6.7]) (Fig. 2.2b). Some other important functional groups were also consistently affected, among which transcriptional regulator genes (i.e., *galR*, *cytR*, *dnaA*, and *ravA*). Consistency in changes of genes related to carbohydrate metabolism correlated well with the improved (and also consistent) responses to carbohydrates of the Biolog plates. Our ancestral strain was unable to utilize galactose, presumably because its genotype has a mutation in one of the *gal* operon genes (*galE15*) that results in a change of an amino acid (Ser to Phe). GalE catalyzes a hydride transfer in the interconversion of UDP-galactose and UDP-glucose as part of galactose catabolism. Interestingly, we found a consistent point mutation in our evolved forms that reverted this mutation to the original state of *E. coli* K-12, able to consume galactose as sole carbon source (Adhya *et al.*, 1969). Moreover, a *galR* point mutation located in nucleotide 2975171 of the genome (C->T), changing the codon TCT to TTT, was consistently observed. As a result, a polar amino acid (serine) was replaced with a nonpolar one (phenylalanine). GalR is the major repressor of the *gal* operon (*galETKM*). Δ *galR* mutants have been found to incite induction of the *gal* operon (3-fold higher expression than the wild type) in the presence of galactose (Tokeson *et al.*, 1991; Geanacopoulos *et al.*, 1997). The consistent *galR* mutation and the consistent metabolic upshift observed in the Biolog results provide ammunition for the hypothesis that these are the basis of the consistently enhanced fitness in all evolved forms. Apparently, *E. coli K12* MC1000 cells evolving in LB medium that have undergone a switch to derepression of the galactose operon are amenable to selection by the force exerted by disaccharides that are present in the medium.

Similarly, mutations were found in genes involved in metal binding. LB is a medium that contains very small amounts of divalent cations (Mg^{2+} and Ca^{2+}) (Wee *et al.*, 1988). Given the importance of acquisition of sufficient metals by organisms evolving in LB medium, this would clearly agree with the presumed importance of fitness-enhancing changes in genes involved in metal binding. Moreover, the changes found in

membrane-related genes might also relate to this hypothesis, as well as to the idea that cells may have evolved to facilitate the uptake of carbon/energy sources through their membranes.

Moreover, a suite of nonconsistent (not occurring in all forms) mutations was found in a nonrandom fashion (Fig. 2.3). The highest numbers of mutations (intragenic and intergenic) were found in treatments A and C, with an average 52 and 47 mutations ($P > 0.05$), respectively. In contrast, treatment B showed a lower average number (10) of mutations (treatment A versus B, $P < 0.01$; treatment A versus C, $P > 0.05$; treatment B versus C, $P > 0.05$, two-tailed t test). The numbers of mutations were found to correlate well with the levels of fitness increase. First, evolved forms from treatment B showed the lowest fitness improvements and the lowest numbers of genetic changes. Second, in evolved forms of treatment groups A and C, the numbers of nonconsistent mutations were high, whereas the raised dN/dS ratios indicated that the numbers of positively selected mutations were higher, which in an overall fashion would lead to their higher fitness.
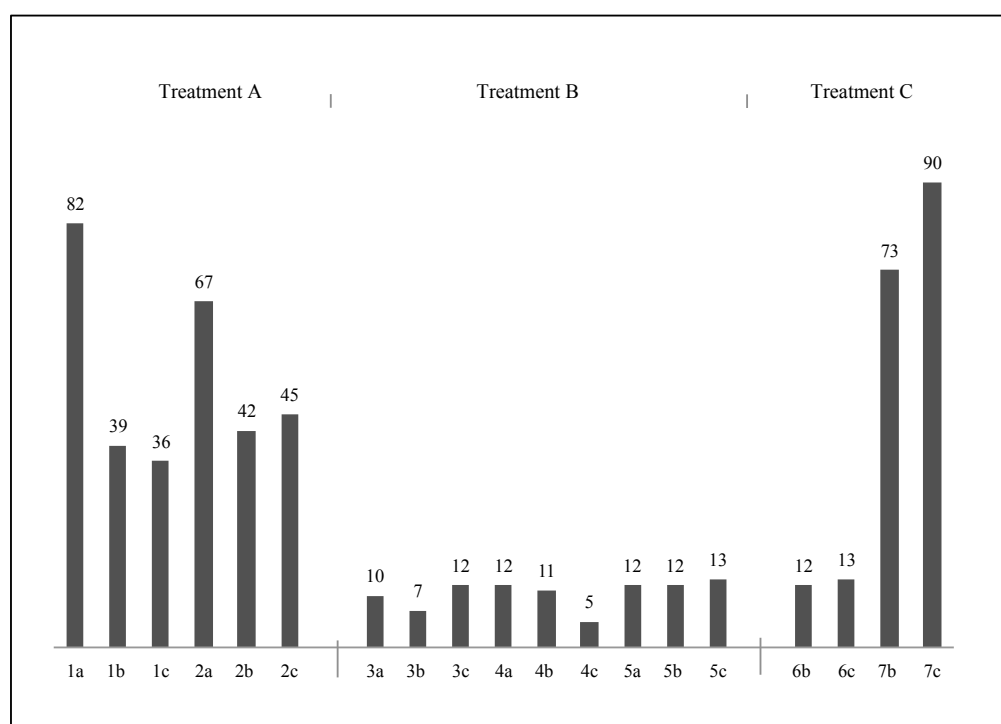
Figure 2.3. Mutations (SNPs and DIPs) nonconsistent among all populations. The total number of nonsynonymous single-nucleotide substitutions, insertions, and deletions (nonconsistent among all treatment groups) found after sequencing using Illumina Hiseq2000 are shown. Intragenic and intergenic mutations are included. There was a significant difference in the number of mutations between treatments A and B (TA versus TB, P < 0.01). There was no significant difference when treatment C was compared to the other two treatments (TA versus TC, P > 0.05; TB versus TC, P > 0.05; two-tailed t tests).

To observe genomic diversification at the population level, the numbers of mutations were compared between replicate populations. No significant differences were found in the average numbers of mutations between replicate populations of treatment groups A and B. However, the difference between the two populations of treatment C, i.e., 69 mutations, was remarkable. The environmental heterogeneity, providing a spatial structure for treatment C, may have led to higher diversification (Habets *et al*., 2006). Another finding was that unique mutations (not found in other forms of the same treatment group or other treatment groups [see Table S2.2 in the supplementary information) were observed. Forms 1a, 2a, and 7c revealed high numbers of unique mutations, i.e., 55, 56, and 21, respectively.

Functional annotation analysis of nonconsistent mutations

Results of the functional annotation analysis for all nonconsistent mutations (Fig. 2.4) showed similarities between treatments as well as treatment group particularities. First, the evolved forms of treatments A and C shared several common mutations. Strikingly, the functional group with the largest number of genes affected with mutations was the group "membrane processes". A second functional group with mutations common to these two treatments was that concerned with hydrolase and transferase proteins. Treatment A apparently incited consistent mutations in the ABC transporter genes *thkH*, *ompA*, *mppA*, *ssuB*, *ssuC*, *ompF*, *macB*, and *rbsA* across all clones. Some of these transporters are modulators of the passage of solutes like sugars, ions, and amino acids. Also, treatment A consistently showed mutations in regulatory genes involved in central metabolism (*arcA* and *arcB*). Mutations in these two-component regulatory system genes (under aerobic conditions) have been found to be responsible for upregulation of genes whose products are involved in the tricarboxylic acid cycle (Flores *et al*., 2007). Treatment group B had the lowest number of mutated genes involved in transcription regulatory processes, but it showed consistency with respect to the gene *rpoA*. This gene encodes the α -subunit of RNA polymerase and was found to carry a single-base transition, leading to a substitution of Asn for His at position rpoA294. In general, an overall comparison of functional annotations between populations of the same treatment group indicated similarities in the functions affected by mutations. Only treatment group C showed less commonality between its two populations.
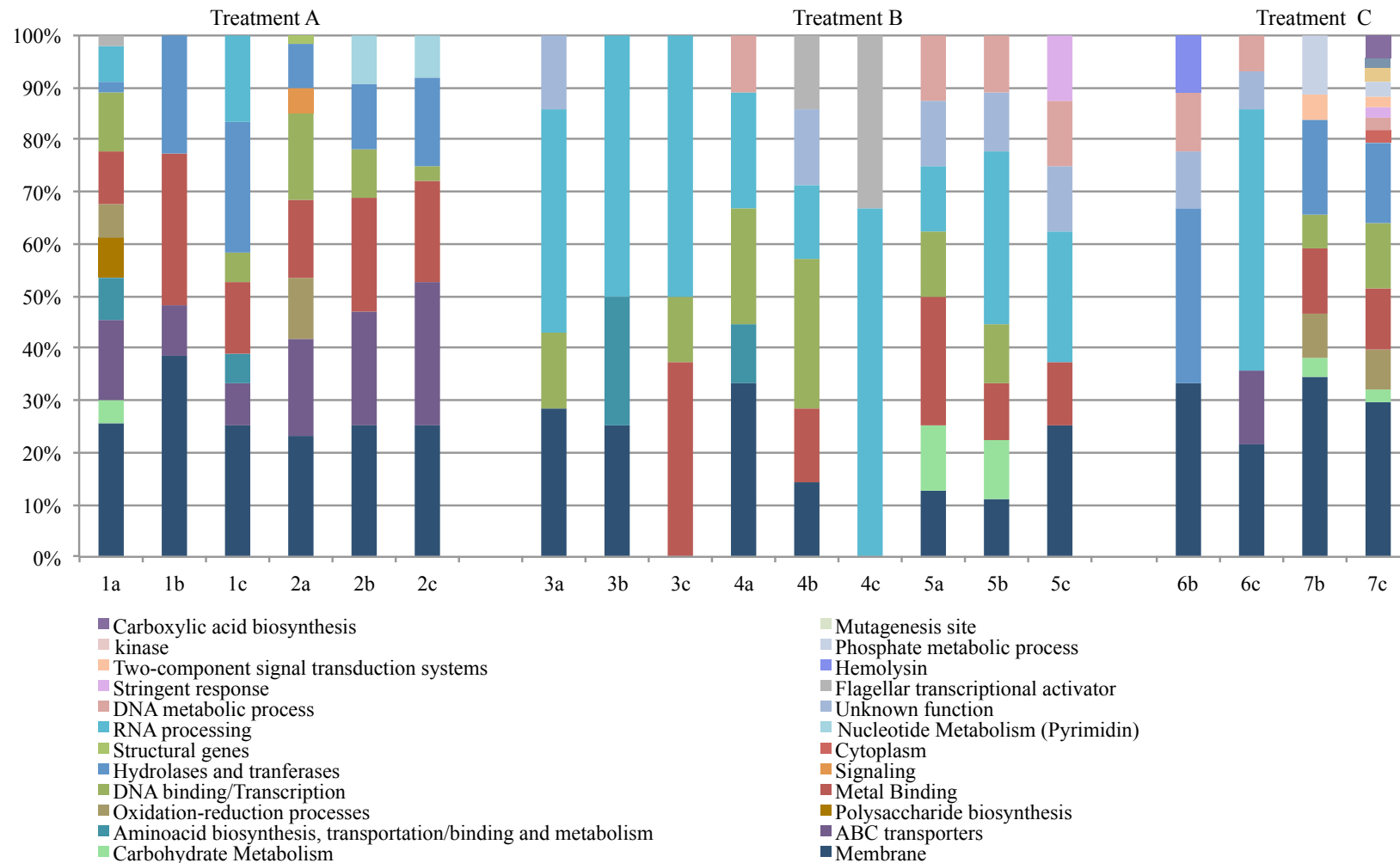
Figure 2.4. Functional groupings of the nonconsistent mutations among all population mutations. The graph shows the different functional categories (using KEGG and GO terms) of the genes affected that displayed no consistent mutations (to any treatments). The percentages correspond to the amount of mutated genes that fell into each functional category. Enrichment analysis was performed with DAVID v.6.7.

The unique mutations found in forms 1a, 2a, and 7c involved genes related to processes such as oxidation-reduction (*trxC*, *aidB*, and *narZ*), carbohydrate metabolism (*aldA*, *leuA*, *fucO*, and *tktA*), polysaccharide biosynthesis (*wzzE, rfc, ispA, rfaG, mppA,* and *plsC*), flagellar transcriptional activation, and DNA binding processes. Phenotypically, these forms appear to have different metabolic profiles compared to their coevolved partners, and these mutations and their epistatic effects may cause these particularities.

An analysis of the extent to which individual genetic variation contributes to changes in phenotype is difficult, due to possible epistatic interactions (Blaby, 2012). Conrad *et al.* in 2009 suggested that an analysis of important regulatory regions needs to be performed, and the impact of changes in such regions has to be evaluated by mutations of wild types (Conrad *et al.*, 2009). Further work will analyze such effects in our evolved forms.

Conclusions

The adaptive response to LB medium is related to selection for galactose derepression.

In this study, we evaluated the adaptation of *E. coli K12* MC1000 following evolution for ~1,000 generations along parallel evolutionary paths in a complex environment defined by medium (LB) and oxygen condition. Moreover, we described, in an integrative way, the effect of environment, i.e., the presence or absence of oxygen and shifts therein of the growth medium, on adaptation during prolonged growth of the organism. Substantial mutational and metabolic differences could be distinguished in the endpoint evolved forms, even within populations. The number of potential niches offered by the medium was likely large enough to allow such diversification. This is in contrast with previous studies performed in simpler (defined) growth media, as these studies found a lack of diversification (Schluter, 2000; Kassen, 2009). Populations that have evolved under temporal environmental variations have less space to diversify than those that have evolved in constant environments. However, possible early fixation of changes might have occurred, leading to the establishment of generalists with possible interactive

strategies and an ability to handle both oxygen conditions.

Collectively, the above phenotypic and genomic data indicate that all evolved forms modulated their metabolism toward the utilization of particular carbohydrates. A major consistent metabolic upshift occurred in the galactose utilization pathway, resulting in the enhanced utilization of particular disaccharides that feed into this pathway. It is plausible that this enhanced metabolism ultimately results in enhanced metabolic flux (shunting into glycolysis and the Krebs cycle); in other words the amount of substrate used per unit of time is increased in all evolved forms.

Previous studies of genomic changes in independent *E. coli* populations grown under different conditions already identified several consistent adaptive mutations in regulatory genes (Kurlandzka *et al*., 1991; Notley-McRobb *et al*., 1999). The positive reversion of the *galE* mutation and the mutation in the *galR* repressor might be involved in the fitness increase of our evolved forms as a key strategy to obtain carbon and energy. Clearly, unique mutations can be responsible for the variability among the populations, possibly being proxies for genetic drift.

The parallel adaptive responses suggested that, even though environment drove adaptation, the medium played the key role in the adaptive response. Thus, *E. coli* modulates its metabolism to cope with the scarcity of easily available carbon and energy sources like glucose in LB medium, and it can respond by activating alternative pathways converging to glycolysis. On the basis of our data, we have inferred that key adaptive responses to diverse nutritional sources, in this case, different types of small sugars, are major determinants of evolutionary adaptions, whereas further diversification may occur as a consequence of medium complexity.

Acknowledgments

colleagues from the Microbial Ecology Department of the University of Groningen for their input and support.

References

Adhya SL, Shapiro JA. (1969). The galactose operon of *E. coli* K-12. I. Structural and pleiotropic mutations of the operon. Genetics 62:231–247.

Baev MV, Baev D, Janeso A, Campbell JW. (2006). Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols and organic acids with transcriptional microarrays. Appl. Microbiol. Biotechnol. 71:310 –316.

Barrett RDH, MacLean RC, Bell G. (2005). Experimental evolution of *Pseudomonas fluorescens* in simple and complex environments. Am. Nat. 166:470 – 480.

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

Bennett A, Hughes BS. (2009). Microbial experimental evolution. Am. J. Physiol. Regul. Integr. Comp. Physiol. 297:17–25.

Bennett A, Lenski RE, Mittler J. (1992). Evolutionary adaptation to temperature. I. Fitness responses of *Escherichia coli* to changes in its thermal environment. Evolution 46:16 –30.

Bennett E. (1996). Evolutionary adaptation to temperature. V. Adaptation of *Escherichia coli* to a temporally varying environment. Evolution 50:493–503.

Blaby I. (2012). Experimental evolution of a facultative thermophile from a mesophilic ancestor. Appl. Environ. Microbiol. 78:144 –155.

Blattner F. (1997). The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1462.

Casadaban MJ, Cohen SN. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138:179 –207.

Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BØ. (2009). Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. 10:R118.1– R118.12.

Cooper TF, Lenski RE. (2010). Experimental evolution with *E. coli* in diverse resource

environments. I. Fluctuating environments promote divergence of replicate populations. BMC Evol. Biol. 10:11. doi:10.1186/1471-2148-10-11.

Dettman JR, Rodrigue N, Melnyk H, Wong A, Bailey SF, Kassen R. (2012). Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol. Ecol. 21:2058 –2077.

Elena SF, Lenski RE. (2003). Microbial genetics: evolution experiments with microorganisms. The dynamics and genetic bases of adaptation. Nat. Rev. Genet. 4:457– 469.

Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen Krüger BE, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Trägårdh C, Fuchs L, Revstedt J, Friberg PC, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin HY, Neubauer P, van der Lans R, Luyben K, Vrabel P, Manelius A. (2001). Physiological responses to mixing in large scale bioreactors. J. Biotechnol. 85:175–185.

Enne VI, Delsol A, Roe JM, Bennett PM. (2004). Rifampicin resistance and its fitness cost in *Enterococcus faecium*. J. Antimicrob. Chemother. 53: 203–207.

Finkel SE, Kolter R. (1999). Evolution of microbial diversity during prolonged starvation. Proc. Natl. Acad. Sci. U. S. A. 96:4023– 4027.

Flores N, Leal L, Sigala JC, de Anda R, Escalante A, Martínez A, Ramírez O. Gosset G, Bolivar F. (2007). Growth recovery on glucose under aerobic conditions of an *Escherichia coli* strain carrying a phosphoenolpyruvate:carbohydrate phosphotransferase system deletion by inactivating *arcA* and overexpressing the genes coding for glucokinase and galactose permease. J. Mol. Microbiol. Biotechnol. 13:105–116.

Gagneux S, Long CD, Small P, Van T, Schoolnik GK, Bohannan BJM. (2006). The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science 312:1944 –1946.

Geanacopoulos M, Adhya S. (1997). Functional characterization of roles of GalR and GalS as regulators of the gal regulon. J. Bacteriol. 179:228 –234.

Habets M, Rozen DE, Hoekstra RF, DE Visser JA. (2006). The effect of population structure on the adaptive radiation of microbial populations evolving in spatially structured environments. Ecol. Lett. 9:1041–1048.

Hanko VP, Rohrer JS. (2000). Determination of carbohydrates, sugar alcohols, and glycols in cell cultures and fermentation broths using high- performance anion-exchange chromatography with pulsed amperometric detection. Anal. Biochem. 283:192–199.

Hanko VP, Rohrer JS. (2004). Determination of amino acids in cell culture and

fermentation broth media using anion-exchange chromatography with integrated pulsed amperometric detection. Anal. Biochem. 324:29 –38.

Hengge-Aronis R. (2002). Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. Microbiol. Mol. Biol. Rev. 66:373–395.

Huang DW, Sherman BT, Lempicki RA. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44 –57.

Huang DW, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, Lempicki RA. (2009b). Extracting biological meaning from large gene lists with DAVID. Curr. Protoc. Bioinformatics Chapter 13:Unit 13.11.

Hughes BS, Cullum AJ, Bennett A. (2007). Evolutionary adaptation to environmental pH in experimental lineages of *Escherichia coli*. Evolution 61:1725–1734.

Kampmann M-L, Fordyce SL, Avila-Arcos MC, Rasmussen M, Willer- slev E, Nielsen LP, Gilbert MTP. (2011). A simple method for the parallel deep sequencing of full influenza A genomes. J. Virol. Methods 178:243–248.

Kassen R, Rainey PB. (2004b). The ecology and genetics of microbial diversity. Annu. Rev. Microbiol. 58:207–231.

Kassen R. (2009). Toward a general theory of adaptive radiation. The year in evolutionary biology. Ann. N. Y. Acad. Sci. 1168:3–22.

Kim BH, Gadd GM. (2008). Bacterial physiology and metabolism. Cambridge University Press, Cambridge, United Kingdom.

King T, Ishihama A, Kori A, Ferenci T. (2004). A regulatory *trade-off* as a source of strain variation in the species *Escherichia coli*. J. Bacteriol. 186: 5614 –5620.

Kurlandzka A, Rosenzweig RF, Adams J. (1991). Identification of adaptive changes in an evolving population of *Escherichia coli*: the role of changes with regulatory and highly pleiotropic effects. Mol. Biol. Evol. 8:261–281.

Lara AR, Galindo E, Ramírez OT, Palomares LA. (2006). Living with heterogeneities in bioreactor, understanding the effects of environmental gradients on cells. Mol. Biotechnol. 34:355–381.

Lenski RE, Rose MR, Simpsom SC, Tadler SC. (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315–1341.

Leroi A, Lenski RE, Bennett A. (1994). Evolutionary adaptation to temperature. III. Adaptation of *Escherichia coli* to a temporally varying. Evolution 48:1222–1229.

Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754 –1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. (2009b). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078 –2079.

Maharjan RP, Seeto S, Ferenci T. (2007). Divergence and redundancy of transport and metabolic rate-yield strategies in a single *Escherichia coli* population. J. Bacteriol. 189:2350 –2358.

Nakatsu C, Korona R, Lenski R, de Bruijn F, Marsh T, Forney L. (1998). Parallel and divergent genotypic evolution in experimental populations of *Ralstonia sp*. J. Bacteriol. 180:4325– 4331.

Notley-McRobb L, Ferenci T. (1999). Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *E. coli* populations. Environ. Microbiol. 1:33– 43.

Pasupuleti VK, Demain A. (2010). Protein hydrolysates in biotechnology. Springer Dordrecht, Heidelberg, Germany.

Pelosi L, Kuhn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D. (2006). Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics 173:1851–1869.

Rainey PB, Travisano M. (1998). Adaptive radiation in a heterogeneous environment. Nature 32:69 –72.

Schluter D. (2000). The ecology of adaptive radiation. Oxford University Press, Oxford, United Kingdom.

SezonovG, Joseleau-PetitD, D'AriR. (2007). *Escherichia coli* physiology in Luria-Bertani broth. J. Bacteriol. 189:8746 – 8749.

Shehata T, Marr A. (1975). Effect of temperature on the size of *Escherichia coli* cells. J. Bacteriol. 124:857– 862.

Silva L, Passarinha F, Sousa F, Queiroz JA, Domingues FC. (2009). Influence of growth conditions on plasmid DNA production. J. Microbiol. Biotechnol. 19:1408 –1414.

Tokeson JPE, Garges S. Adhya S. (1991). Further inducibility of a constitutive system: ultrainduction of the gal operon. J. Bacteriol. 73:2319–2327.

Tripathi N, Shrivastva A, Biswal KC, Rao PVL. (2009). Optimization of culture medium for production of recombinant dengue protein in *Escherichia coli*. Ind. Biotechnol. 5:179 –183.

Versalovic J, Koeuth T, Lupski JR. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. 19:6823– 6831.

Versalovic J, Schneider M, de Brulin FJ, Lupski JR. (1994). Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. 5:25– 40.

Wang D, Li Q, Yang M, Zhang Y, Su Z, Xing J. (2011). Efficient production of succinic acid from corn stalk hydrolysates by a recombinant *Escherichia coli* with ptsG mutation. Process Biochem. 46:365–371.

Wee S, Wilkinson B. (1988). Increased outer membrane ornithine- containing lipid and lysozyme penetrability of *Paracoccus denitrific*ans grown in a complex medium deficient in divalent cations. J. Bacteriol. 170:3283–3286.

Wichman HA, Yarber CD, Scott LA, Bull JJ. (2000). Experimental evolution recapitulates natural evolution. A remarkable study of parallel and convergent molecular evolution that is based on the whole-genome sequencing of viruses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355:1– 8.

Wimpenny JWT, Abdollahi H. (1991). Growth of mixed cultures of *Paracoccus denitrificans* and *Desulfovibrio desulfuricans* in homogeneous and in heterogeneous culture systems. Microb. Ecol. 22:1–13.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. (2006). Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 103:9107–9112.

Yang Z, Bielawski J. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15:496 –503.

Chapter 3

Comparative Genomics and Transcriptomics Analysis of Experimentally Evolved *Escherichia coli K12* MC1000 in Complex Environments

Pilar Eliana Puentes-Téllez, Ákos T. Kovács,

Oscar P. Kuipers , Jan Dirk van Elsas

Abstract

It has recently become feasible to study the basis and nature of evolutionary changes in bacteria in an experimental setting using defined media. However, assessment of adaptive changes in complex environments has been scarce. In an effort to describe the responses in such environments, we unravel, in a comparative approach, the transcriptional and genetic profiles of 19 *Escherichia coli* strains that evolved in Luria Bertani medium under three different oxygen regimes over 1000 generations. A positive relationship between upregulation of gene expression and the number of mutations was observed, suggesting that a number of metabolic pathways were activated. Phenotypic polymorphisms were observed in parallel cultures, of which some were related with mutations at the regulatory level. Non-parallel responses were observed at the intrapopulational level, which is indicative of diversifying selection. Parallel responses encompassed transcriptome diversity, and their effects were directly affected by differing genomic backgrounds. A fluctuating selective force produced higher phenotypic diversity compared with constant forces. This study demonstrates how phenotypic innovations may depend on the relationship between genomic changes and local ecological conditions. Using both comparative genomics and transcriptomics approaches, the results help elucidating various adaptive responses in cultures in unexplored complex environments.

80

Introduction

Evolution experiments that address adaptation in bacteria to conditions in controlled environments constitute excellent approaches that allow to elucidate the nature of the genetic and phenotypic adaptations that occur (Lenski *et al*., 1991; Zhang and Ferenci, 1999; Elena and Lenski, 2003; Maharjan *et al*., 2006; Velicer *et al*., 2006). In such studies, fitness, i.e. the rate at which offspring is produced, has been a key assessment criterion. Hand in hand with such fitness assessments, the key phenotypic traits and genomic consequences of the underlying adaptive processes have received scrutiny (Riehle *et al.,* 2003; Hindré *et al*., 2012).

In general, genes of interest, which are differentially expressed in an evolved form as compared with the ancestor and potentially involved in fitness increases, may be amplified for sequencing, analysis of mutations and expression (Cooper *et al.,* 2003; Elena and Lenski, 2003). However, such an approach is limited given the fact that phenotypically relevant epistatic interactions as well as novel candidate genes may be disregarded. Moreover, homogeneity of the populations under study is often assumed to occur in spite of the fact that several studies have indicated that spatially structured (Rainey and Travisano, 1998) and even unstructured environments (in which direct competition is assumed to be the key driver) (Helling *et al*., 1987; Rosenzweig *et al*., 1994; MacLean *et al*., 2005; Maharjan *et al*., 2006) can incite adaptive diversification. With respect to the latter, only few studies have thus far described such within-population diversification on a genome-wide basis (Herron and Doebeli, 2013; Puentes-Téllez *et al*., 2013).

Currently, achievable 'omics' data sets allow us to access bacterial genomes and to go from targeted, low- scale observations to broader, whole-genome perspectives, allowing to unravel associations between genotype and phenotype (Bantinaki *et al*., 2007; Crozat *et al*., 2010; Papp *et al*., 2011; Kawecki *et al*., 2012; Tenaillon *et al*., 2012). Moreover, the consequences of genomic changes during adaptive evolution can be addressed (Herring *etal*., 2006; Gresham *etal*., 2008; Brockhurst *etal*., 2011). Thus,

comparative (genomic and transcriptomic) studies of whole genomes can now be performed at several levels. The most elementary level encompasses the assessment of effects of changes in structural (coding, non-regulatory products) genes, whereas the second level addresses those of regulatory genes. However, we now understand that some genes may have mixed function.

Previous studies in this area have found that evolved strains quite often harbour just few genomic changes, whereas higher numbers of genes showed altered expression (Cooper *et al*., 2003; Fong *et al*., 2005; Gresham *et al*., 2008). These analyses might provide a proxy for behavior, adaptation and diversification under selection. Moreover, alterations in regulatory genes can result in profound phenotypic changes because of their pleiotropic effects on a number of metabolic pathways. Indeed, regulatory genes can have positive or negative effects on tens or even hundreds of genes (major regulatory genes) and adaptation is often achieved through changes in regulatory rather than structural genes (Cooper *et al*., 2003; Weber *et al*., 2005; Romero *et al*., 2012; Wang *et al*., 2012). Such high impact of regulatory genes highlights the importance of focusing on them in studies on bacterial adaptation and evolution (Philippe *et al*., 2007).

Despite considerable progress, there are several issues that delay the definition of a robust theory with respect to the link between genotype and phenotype (fitness) (Bantinaki *et al*., 2007; Orr, 2010; Papp *et al*., 2011). A first issue is the paucity of knowledge on the distribution of the effects of beneficial mutations (Pepin *et al*., 2006; Orr, 2010), a second one is that most of our understanding is based on adaptive events chosen in the light of the selected single environmental pressures. Indeed, only few studies have addressed adaptive evolution under the multiple selective pressures furnished by environmental complexity. Recently, simulations of natural environments have indicated the complex (parallel and non-parallel) genomic bases of adaptation in microorganisms (Wong *et al*., 2012). There is a perceived need to include more environmental complexity to controlled experiments and to study adaptation in a manner closer to the dynamics of natural or industrial settings in order to elucidate the potential of microorganisms to deal with more nature-like environments (Dettman *et al*., 2012).

In a recent experiment, parallel populations of *Escherichia coli K12* MC1000 were allowed to evolve during 150 days (approximately 1000 generations) in sequential batch cultures in a complex growth medium, i.e. Luria Bertani (LB) broth, under three different oxygen regimes. The oxygen regimes were aerobic (populations 1 and 2), sequentially altered aerobic-anaerobic (populations 3, 4 and 5) and anaerobic (populations 6 and 7). At the end of the study, different evolved forms were selected based on colony morphologies as follows: (i) small colonies ('a' type) with diameters < 1 mm, (ii) large/rough/irregular colonies ('b' type) with diameters > 1 mm and (iii) large/ smooth/regular colonies ('c' type) with diameters > 1 mm (Puentes-Téllez *et al.*, 2013). Remarkable numbers of genomic differences in the 19 evolved forms, as compared with the ancestor, were found (Puentes-Téllez *et al.*, 2013). Moreover, several consistent mutations were found across all treatments, at the same nucleotide position, whereas others were unique for each population or form. The first finding suggested that key forces of natural selection, resulting in adaptive evolution, were similar between all systems possibly triggered by the environmental factor shared by all systems, i.e. the growth medium (Puentes-Téllez *et al.*, 2013). In particular, a specific shared metabolic innovation, i.e. active galactose metabolism, was found. In contrast, the non-consistent mutations that were also found suggested the occurrence of diversification processes even within the respective populations.

In an attempt to understand the broad patterns of evolution of the different forms, we here assess their phenotypic (gene expression) responses to the growth medium in comparison with those of the ancestral strain. We specifically focus on the question whether a fluctuating selective force would incite the emergence of a more heterogeneous phenotypic response than a non-fluctuating selective force. We also determined phenotypic parallelism and heterogeneity within and among populations by observing the gene expression profiles, allowing to infer to what extent environmental conditions influence the phenotypic responses and potential diversification. Being aware of the current difficulties to link genotype to phenotype at a larger scale, we compared specific variations in gene expression patterns with genomic changes at the regulatory level.

Materials and Methods

The long-term evolution experiment

A long-term experiment was performed on seven bacterial populations of *E. coli K12* MC1000 that has been described in detail in our previous work (Puentes-Téllez *et al.*, 2013). Briefly, parallel populations derived from a unique overnight culture were daily propagated under three different environmental conditions in LB medium. Two populations were grown at 37°C under shaking conditions (aerobic treatment: populations 1 and 2), three under a cycling environment (shaking > static > shaking, fluctuating treatment: populations 3, 4 and 5) and two in screw-capped bottles allowing static, limited oxygen conditions (anaerobic treatment: populations 6 and 7). The composition of LB includes a complex mixture of amino acids, oligopeptides, vitamins and few sugars. After 150 days, a total of ~1000 generations was obtained (Puentes-Téllez *et al.*, 2013). Tubes containing end- point populations (evolved strains) were placed in 20% glycerol at –80°C for further analyses. In order to investigate the direct effects of each environment, a rifampicin mutant ancestor (tested marker) was competed 1:1 against morphologically different evolved forms. These were selected after platting each population on LB agar, discriminating colonies by size, shape and surface characteristics, and named as 'a', 'b' or 'c' following each population's number. Competitions took place in each evolving environment, resulting in a direct fitness comparison. Relative fitness values (*w*) of a total of 19 evolved forms were estimated by determining the end population of each competitor after competition by calculating the Malthusian parameter (*m*) of each competitor. *m* = ln[(density at end of competition)/(density at time zero of competition)]. Relative fitness values were calculated as *w* = *m* of evolved form/*m* of the ancestor.

Genome sequencing, comparative analysis and screening for mutations

All 19 evolved forms plus the ancestor strain were subjected to genomic DNA extraction procedures (in detail description: Puentes-Téllez *et al.*, 2013). We sought to identify all accumulated mutations in all these evolved strains. Thus, we used the sequenced and annotated GenBank file of *E. coli K12* MG1655 (GenBank: U00096.2) as reference and mapped high-quality reads of all sequenced genomes (including the ancestor). Mutations were confirmed by single-nucleotide polymorphism and deletion-

insertion polymorphisms calling using Breseq v.014 (Barrick *et al*., 2009) and SAMtools (Li *et al*., 2009).

DNA microarray analysis

Growth curves of all selected colony types and the ancestor were obtained by growing them under their evolved environmental conditions. For this, optical density was recorded at 600 nm using a microplate reader (VersaMax Microplate reader, Molecular Devices Corp, Sunnyvale, CA, USA). Cells were harvested for the microarray experiment in a late logarithmic growth phase [between optical density (OD) 0.6 and 0.7] from cultures that have been inoculated at low density. Two biological replicates per colony type and also of the ancestor were used during all the microarrays experiments. Pellets harvested by centrifugation (10.300 *g*, 1 min) were immediately used for RNA isolation with NucleoSpin RNA II isolation kit (Macherey-Nagel, Biokè, Leiden, the Netherlands). RNA concentrations and purity were assessed using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Rockford, IL, USA). RNA samples were reverse- transcribed into cDNA with the Superscript III reverse transcriptase kit (Invitrogen, Carlsbad, CA, USA) and labelled with Dylight Amine-Reactive Dyes Dylight-535 and Dylight- 635 (Thermo Scientific, Rockford, IL, USA). Purified and labelled cDNA samples (NucleoSpin Gel and PCR Clean-up kit, Biokè) were hybridized in a dye-swap design using the Agilent Gene Expression Hybridization Kit (Agilent Technologies, Santa Clara, CA, USA) at 65°C for 17 h to genome-wide multistrain *E. coli* (8x15K) microarrays of 60-mer oligonucleotides (Agilent Technologies). After hybridization, slides were washed for 5 min in 6× SSC and 0.005% Triton X-102, and 10 min in 0.1× SSC, 0.005% Triton X-102, air-dried and scanned with a GenePix 4200AL confocal laser scanner (Molecular Devices, Union City, CA, USA). Determination of the individual intensities was performed with GenePix Pro 6.1 (Molecular Devices). Statistical analysis was performed using R Limma package (Linear Models for Microarrays Data) (Smyth, 2004). Multiple-gene probes results were merged and normalized using the MA table conversion tool available on the MOLGEN Bioinformatics Server (http://server.molgenrug.nl/) (Molecular Genetics, University of Groningen, Groningen, the Netherlands). Normalization of the data, calculation of *P* values and examination of absolute gene expression numbers reduced possible noise

caused by the backgrounds and diminished bias affecting the power of the data. Genes with a *P* value of <0.05 and minimum of twofold differentially expressed compared with the ancestor were considered significantly changed. The data set has been deposited in GEO Database (Accession number GSE44614).

Quantitative RT-qPCR

A total of seven genes with both large and subtle expression changes were selected in order to confirm the results of the microarray gene expression data across all samples using RT-qPCR. The RNA samples described earlier were treated with DNase I (Thermo Scientific) at 37°C for 1h and repurified using NucleoSpin RNA II isolation kit (Macherey- Nagel). Briefly, 2 µg of total RNA was mixed with 2 ul of random hexa/nonamers and kept during 5 min at 70°C. Two microlitres of 2.5 mM of deoxynucleotide triphosphate (dNTP), 4 µl of the buffer and 0.5 µl of the Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV RT) enzyme (Thermo Scientific, Waltham, MA, USA) were added to the initial mix, and final volume was adjusted with Diethylpyrocarbonate (DEPC) treated water to 20µl. The final mix was taken to 25°C × 10 min, 42°C for 60 min and 70°C for 10 min using a thermocycler. The qPCR reaction mixture was prepared adding 10 µl of Thermo Fisher Maxima SYBR Green qPCR Master Mix to 1 µl of cDNA containing 2.0 µl of each primer (1.5 µM) and 5 µl of water. Mixtures were cycled using Bio- Rad iQ5 (Bio-Rad Laboratories, Richmond, CA, USA). Relative amounts of each transcript were calculated using the $2\text{-}\Delta\Delta\text{CT}$ method (Schmittgen and Livak, 2008). The efficiency between the control gene (*rpoA*) and tested genes was measured and found to be less than 10% of each other. Primers were designed using the default parameters of Clone Manager Suite (Sci-Ed Software, Durham, NC, USA). The sequences of the primers used are shown in Table S3.1.

Auto-aggregation assay

Overnight cultures of the strains in LB medium were adjusted to an OD600 of 0.05, a volume of 0.8 µl of each adjusted concentration was inoculated to 10 ml of LB in a sterile 20 ml tube. All cultures were vigorously shaken for 10 s and incubated statically at 37°C. When positive, auto-aggregation was observed at the bottom of the tubes. A negative result was observed when there was lack of sediment and higher turbidity in the

culture.

Functional annotation

We used the String 9.0 database to look for known functional interactions of the upregulated genes (http://string-db.org/). Following the default settings, we found the significant KEGG pathways and GO biological processes terms represented by the lists of genes. *P*-value scores correspond to strength of the enrichment analysis (translated in probability) of finding the genes within the same pathway or process (Jensen *et al*., 2009).

Results and Discussion

In order to observe the scope of the adaptive responses of *E. coli K12* MC1000 in LB medium, we used a total of seven independent populations that had previously evolved under three oxygen availability regimes, i.e. aerobic, fluctuating (aerobic/anaerobic) and anaerobic conditions. We previously evaluated the fitness levels and genomic differences of forms selected from these seven populations (Puentes-Téllez *et al*., 2013). Here, we evaluated the gene expression patterns of all 19 clones derived from a same original strain (selected from all seven end-point populations) in comparison with that of their ancestor. We performed whole-genome microarray analyses to detect differential gene expression between all 19 evolved forms relative to the ancestor. Late logarithmic phase cultures obtained after growing each form in their evolving environment were harvested to perform RNA extraction, as late log phase is a stage at which cells consumed the primary energy sources of LB medium. In fact, growth rates have been suggested to decrease by 40%, and cells may start to consume multiple 'second tier' substrates (Baev *et al*., 2006a,b,c). This would allow us to observe differential preferences for alternative energy sources or to observe the differential mechanisms populations use to deal with the adversities of this growth phase. For a selected set of genes, the microarray gene expression data were confirmed using RT-qPCR. This analysis showed that reverse transcription polymerase chain reaction (RT-qPCR) coincided strongly with the transcriptional changes observed by the microarray results, confirming the transcript measurements obtained with the Agilent slides (data not shown, for the list of genes, see Table S3.1.

Upregulated and downregulated transcript levels of evolved forms relative to the ancestor

Differential transcriptional responses were analysed using the genes with the most significant alterations in expression versus the ancestor (ranked by *P*-value). After normalization of the data, we selected the first 150 genes per strain of this list. This top 150 list (including upregulated and downregulated genes) allowed an integrated initial analysis of the differences observed. Fig. 3.1a shows the upregulated and downregulated genes. Noteworthy is the fact that differences were observed between forms even within the populations. Particularly, two forms of the aerobic and anaerobic treatments (2a and 7c, respectively) showed a remarkable percentage of down- regulated genes compared with the rest of their coevolving partners. Despite this particularity, high percentages of upregulated genes were obtained in the aerobic and anaerobic treatments (on average, about 100/150 and 91/150 genes), whereas the fluctuating treatment revealed on average 67/150 upregulated genes. The oxygen availability shifts in the fluctuating treatment may thus have incited less prevailing metabolic drivers compared with the other two treatments, resulting in gene expression patterns being 'more similar' to the ancestor (in terms of activation of metabolic processes) and different from the forms of the constant aerobic and anaerobic treatments. Downregulation in this treatment suggests the shutting down of perhaps now obsolete pathways relative to the ancestor. Earlier experimental evolution experiments have demonstrated that adaptation to novel conditions can involve phenotypic innovations, i.e the emergence of novel traits (either earning or losing capabilities) in the evolved strains relative to the ancestor (Bennett *etal*., 1992; Lenski and Bennett, 1993; McKenzie *et al*., 2000; Cooper *et al*., 2003; Riehle *et al*., 2003; Hughes *et al*., 2007; Lee and Palsson, 2010). Some innovations may have appeared relatively early in the experimental time (within few hundreds of generations) and some of them later. The extent to which phenotypic differences show up in these experiments might depend on the number of available niches in the environment, as well as on the point in time at which they are measured. It has recently been indicated that often adaptive evolution in experimental populations of microbes involves the fine- tuning of the entire transcriptional program of the cell (Dettman *et al*., 2012). Because of the complexity of our system, we expected the occurrence of considerable differences in gene

expression in the evolved forms relative to the ancestor. However, from a global perspective, the number of genes with significantly different expression changes did not exceed 10% of the total gene count (on the basis of strain K12, our *E. coli* ancestor contained 4414 predicted genes). Interestingly, although the top list thus covered only up to about 4% of the total genome (150 genes), treatments could be differentiated with this initial analysis. This indicated that the parameters of fold- change and the significance were sufficiently informative. Moreover, although the link between genomic and gene expression changes is far from resolved, we examined the putative link between these, comparing the number of upregulated and downregulated genes (top 150 list) with the total numbers of non-synonymous mutations by treatment. We observed a similarly shaped pattern when considering the numbers of upregulated genes (aerobic > fluctuating < anaerobic) and the numbers of mutations (aerobic > fluctuating < anaerobic) (Fig. 3.1b) in comparison with the one of downregulated genes (aerobic < fluctuating > anaerobic). This finding suggests an influence of these mutations on active metabolism perhaps because higher numbers of non-synonymous genomic changes can cause higher numbers of pleiotropic effects.

A transcriptional change marks the entrance of a cell into a different physiological state. This change is often defined by the nutrient status of the medium and can then be characterized by specific patterns of substrate utilization (Baev *et al*., 2006a). In our complex medium, the patterns of substrate utilization may have been very diverse in all populations compared with the ancestor and the nature of these responses was clearly defined by the type of oxygen regime.
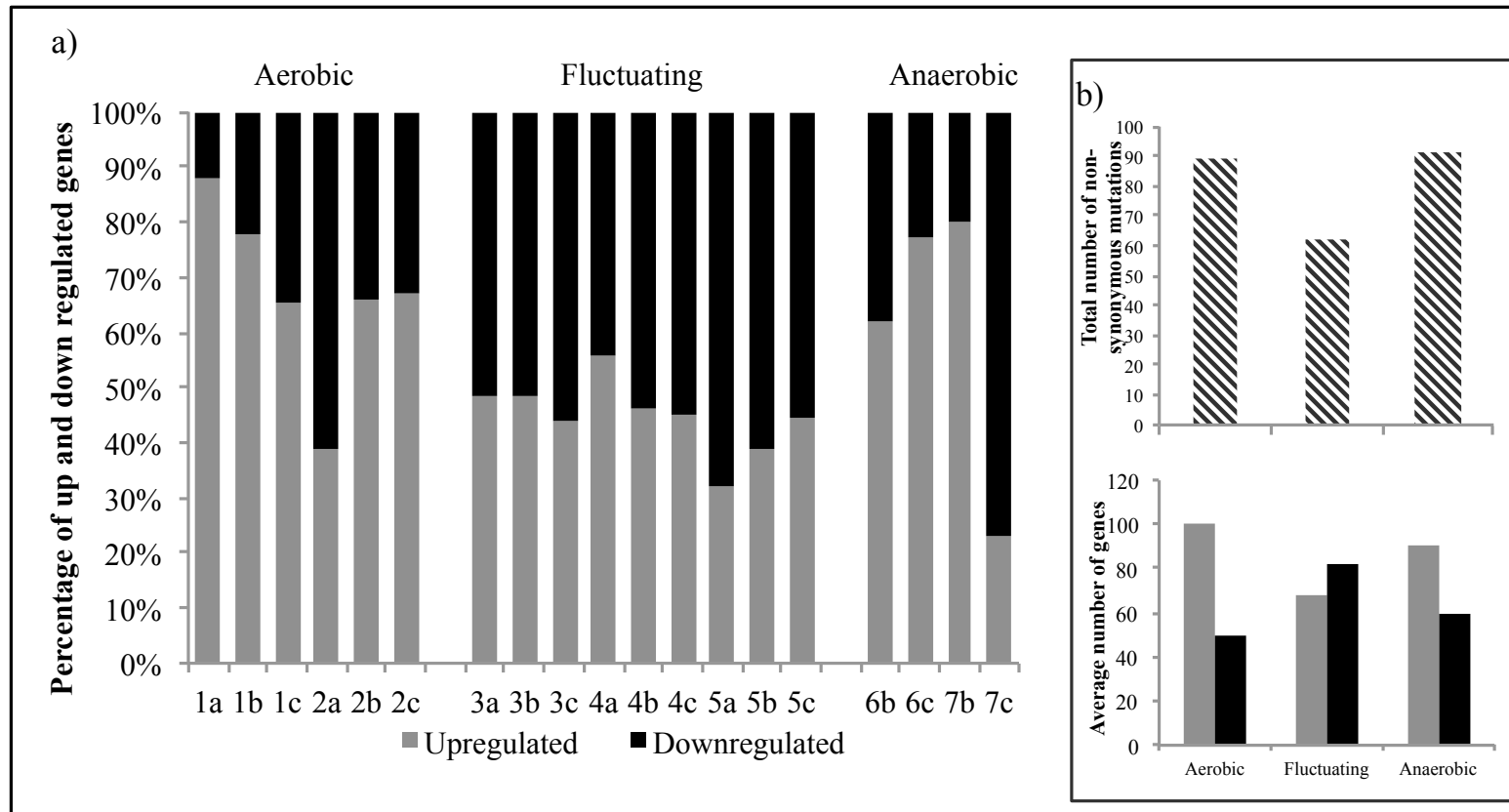
Figure 3.1. Upregulation and downregulation relative to the ancestor in the top 150 significantly changed genes.

a). Percentage of upregulated and downregulated genes by form across all treatments. Numbers in the X axis represent the populations (1–7) for all treatments and letters the different morphologies: (i) small colonies ('a' type) with diameters < 1 mm, (ii) large/rough/irregular colonies ('b' type) with diameters > 1 mm and (iii) large/smooth/regular colonies ('c' type) with diameters > 1 mm.

b). Above, total number of non-synonymous (intragenic) mutations by treatment. Below, total number of upregulated and downregulated genes from the top 150 genes list by treatment. The number of mutations resembles the upregulated gene expression profile of the three treatments.

Differential gene expression indicates the presence of parallel and non-parallel responses

In order to focus on the most significant differences among all the forms and reduce noise, we established a sublist of 50 genes per evolved form from our list. To pinpoint and examine candidate genes that may have remarkably contributed to the adaptive evolution and have participated in the active metabolic innovations (upregulation), we sought to cluster genes of which the expression had changed in parallel or not. For this purpose, a matrix was constructed with only the upregulated genes, discriminating their presence or absence by form. From this cluster analysis, we could observe the genes that are involved in diversification at the interpopulational and intrapopulational level, next to the consistent changes and key variables explaining differences in our multidimensional data sets.

Fig. 3.2 shows a principal component analysis (PCA) (Canoco v.4.5, Lepš, 2003) with the expression profiles of the 19 evolved forms, in which principal components 1 (PC1) and 2 (PC2) explain (in total) 21% of the variance. There was no clear correspondence between the gene expression results and the different colony morphologies and populations. The first component of the PCA discriminated the data in two parts according to the environmental conditions. The oxygen-limited conditions (found in the fluctuating and anaerobic treatments) possibly incited more similar expression patterns than the aerobic conditions. On the other hand, data points along axis PC2 roughly separated the anaerobic treatment forms from the other two treatments. The evolved forms from the fluctuating and anaerobic treatments appeared to be more heterogeneous in terms of gene expression than those of the aerobic treatment (except for two outliers: forms 2a and 2b).

Although the aforementioned PCA (with supplementary information in Table S3.3) only explained a low percentage of the total variance, it confirmed that environment strongly affected the adaptive response to LB medium relative to the ancestor. In order to analyse the changes in expression levels in detail, we first examined the responses in the main 'clouds' observed in the PCA (mainly represented by treatment). We screened the

statistically overrepresented functional categories of the genes contributing to the main clusters (treatments) and to those appearing isolated from their treatment using String 9.0. Table 3.1 shows the functional impact of the upregulated individual genes clustered in the significant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and the main gene ontology (GO) biological processes involved.

Figure 3.2. Principal component analysis of upregulated genes obtained in the 50 genes top list. (■) Aerobic treatment forms, (▲) fluctuating treatment forms, (●) anaerobic treatment forms. Shady ellipses show the main clusters. Numbers in the graph correspond to the genes listed in supplementary material (Table S3.3).

The major group of upregulated genes in the aerobic treatment cloud clustered in the category 'oxidative phosphorylation' containing genes belonging to the aerobic respiratory chain of *E. coli*. For example, several genes from the *cyoABCDE* operon, encoding the cytochrome *bo*3, one of two main terminal oxidases in the respiratory chain (Cotter *et al*., 1990), were upregulated up to fivefold. Essential genes (belonging to this cluster) for energy transduction and proton translocation (*nuoM*, *nuoG* and *nuoE*) were also differentially upregulated. Finally, genes from the *atpIBEFHAGDC* operon, which catalyses the synthesis of Adenosine-5′- triphosphate (ATP) under aerobic cell growth (Kasimoglu *et al*., 1996), were also overexpressed. The expression of all these genes was previously found to be higher when tricarboxylic acid (TCA) cycle intermediates, such as fumarate, malate and succinate, were used or when acetate and pyruvate, the metabolic precursor of the cycle, were used (Kasimoglu *et al*., 1996). This correlates nicely with the second and third KEGG pathway from the annotation analysis, which involves genes involved in pyruvate metabolism and the TCA cycle. This analysis suggests that in the aerobic treatment forms, there was a more active regulation of processes related to cellular respiration/ central metabolism during aerobic respiration and energy derivation by oxidation of organic compounds compared with the ancestor.

Even though they appeared to be quite distant in the PCA, there were interesting commonalities between the aerobic and fluctuating treatments. The list of most upregulated genes of both treatments contained the *yahO* and *yqhD* genes (up to 50-fold). The *yahO* gene encodes the σ^S subunit of RNA polymerase. σ^S is the product of *rpoS*, which controls the expression of genes responding to starvation and stress (Lacour and Landini, 2004). Thus, σ^S is a major regulator of genes required for adaptation to stationary phase. However, recent studies discovered important physiological roles of RpoS in growing cells, regulating genes that are involved in oxidative stress responses as a consequence of aerobic metabolism (Dong *et al*., 2008). Also, overexpression of *yqhD* leads to increased resistance to reactive oxygen-generating compounds (Pérez *et al*., 2008). Notoriously, all *yqhD* expression levels in the aerobic treatment were >20-fold enhanced, except for form 2a, which revealed only a 6.7-fold enhancement relative to the ancestor. This difference suggests a particular physiological behavior of the latter form.

The aerobic and fluctuating treatments revealed another outstanding commonality, i.e. the expression of the *flu* gene was upregulated to up to 30-fold in most forms from these treatments. In contrast, populations of the anaerobic treatment showed no difference in the *flu* gene expression compared with the ancestor (Fig. 3.3a), and so this phenotypic response appears to be triggered by the selective force prevalent in the oxygen-receiving environments. The *flu* gene encodes an outer membrane protein with a molecular mechanism that controls expression by phase variation, antigen 43 (Ag43). The Ag43-encoding gene was identified as a result of the auto-aggregative property it confers on the bacterial population and consequent biofilm formation (Henderson *et al*., 1997; Kjaergaard *et al*., 2002). The benefits and relevance of this protein for a bacterial cell population are not well defined (Van der Woude and Henderson, 2008) but could lie in particular stress conditions in the liquid oxygen-receiving systems in which living in an aggregate confers an ecological advantage.

a)

Treatment	Population	Differential auto-aggregation within the populations
Aerobic	1	+
	2	+
Fluctuating	3	+
	4	+
	5	+
Anaerobic*	6	-
	7	-
Ancestor	-	-

b)
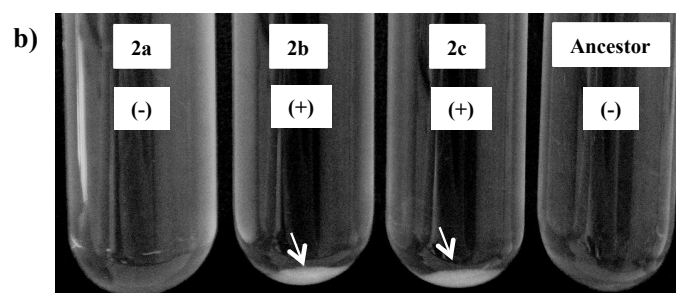
Figure 3.3. a). Populations with forms performing auto-aggregation after 18 h of growth in a static LB medium. b). Differential auto-aggregation in population 2 (forms 2a, 2b and 2c of the aerobic treatment), negative auto-aggregation of the ancestor is shown. (+) Positive auto-aggregation; (−) negative auto-aggregation. *Anaerobic treatment forms do not auto-aggregate.

Analysis of the KEGG pathways to which the expressed genes of the fluctuating treatment forms belonged revealed a low probability to find genes in the same pathway (Table 3.1) representing a phenotypically diverse response ($P > 0.05$). Moreover, the PCA analysis showed that – even though all forms clustered together as a treatment – particular differences were still detectable. The String 9.0 analysis revealed that the highest number of upregulated genes encoded ABC transporter systems. Such genes ease the translocation of substrates like vitamins (*thiP* gene in our results), dipeptides (*ddpC*), oligopeptides (*oppF*, *oppA*, *oppB*) and defined amino acids *(argT)* through the membrane. Our data set revealed these genes to cluster significantly in the biological processes 'ion transport' group ($P < 0.05$). The fluctuating treatment forms further differentially expressed genes related to amino acid biosynthesis (*pheA*, *aroF*, *hisC*, *trpB*, *ubiB*, *ubiD*) and metabolism (*carA*, *nadB*, *ansB*, *nagK*, *yagH*, *ptsG*, *murA*). Thus, we observed that the upregulated genes of the fluctuating treatment were – in general – involved in processes related to the uptake of diverse substrates and the maintenance of an internal steady state of ions (homeostasis) by biosynthetic processes. Fluctuating treatment forms seem to be phenotypically more polymorphic ($P > 0.05$), and this might have been triggered by the particular dynamism of their fluctuating environment.

Table 3.1. Functional annotation by treatment of the top 50 significant upregulated genes.

	KEGG Pathways	Number of genes	*P-value*	Biological processes	Number of genes	*P-value*
Aerobic	Oxidative phosphorylation	14	0,001	Oxidation-reduction Process	33	0,001
	Pyruvate metabolism	10	0,001	Generation of precursor metabolites and energy	21	0,001
	Citrate cycle (TCA cycle)	8	0,001	compounds	19	0,001
	Ribosome	9	0,002	Cellular respiration	17	0,001
	Glyoxylate and dicarboxylate metabolism	6	0,010	Proton transport	9	0,001
	Nitrogen metabolism	5	0,033			
	Glycolysis / Gluconeogenesis	5	0,050			
Fluctuating	Phenylalanine, tyrosine and tryptophan biosynthesis	4	0,034	Ion transport	17	0,022
	Ubiquinone and other terpenoid-quinone biosynthesis	3	0,079	Cellular homeostasis	7	0,021
	Alanine, aspartate and glutamate metabolism	4	0,094	Anion transport	7	0,029
	ABC transporters	15	0,113	Response to inorganic substance	7	0,046
	Sulfur metabolism	2	0,184	Coenzyme biosynthetic process	7	0,059
	Aminoacyl-tRNA biosynthesis	3	0,206			
	beta-Alanine metabolism	2	0,229			
	Lysine biosynthesis	2	0,253			
	Amino sugar and nucleotide sugar metabolism	4	0,275			
	Nicotinate and nicotinamide metabolism	2	0,276			
	Phenylalanine metabolism	2	0,299			
	Nitrogen metabolism	3	0,357			
Anaerobic	Phosphotransferase system (PTS)	7	0,001	Response to stimulus	32	0,001
	Glycine, serine and threonine metabolism	4	0,011	Transcription, DNA-dependent	15	0,001
	Starch and sucrose metabolism	4	0,011	Response to stress	21	0,001
	Fructose and mannose metabolism	4	0,021	Single-species biofilm formation	6	0,001
	Glycolysis / Gluconeogenesis	4	0,023	Response to pH	5	0,001
	Citrate cycle (TCA cycle)	3	0,043			
	Alanine, aspartate and glutamate metabolism	3	0,047			

The functional annotation analysis corresponds to the genes clustered in the PCA. P-value is the result of a gene ontology enrichment analysis representing the strength of the statistic (translated in probability) to find genes in the same pathway. The fluctuating treatment has only one significant cluster in the KEGG pathway annotation; however, although not significant, the list of first 11 clusters (based on P-value) are shown for information demonstrating a heterogeneous upshift of metabolic pathways in these forms. The first five significant GO Biological processes are shown based on number of clustered genes. The genes that appear as 'unique' and are responsible for isolating the outliers in the aerobic and anaerobic treatments (2a, 2b and 7b) have been excluded and analyzed separately.

On the other hand, in our previous work, we found that even though the fluctuating treatment forms outcompeted the ancestor strain when growing together, they showed the lowest (averaged) fitness increase ($w = 1.114$) compared with the forms of aerobic and anaerobic treatments ($w = 1.162$ and $w = 1.351$, respectively, Fig. 3.4) (Puentes-Téllez *et al.*, 2013). Previous studies have demonstrated that the uptake of free amino acids and peptides by *E. coli* growing in LB medium takes mainly place in the early stages (before 3.5 h) of growth. After these have been taken up and consumed, biosynthesis of amino acids and nucleotides takes place, next to other biosynthetic processes, decreasing growth rate (Baev *et al.*, 2006b). This indicates that the enhanced investment by the fluctuating treatment forms in such biosynthesis is costly to the cell.
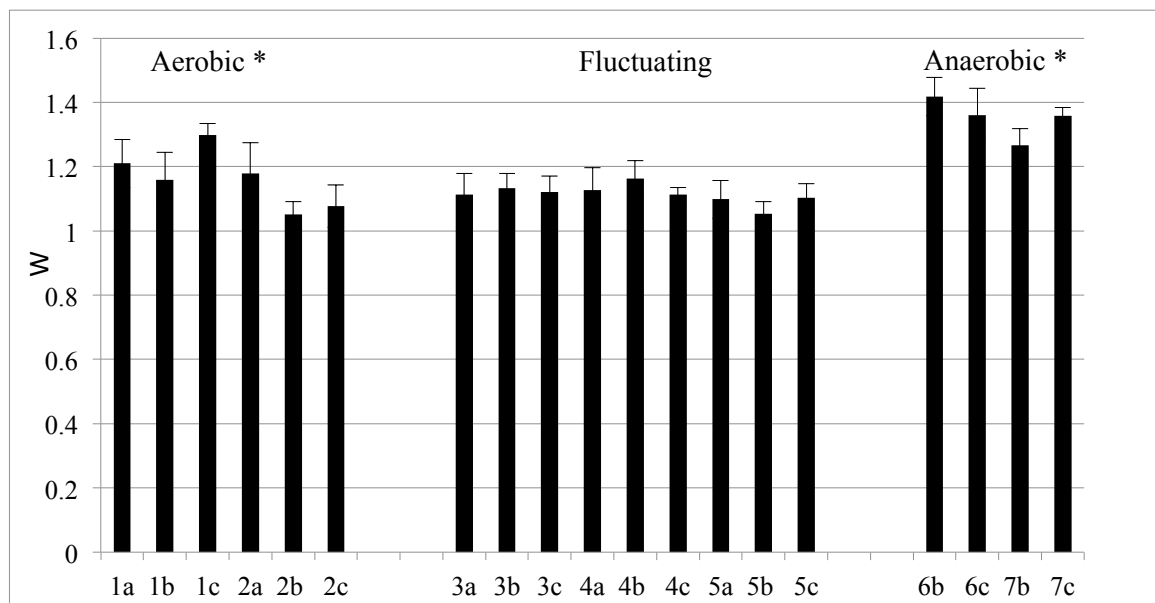
Figure 3.4. Direct fitness response. Direct fitness response relative to the ancestor of all the colony types in all treatments after 150 days of long-term experiment. *Treatments with significantly different fitness between populations ($P < 0.001$).

Functional analyses of the anaerobic treatment forms revealed that the main group of genes with altered (upregulated) gene expression was related to carbohydrate-active transport systems. Genes involved in phosphorylation of incoming 'alternative' sugars such as maltose (*malX, manZ, manX*), trehalose (*treC, treB*) and galactitol (*gatB, gatA*, only in population 7) possibly facilitated their translocation across the cell membrane. However, metabolism of amino acids and purines appeared to play a major role as well because upregulation of genes involved in the metabolism of serine (*serA*), threonine

(*thrB*, *thrC*), glycine (*ipdA*), purine (*purA*) and glutamate (*glsA* and *gadB*) was observed. Several genes with important roles in the general stress response (acid, oxidative and carbon-starvation stress) were also upregulated (*rpoH*, *gadE* and *gadX*). Anoxic habitats often have low pH and high concentrations of acids produced by fermentation. Under such conditions, genes that show pH dependency are co-induced (Yohannes *et al*., 2004). Our results suggest that under the applied oxygen-limited conditions, several acid-regulated pathways of catabolism and stress response were more induced in the evolved forms than in the ancestor.

Overall, the list of differentially upregulated genes in the anaerobic treatment showed few commonalities between the populations (populations 6 and 7). Analysis of the unique genes from population 6 by String 9.0 revealed the coexpression of several genes involved in the assembly and mobilization of iron-sulphur clusters (*iscA*, *iscS*, *hscA*, *hscB*, *sufA*, *sufD*, *nifU* and *erpA*), assisters of biochemical transformation of proteins (Giel *et al*., 2006), specific genes for stress response or acid resistance (*hdeA*, *hdeD*, *rpoE*) and other oligopeptide transporter genes (*oppBCDF*). In contrast, population 7 uniquely expressed genes involved in the response to amino acid starvation (*relBE*), next to transcriptional regulators involved in resistance to DNA damage (*uspB*, *uspD*, *uspE*). All these genes are typically overexpressed when cells respond to various stresses. Differences between populations 6 and 7 were remarkable among all data in accordance with the difference in the numbers of non-synonymous mutations between them (Table S3.2). As observed, population 7 had been subjected to a higher number of mutations that presumably generated the activation of stress resistance responses.

In general, forms of the aerobic and anaerobic treatments showed not only a more specialized response to oxygen availability or depletion, but also a more 'alleviating' one, with the activation of several stress response systems. Hence, these two treatments seem to have posed more adverse conditions to the ancestor than the fluctuating treatment. Activation of stress response genes is probably a very general feature in laboratory populations of microbes, and their inactivation might be favourable for particular forms. The reason is that stress responses may be costly, favouring any mutation that tune them down (Dettman *et al*., 2012). Thus, less stressful conditions in an environment would lead to lowered activation of stress response genes, inciting less phenotypic changes that

generally occur in transitions from exponential growth to (starvation) survival.

The presence of parallelism in independent populations indicates that populations may evolve to have the same genomic and phenotypic traits. Clearly, we found parallelism driven by the type of treatment; at the same time, high strain divergence was detected in independently evolving populations with parallel responses, indicating that although under a same treatment, populations had different (combined) evolutionary responses. This parallelism suggests that independent populations can adapt to similar environments via the same (or different) mutations (Wong *et al*., 2012). A hallmark of adaptive evolution and positive selection is the presence of parallel responses in independently evolving populations propagated in similar environments (Cooper *et al*., 2003; Woods *et al*., 2006; Ostrowski *et al*., 2008; Crozat *et al*., 2010; Hindré *et al*., 2012; Wong *et al*., 2012).

Remarkable within-population phenotypic diversity

Although parallelism was clearly present and driven by the treatments, the presence of interpopulational and intrapopulational heterogeneity was conspicuous. It may indicate an emerging ability upon growth under the respective treatment to colonize niches that had not been used by the ancestor. Thus, phenotypic heterogeneity was observed in all populations, and our selection of just a few forms per population may have been just the tip of the iceberg of the high degree of diversification taking place in each environment. One of the most evident examples of such diversification was given by the outliers of the oxic environment (aerobic treatment; Fig. 3.2).

We examined the phenotypic heterogeneity in one of the populations with outliers, specifically population 2. This population, interestingly, was found to have an outlier form that expressed a group of genes of the fimbrial complex (*fim* genes). Other studies have found that differential expression of such genes incites the emergence of colony polymorphisms (phase variation; Hasman *et al*., 2000). The physical presence of fimbriae on the cell surface prevents auto-aggregation (Hasman *et al*., 2000). Such *fim* cluster genes were upregulated in form 2a but not in forms 2b and 2c. Auto-aggregation assays (Hasman *et al*., 1999) revealed differential auto- aggregation in this population among

population 2 forms (Fig. 3.3b). The ancestral strain, which was included in the assay, did not display auto-aggregation.

In addition, forms 2a and 2b revealed other particular phenotypic behaviors representing excellent candidates to be characterized as having specific differential intrapopulational niche occupations. Form 2a revealed a remarkable pH resistance response, as several genes involved in the maintenance of pH homeostasis were induced (*hdeB*, *hdeA*, *hdeD*, *gadA*, *gadE*). Furthermore, it uniquely expressed the *spy* gene that is induced under cell wall stress, preventing protein aggregation and assisting protein refolding (Kwon *et al*., 2010). The unique behavior of form 2a suggested selection for pH resistance mechanisms. On the other hand, form 2b revealed an interesting 'anaerobic-like' behavior. Several genes involved in anaerobic respiration, i.e. those that allow the use of nitrate as the terminal electron acceptor (*narG*, *narH*, *narZ*) and reductase enzyme complex genes (*fdnH*, *fdnI*, *fdnG*) were upregulated. Even though well- oxygenated, the aerobic treatment milieu may have had anoxic conditions, allowing to respire anaerobically but also to produce fermentation products such as acetate and formate during early log-phase growth, which can reach high concentrations and lower the pH (Yohannes *et al*., 2004). Form 2b may have adaptively developed abilities to cope with such products. Another uniquely expressed group of genes in this form was given by those related to glycerol uptake and metabolism (*glpACB*, *glpFK*, *glpTQ* and *glpD*). Such genes function mostly under anaerobic conditions and offer the possibility to use glycerol as the carbon source (Murarka *et al*., 2008). The anaerobic-like metabolic behavior of form 2b and the resistant and persisting profile of form 2a confirm that these forms may deal with the environment in different fashions, having unique energy sources and therefore likely occupying different niches in the population.

Such non-parallel changes representing phenotypic polymorphisms are suggestive of genetic drift and evolutionary branching under sympatry (Herron and Doebeli, 2013). This diversification may have been caused by interactions between individuals (leading to coexistence) or driven by the environment perceived by the population, involving the main environmental factors (treatment effect or multiplicity of energy sources), resulting in internal ecological dynamics (Herron and Doebeli, 2013). Also, an emerging ability upon growth under the respective treatment to colonize niches that were left unoccupied

by the ancestor could have taken place. Although not tested here, empirical evidence has suggested that frequency- dependent selection, together with mechanisms such as cross-feeding and mutation selection balances, can maintain such diversity in apparent contravention of the competitive exclusion principle (Rainey *et al.*, 2000). Although the stability of the observed polymorphism has not been examined, it presumably is maintained by the heterogeneous nature of the environment. Ecological theory suggests that diversity is supported by the heterogeneity of an environment, as perceived by the organisms that occupy it and possibly related to differential availability of energy and carbon sources (e.g. multiplicity of substrates, cross-feeding) (Kassen, 2002).

Parallel regulatory responses leading to fitness improvement

The top 150 list of upregulated genes was processed in the MOLGEN server (http://server.molgenrug.nl/) to obtain the regulatory genes involved, without discrimination of their regulatory function (inducers or repressors). With this data set, we sought to analyse the regulatory similarities between the evolved forms and to look for a link with fitness. Fig. 3.5 shows a PCA analysis plotting all regulatory genes found from this analysis. PC1 explains almost 20% of the variation between all forms and clearly separates the aerobic and anaerobic treatments (right) from the fluctuating treatment (left). With this analysis, we found the main regulons involved in parallel upregulatory events, implying adaptive selective responses. With the exception of a few outliers, the forms clustered by treatment. Thus, across all treatments, similar genes are major regulators (controlling the expression of a high number of genes on the genome) relative to the ancestor. For example, the CRP gene regulates the expression of about 180 genes of *E. coli* involved in several central metabolic processes (transport of carbon sources, glycolysis-gluconeogenesis switching to TCA cycle, pyruvate dehydrogenase pathway and aerobic respiration). Interestingly, it has been found that CRP plays a key role in the activation of a high number of genes for utilization of carbon sources other than glucose (Shimada *et al.*, 2011), especially in environments where glucose is limiting. This correlates nicely with the used glucose- limited environment. Other commonly involved major regulators are FNR, FIS and H-NS (that regulate up to 21% of the genes in the *E. coli* genome), which act in the transcription of genes involved in oxygen transitions, acid

resistance, DNA binding, chemotaxis, cell structure, and molecular biosynthesis (Salmon *et al.*, 2003) and maintenance of nucleotide structure (Cho *et al.*, 2008; Wang *et al.*, 2012). These results thus suggest that the evolved forms exhibited a more active and dynamic behavior of central pathways relative to the ancestral strain.

Moreover, a putative relation was found between the positive effects of the regulators and the increase in fitness (Figs 3.4 and 3.5). The highest and most significant fitness enhancement after the long-term cultivation was observed in the forms of the aerobic and anaerobic treatments (w = 1.162 and w = 1.351). The forms within these treatments shared 55% (P > 0.05) of the regulatory genes involved in upregulation and appeared in the same site of the most explanatory axis. This result suggested an activation of regulatory genes that are likely involved in fitness improvement under both type of conditions.

Mutations in regulatory genes

Because even small changes in regulatory status can have a substantial impact on the phenotypic divergence associated with fitness effects (Ludwig *et al.,* 2005), we sought to examine which of the regulatory genes involved in upregulation may have undergone mutational events in order to elucidate any link between the observed genotype and phenotype (improved fitness relative to the ancestor, Fig. 3.4) and polymorphisms. Thus, the mutations across all genomes (Table S3.2) were compared with the list of regulatory genes found to be involved in increased gene expression. A diverse set of six regulatory genes were shown to have acquired mutations (highlighted with boxes in Fig. 3.5). Only one of these (*arcA*) was found to have undergone a mutation in only one of the treatments (aerobic treatment: populations 1 and 2), whereas the other five had consistently mutated in all forms across all treatments (same point mutations). The *arcA* gene product is a dual transcriptional regulator that represses or activates about 168 genes in the genome that are mainly involved in redox metabolism. Phosphorylation of *arcA* represses TCA cycle genes, hence less acetate may be formed. Interestingly, the mutation found (causing a replacement of alanine by valine) was located in a rather conserved region (A76V). However, the mutation might have consequences for one or more of the *arcA* regulated genes. When observing the genes upregulated by *arcA* in both populations (1 and 2) of

the aerobic treatment, we observed parallelism at the intrapopulation level. Candidate genes approaches have found parallelism in evolution of regulatory genes such as *spoT* (Cooper *et al.*, 2003). The majority of the upregulated genes in population 1 was related to oxidative phosphorylation (*cyoABCDE*) and galactose metabolism (*gatABCDZ*), whereas in population 2, these included more diverse results from anaerobic metabolism-related genes (form 2b, *glpABCD*) to hydrogen uptake-related genes (*hybB*, *hybC*, *hybO*). Thus, there was a clear distinction of the *arcA* upregulated genes between populations. Interestingly, in contrast with the rest of the forms of the aerobic treatment, only form 2a did not contain the *arcA* mutation, which is likely to be related to the specific niche occupation by this form in the population. Fitness results showed significant differences of this form (average $w = 1.180$) compared with its coevolved partners ($w = 1.050$ and 0.076) ($P < 0.05$). The impact of not having the *arcA* mutation will be further investigated. The exclusivity of this mutation to the aerobic treatment forms suggested a selective adaptive response. However, in our system, the different genomic background of each population may have affected the pleiotrophy of such mutation.
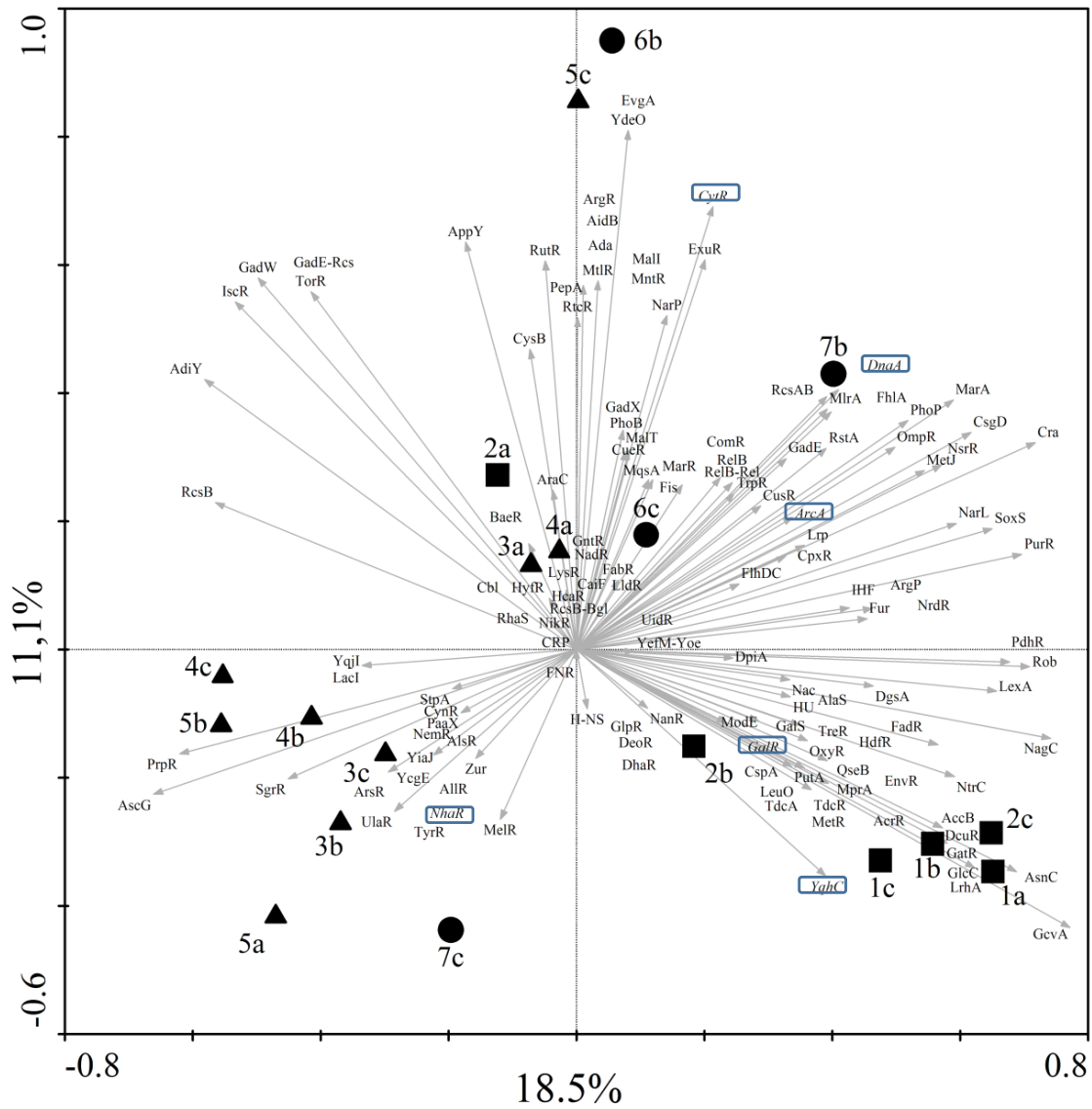
Figure 3.5. PCA of the regulatory genes involved in upregulated expression of the top 150 genes list. (■) Aerobic treatment forms, (▲) fluctuating treatment forms, (●) anaerobic treatment forms. Six major regulatory genes were shown to have acquired mutations (highlighted with boxes).

The other five regulatory genes with mutations (*yqhC, galR, cytR, dnaA, nhaR*) revealed consistent mutations across all treatments. Nevertheless, they did not appear to have a comparative effect on the gene expression response. *YqhC,* however, appeared as a consistent regulator involved in the upregulation in all forms of the aerobic treatment and in specific forms of fluctuating treatment (4b and 4c). This gene has been found to be involved in upregulation of *yqhD* and *dkgA* in another study (Pérez *et al*., 2008). Both of these genes are upregulated in the aerobic and fluctuating treatment forms; together, they

increase the resistance to reactive oxygen compounds in aerobic environments (Pérez *et al*., 2008). This result suggests that this mutation, which was likely triggered by the LB medium, is only positively selected in specific environmental conditions (where oxygen is involved), whereas it was rather neutral in the non-aerated environment.

Overall, such specific genomic and phenotypic changes could go with the upregulation of certain genes under certain environmental conditions but downregulation under other conditions. Comparative analysis of phenotypes from a genomic regulatory level is thus useful to elucidate certain phenotypic responses. However, this analysis can become tedious in a complex environment because of the high number of variables and because pleiotropic effects depend on the genomic backgrounds. The mechanisms involved in the regulatory changes are possibly part of a big network of adaptive responses triggered by the high number of mutational events. Moreover, this pleiotropy may, as a final outcome, produce convergent phenotypes (in this study, meaning enhanced-fitness forms).

Conclusions

From a global perspective, in this study, we could observe how phenotypic innovations depend on the relationship between genomic changes and ecological conditions. The number and nature of the genetic and gene expression changes correlated with a better ability to deal with the environmental traits. Environment appeared as the main driving force for adaptation, being – in our experimental setup – that constant and perhaps more stressful environments cause a more specialized and positive adaptive response compared with a more generalist and costly one in the fluctuating condition. The influence of the multiplicity of substrates in the medium and possibly emergent ecological dynamics are remarkable for heterogeneity. This study is a groundbreaker in describing the adaptive responses of *E. coli* to a complex environment in which a multiplicity of substrates is available and in empirically demonstrating phenotypic diversification at an intrapopulational level under these conditions. We found that parallel responses occur in independent populations in similar ecological settings, which may indicate a certain predictability of their evolution. However, the presence of such

parallelism may not do away with diversification as differential genomic backgrounds can further affect it. Although tedious, further evolution experiments in complex environments in which larger numbers of evolved forms are screened genotypically and phenotypically will elucidate the flexibility versus parallelism of the adaptive outcomes and address how selection in complex systems acts on complex organisms.

Acknowledgements

References

Baev, M.V., Baev, D., Radek, A.J., and Campbell, J.W. (2006a) Growth of *Escherichia coli* MG1655 on LB medium: determining metabolic strategy with transcriptional microarrays. *Appl Microbiol Biotechnol* 71: 323–328.

Baev, M.V., Baev, D., Radek, A.J., and Campbell, J.W. (2006b) Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of amino acids, peptides, and nucleotides with transcriptional microarrays. *Appl Microbiol Biotechnol* 71: 317–322.

Baev, M.V., Baev, D., Radek, A.J., and Campbell, J.W.(2006c) Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols, and organic acids with transcriptional microarrays. *Appl Microbiol Biotechnol* 71: 310–316.

Bantinaki, E., Kassen, R., Knight, C.G., Robinson, Z., Spiers, A.J., and Rainey, P.B. (2007) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* 176: 441–453.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., *et al*. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243–1247.

Bennett, A.F., Lenski, R.E., and Mittler, J.E. (1992) Evolutionary adaptation to temperature . I . Fitness responses of *Escherichia coli* to changes in its thermal environment. *Evolution* 46: 16–30.

Brockhurst, M.A., Colegrave, N., and Rozen, D.E. (2011) Next-generation sequencing as a tool to study microbial evolution. *Mol Ecol* 20: 972–980.

Cho, B.K., Knight, E.M., Barrett, C.L., and Palsson, B. (2008) Genome-wide analysis of *Fis* binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome res* 18: 900–910.

Cooper, T.F., Rozen, D.E., and Lenski, R.E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci USA* 100: 1072–1077.

Cotter, P.A., Chepuri, V., Gennis, R.B., and Gunsalus, R.P. (1990) Cytochrome o (cyoABCDE) and d (cydAB) oxidase gene expression in *Escherichia coli* isregulated by oxygen, pH, and the fnr gene product. *J Bacteriol* 172: 6333– 6338.

Crozat, E., Winkworth, C., Gaffé, J., Hallin, P.F., Riley, M.A., Lenski, R.E., and Schneider, D. (2010) Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Mol Biol Evol* 27: 2113– 2128.

Dettman, J.R., Rodrigue, N., Melnyk, A.H., Wong, A., Bailey, S.F., and Kassen, R. (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol Ecol* 21: 2058–2077.

Dong, T., Kirchhof, M.G., and Schellhorn, H.E. (2008) RpoS regulation of gene expression during exponential growth of *Escherichia coli* K12. *Mol Genet Genomics* 2008: 267– 277.

Elena, S.F., and Lenski, R.E. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4: 457–469.

Fong, S., Joyce, A.R., and Palsson, B. (2005) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* 15: 1365–1372.

Giel, J.L., Rodionov, D., Liu, M., Blattner, F.R., and Kiley, P.J. (2006) IscR-dependent gene expression links iron-sulphur cluster assembly to the control of O2-regulated genes in *Escherichia coli*. *Mol Microbiol* 60: 1058–1075.

Gresham, D., Desai, M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., *et al*. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 4: e1000303.

Hasman, H., Chakraborty, T., and Klemm, P. (1999) Antigen- 43-mediated autoaggregation of *Escherichia coli* is blocked by fimbriation. *J Bacteriol* 181: 4834–4841.

Hasman, H., Schembri, M.A., and Klemm, P. (2000) Antigen 43 and type 1 fimbriae determine colony morphology of *Escherichia coli* K-12. *J Bacteriol* 182: 1089–

1095.

Helling, R.B., Vargas, C.N., and Adams, J. (1987) Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* 116: 349–358.

Henderson, I.R., Meehan, M., and Owen, P. (1997) Antigen 43, a phase-variable bipartite outer membrane protein, determines colony morphology and autoaggregation in *Escherichia coli* K-12. *FEMS Microbiol Lett* 149: 115– 120.

Herring, C.D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M.K., Joyce, A.R., *et al*. (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38: 1406–1412.

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

Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. (2012) New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nat Rev Microbiol* 10: 352– 365.

Hughes, B.S., Cullum, A.J., and Bennett, A.F. (2007) Evolutionary adaptation to environmental pH in experimental lineages of *Escherichia coli*. *Evolution* 61: 1725–1734.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., *et al*. (2009) STRING 8 – a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37: D412–6.

Kasimoglu, E., Park, S.J., Malek, J., Tseng, C.P., and Gunsalus, R.P. (1996) Transcriptional regulation of the proton-translocating ATPase (atpIBEFHAGDC) operon of *Escherichia coli*: control by cell growth rate. *J Bacteriol* 178: 5563–5567.

Kassen, R. (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity. *J Evol Biol* 15: 173–190.

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M.C. (2012) Experimental evolution. *Trends Ecol Evol* 27: 547–560.

Kjaergaard, K., Hasman, H., Schembri, M.A., and Klemm, P. (2002) Antigen 43-mediated autotransporter display, a versatile bacterial cell surface presentation system. *J Bacteriol* 184: 4197–4204.

Kwon, E., Kim, D.Y., Gross, C.A., Gross, J.D., and Kim, K.K. (2010) The crystal structure *Escherichia coli* Spy. *Protein Sci* 19: 2252–2259.

Lacour, S., and Landini, P. (2004) S -dependent gene expression at the onset of stationary phase in *Escherichia coli: f*unction of S -dependent genes and identification of their promoter sequences. *J Bacteriol* 186: 7186–7195.

Lee, D.-H., and Palsson, B.Ø. (2010) Adaptive evolution of *Escherichia coli* K-12

MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. *Appl Environ Microbiol* 76: 4158–4168.

Lenski, R.E., and Bennett, A. (1993) Evolutionary response of *Escherichia coli* to thermal stress. *Am Nat* 142: 47–64.

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

Lepš, J. (2003) *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge, UK: Cambridge University Press.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005) Functional evolution of a cis-regulatory module. *PLoS Biol* 3: 588–598.

McKenzie, G.J., Harris, R.S., Lee, P.L., and Rosenberg, S.M. (2000) The SOS response regulates adaptive mutation. *Proc Natl Acad Sci USA* 97: 6646–6651.

MacLean, C., Dickson, A., and Bell, G. (2005) Resource competition and adaptive radiation in a microbial microcosm. *Ecol Lett* 8: 38–46.

Maharjan, R., Seeto, S., Notley-McRobb, L., and Ferenci, T. (2006) Clonal adaptive radiation in a constant environment. *Science* 113: 514–517.

Murarka, A., Dharmadi, Y., Yazdani, S.S., and Gonzalez, R. (2008) Fermentative utilization of glycerol by *Escherichia coli* and its implications for the production of fuels and chemicals. *Appl Environ Microbiol* 74: 1124–1135.

Orr, H.A. (2010) The population genetics of beneficial mutations. *Philos Trans R Soc Lond B Biol Sci* 365: 1195–1201.

Ostrowski, E.A., Woods, R.J., and Lenski, R.E. (2008) The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proc Biol Sci* 275: 277–284.

Papp, B., Notebaart, R.A., and Pál, C. (2011) Systems- biology approaches for predicting genomic evolution. *Nat Rev Genet* 12: 591–602.

Pepin, K.M., Samuel, M.A., and Wichman, H.A. (2006) Variable pleiotropic effects from mutations at the same locus hamper prediction of fitness from a fitness component. *Genetics* 172: 2047–2056.

Pérez, J.M., Arenas, F.A., Pradenas, G.A., Sandoval, J.M., and Vásquez, C.C. (2008) *Escherichia coli* YqhD exhibits aldehyde reductase activity and protects from the harmful effect of lipid peroxidation-derived aldehydes. *J Biol Chem* 283: 7346–

7353.

Philippe, N., Crozat, E., Lenski, R.E., and Schneider, D. (2007) Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *BioEssays* 9: 846–860.

Puentes-Téllez, P.E., Hansen, M.A., Sørensen, S.J., and van Elsas, J.D. (2013) Adaptation and heterogeneity of *Escherichia coli* MC1000 growing in complex environ- ments. *Appl Environ Microbiol* 79: 1008–1017.

Rainey, P., Buckling, A., Kassen, R., and Travisano, M. (2000) The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends Ecol Evol* 15: 243–247.

Rainey, P.B., and Travisano, M. (1998) Adaptive radiation in a heterogeneous environment. *Nature* 32: 69–72.

Riehle, M.M., Bennett, A.F., Lenski, R.E., and Long, A.D. (2003) Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature. *Physiol Genomics* 14: 47–58.

Romero, I.G., Rubinsky, I., and Gilad, Y. (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* 13: 505–516.

Rosenzweig, R.F., Sharp, R.R., Treves, D.S., and Adams, J. (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* 137: 903–917.

Salmon, K., Hung, S.P., Mekjian, K., Baldi, P., Hatfield, G.W., and Gunsalus, R.P. (2003) Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J Biol Chem* 278: 29837–29855.

Schmittgen, T.D., and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 3: 1101–1108.

Shimada, T., Fujita, N., Yamamoto, K., and Ishihama, A. (2011) Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS ONE* 6: 1–11.

Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: 1–25 Article 3.

Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., *et al.* (2012) The molecular diversity of adaptive convergence. *Science* 335: 457–461.

Van der Woude, M.W., and Henderson, I.R. (2008) Regulation and function of Ag43 (flu). *Annu Rev Microbiol* 62: 153–169.

Velicer, G.J., Raddatz, G., Keller, H., Deiss, S., Lanz, C., Dinkelacker, I., and Schuster, S.C. (2006) Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proc Natl Acad Sci USA* 103: 8107–8112.

Wang, W., Li, G.W., Chen, C., Xie, X.S., and Zhuang, X. (2012) Chromosome organization by a nucleoid- associated protein in live bacteria. *Science* 333: 1445–1449.

Weber, H., Polem, T., Heuvelin, J., Wendisch, V., and Hengge, R. (2005) Genome-wide analysis of the general stress response network in *Escherichia coli*: δs-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* 187: 1591–1603.

Wong, A., Rodrigue, N., and Kassen, R. (2012) Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. *PLoS Genet* 8: e1002928.

Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A., and Lenski, R.E. (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci USA* 103: 9107–9112.

Yohannes, E., Barnhart, D.M., and Slonczewski, J.L. (2004) pH-dependent catabolic protein expression during anaerobic growth of *Escherichia coli* K-12. *J Bacteriol* 186: 192– 199.

Zhang, E., and Ferenci, T. (1999) OmpF changes and the complexity of *Escherichia coli* adaptation to prolonged lactose limitation. *FEMS Microbiol Lett* 176: 395–401.

Chapter 4

Sympatric Metabolic Diversification of Experimentally Evolved *Escherichia coli* in a Complex Environment

Pilar Eliana Puentes-Téllez, Jan Dirk van Elsas

Abstract

Sympatric diversification in bacteria has been found to contravene initial evolutionary theories affirming the selection of the fittest type by competition for the same resource. Studies in unstructured (well-mixed) environments have discovered divergence of an ancestor strain into genomically and phenotypically divergent types growing both on single and mixed energy sources. This study addresses the metabolic diversification in an *Escherichia coli* population that evolved over ~1000 generations under aerobic conditions in the nutritional complexity offered by Luria-Bertani (LB) broth. The medium lacked glucose but contained a variety of other resources. Two distinct metabolically-diverged types, coinciding with colony morphologies, were found to dominate the populations. One type was an avid carbohydrate consumer, which could quickly utilize the available (alternative) substrates feeding into glycolysis. The second type was a slow grower, which was able to specifically consume acetate. The capacity to utilize acetate might be providing an advantage to this second type, suggesting an increased capability to deal with adverse conditions that occur in the later stages of growth. The diverged metabolic preferences of the two forms suggested coexistence in the complex medium and the presence of an interactive metabolism. We postulate that these types used different alternative metabolic strategies occupying different niches in a sympatric manner as an outcome of adaptation to the complex environment.

Introduction

Recent experimental evolution studies have placed a focus on understanding the diversification of bacteria occurring in sympatry. Such studies have found a direct relationship between the extent of diversification and environmental heterogeneity. Rainey and Travisano (1998) first observed rapid diversification occurring in structured conditions in *Pseudomonas fluorescens* growing in multiple-substrate medium. They concluded that spatial structure, triggering an uneven distribution of resources, was required to induce observable diversification (Rainey *et al*., 1998). Other studies using spatially-structured environments also demonstrated diversification processes in other bacteria (Korona *et al*., 1996; Reboud *et al*., 1997; Ponciano *et al*., 2009). On the other hand, in unstructured (well-mixed) environments, the "competitive exclusion" principle is thought to place limits on the diversity of the population in the system, as competition for the same resource would select for just one (fittest) variant (Hardin, 1960). However, contradictory evidence on this principle was recently produced in bacteria growing under sympatric conditions with just one limiting resource (Kinnersley *et al*., 2009). MacLean *et al*. (2005) re-examined the outcome of selection in unstructured conditions in *Pseudomonas* and found metabolic differentiation of different genotypes that had grown sympatrically, competing for mixed resources during 24 days of cultivation (MacLean *et al*., 2005). Moreover, several studies confirmed the diversification of *Escherichia coli* growing in mixed resources (Friesen *et al*., 2004; Hall *et al*., 2007; Le Gac *et al*., 2008; Puentes-Téllez *et al*., 2013, 2014). Hence, media containing mixes of resources can be seen to provide a form of "structure" under which an organism may differentially specialize (Kassen *et al*., 2002, 2004).

Diversification may thus occur in an environment with multiple resources as a consequence of intense competition for commonly-used energy sources, yielding specialized forms that shift their nutritional preference and are better able to use as-yet-unexploited resources (MacLean *et al*., 2005). New niches may thus emerge through the actions of the organisms themselves. In such instances, diversity is created which can lead

to coexistence maintained through interactions between the different types that are present.

Early work has already revealed particular emerged diversities on the basis of the rates of uptake of specific substrates (Claassen *et al*., 1986, Helling 1*et al*., 1987), patterns of gene expression (Kurlandzka *et al*., 1991; Rosenzweig *et al*., 1994; Kinnersley *et al*., 2009) and genetic changes (Adams *et al*., 1992; Herron *et al*., 2013). However, analyses of diversification in sympatric conditions have been limited to the use of only glucose or mixtures of glucose with just a few energy sources (Helling *et al*., 1987; Friesen *et al*., 2004). It is felt that greater environmental and ecological complexity should be integrated into the studies to get closer to the complexity of natural and industrial settings (Dettman *et al*., 2012). In the current study, we included high nutritional complexity in the factor environment by utilizing a medium, Luria-Bertani (LB) broth, that contains a complex mixture of energy sources and lacks glucose (Hanko *et al*., 2004). This is, to our knowledge, the first study reporting the outcome of evolution in this highly used and complex medium, which is composed of complex substrates commonly used in industrial fermentations. Thus, we studied diversification in an *E. coli* population that had evolved over ~1000 generations under aerobic conditions and was previously shown to possess genomic and transcriptomic heterogeneities (Puentes-Téllez *et al*., 2013, 2014). The scope of this work includes the investigation of metabolic patterns emerging in such heterogeneous populations, next to the dynamics of diversification, using colony morphology and differential metabolic behavior as the criteria.

Materials and Methods

Strains and classification based on colony morphology.

A stored sample from an end-point population (day 150, ~1000 generations) of an *E. coli* K12 MC1000 culture that had undergone selection in serial-batch growth under aerobic conditions in LB broth (Puentes-Téllez *et al*., 2013), was spread onto 3 different LB agar plates to isolate colonies. After overnight growth at 37°C, a total of 30 colonies were randomly selected from all the plates. These colonies were observed using a

stereoscope (40x) and classified using colony size and morphology. Cells belonged to three morphological types (defined here as "forms"), denoted "a", "b and "c". All colonies were then sampled to be used for PCR fingerprinting using ERIC PCR primers (ERIC-I: ATgTAAgCTCCTggggATTCAC; ERIC-II: AAgTAAgTgACTggggTgAgCg) (Versalovic *et al.*, 1991) and run in an agarose gel to confirm identity relative to the ancestral strain.

Metabolic assessment

The metabolic responses of cells from all selected colonies were assessed using Biolog GN2 plates (Biolog, Hayward, CA, USA). To this end, a clone originating from each of the 30 colonies was obtained by re-isolation on LB agar. Cells were resuspended in inoculation fluid (IF) provided by Biolog to an optical density (OD, 590nm) of 0.070 $\pm$0.005. The microplates were then inoculated and incubated at 37°C for 72 h. Absorbance values (turbidity at OD_{590}) were obtained at 24, 48 and 72 h of incubation. Data were interpreted as utilization versus no utilization of each substrate using a cutoff optical density of 0.2 (Maharjan *et al.*, 2007). Data were then analyzed by constructing a multidimensional scaling (MDS) plot as well as by principal components analysis (PCA) of substrates categorized by their nature (Primer 6.1.13) (Clarke *et al.*, 2006). The metabolic information of the ancestor was included in the analyses using the same setup. In the PCA, compounds that could not be classified within the main metabolic groups were assigned as "other compounds".

Maximal growth rate in specific substrates

Using triplicates, the growth of each form in 0.3% of specific carbon sources (galactose, glucose, fructose, glycerol, mannose, L-arabinose, trehalose and acetate (0.03%)) and in selected amino acids (arginine (0.6mM), proline (0.4mM), lysine (0.4mM)) was tested in 100ml flasks containing 30ml of M9 minimal medium ($MgSO_4$ (3mM), $CaCl_2$ (0.12mM), and 24% of M9 minimal salts 5X (Na_2HPO_4 (3.39%), KH_2PO_4 (0.15%), NaCl (0.25%), NH_4Cl (0.5%) ,0.0002% thiamine and 0.1% of a trace elements mix (2.5 g/l EDTA; 1.5 g/l FeSO4; 0.025 g/l CoCl2; 0.025 g/l ZnSO4; 0.015 g/l MnCl2; 0.015 g/l NaMoO4; 0.01 g/l NiCl2; 0.02 g/l H3BO3; 0.005 g/l CuCl2) (Jiménez *et al.*, 2013), at 37°C, with shaking at 200 rpm during 5 d. Growth was measured every 4 h (CFU/ml) by plating dilution series of each sample onto LB agar. The ancestor was

included in the analyses. The maximum growth rate (V_{Max}, (Vasi *et al.*, 1994)) under each specific substrate was calculated as the log transformed CFU/ml value against time during exponential growth phase; three estimates of V_{Max} were obtained per form in each substrate. In addition, growth of each form in LB broth was recorded every h (6 fold) during 24 h at 600nm using a microplate reader set with shaking before each read and at 37°C (VersaMax, Molecular Devices Corp). After 24 h of growth, a sample of each well was plated onto LB agar to obtain the average number of cells per type (CFU/ml).

Results and Discussion

We assessed the metabolic capabilities of individual morphotypes in an *E. coli* K12 population that evolved for ~1,000 generations in serial-batch LB broth cultures under aerobic conditions. Thirty colonies of the end-population were randomly picked from three different LB agar plates. The colonies that were picked showed different morphologies, which were classified in 3 forms, as follows: (1) small colonies with diameters <1 mm, pronounced center and irregular shape ("fried egg" shaped) ("a" form), (2) large/rough/irregular colonies with diameters >1mm ("b" form) and (3) large/smooth/irregular colonies with diameters > 1mm ("c" form) (See Fig. 4.1 and Puentes Tellez *et al*, 2013). In our selection, 56% of the colonies were form "a", whereas 20 and 23% were forms "b" and "c" respectively. All picked colonies revealed ERIC-PCR patterns which were indistinguishable from those of the ancestral *E. coli* type, confirming their identity as evolved forms from this ancestor and discarding possible contamination (Fig. 4.1).
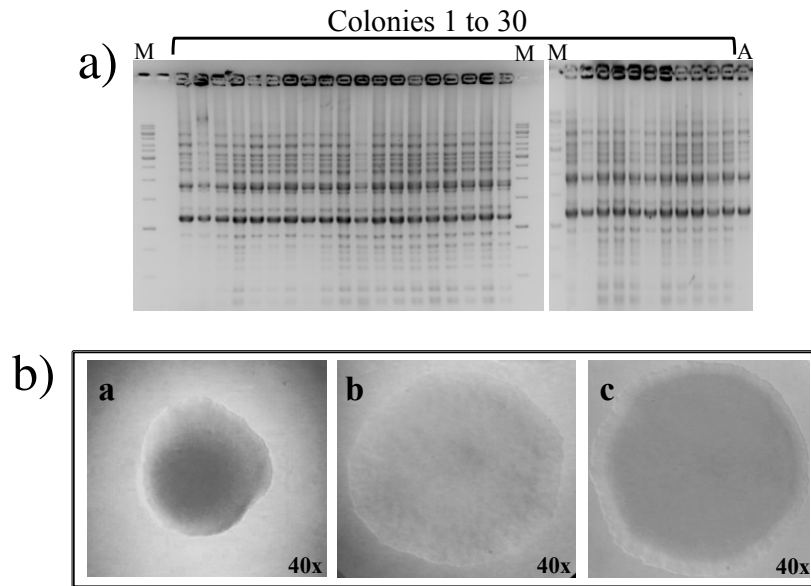
Figure 4.1. a) ERIC-PCR patterns of selected colonies (1 to 30) and the ancestor (A), M: 1kb Marker. b) Photography (40x) of the three observed morphologies (Forms "a", "b" and "c").

In previous work, although a significant level of consistency in genomic and phenotypic changes was found among the forms of this population, colony morphology together with other properties was found to correspond to differential genomic as well as gene expression backgrounds. For example, genomically, form "a" showed a higher number of "non-consistent" (not found in all the forms) mutations compared to forms "b" and "c". The latter two forms had a specific mutation in the major regulatory gene *arcA*. Gene *arcA* is a regulator with effects on nearly 170 genes, and it is involved in oxygen transitions. It is normally inactive during aerobic growth, but represses *rpoS* when active. The effect of differential changes in a high-impact regulatory gene like *arcA* are suggestive of differential phenotypic behaviors and potential niche partitioning within the population. Moreover, higher numbers of genes were upregulated in forms "b" and "c" compared to form "a". However, form "a" showed a increased upregulation of key genes related to resistance to adverse conditions as well as those strongly related to colony morphology (e.g. *fim*) (Puentes-Téllez *et al.*, 2013, 2014).

Diversity of metabolic patterns

Analyses of the metabolic capabilities of all 30 evolved forms, next to the ancestral type, were performed using Biolog GN2 plates. The collective data obtained after 72 h were used to construct a (normalized) multidimensional scaling (MDS) plot (Fig. 4.2.) The analysis revealed that all forms, including the "a", "b" and "c" forms, clustered in two large groups, away from the ancestral type, on the basis of the degree of utilization of the 95 substrates. However, the ancestor appeared to be more similar to the metabolic "cloud" formed by form "a". On the other hand, all clones encompassing forms "b" and "c" clustered together, indicating a fairly similar metabolic behavior. The analysis thus provided evidence for a presumed correlation between colony morphology and metabolic behavior, with forms "b" and "c" (larger and irregular colonies) being more similar and clearly different from form "a" (smaller-"fried-egg" shaped) and the ancestor.
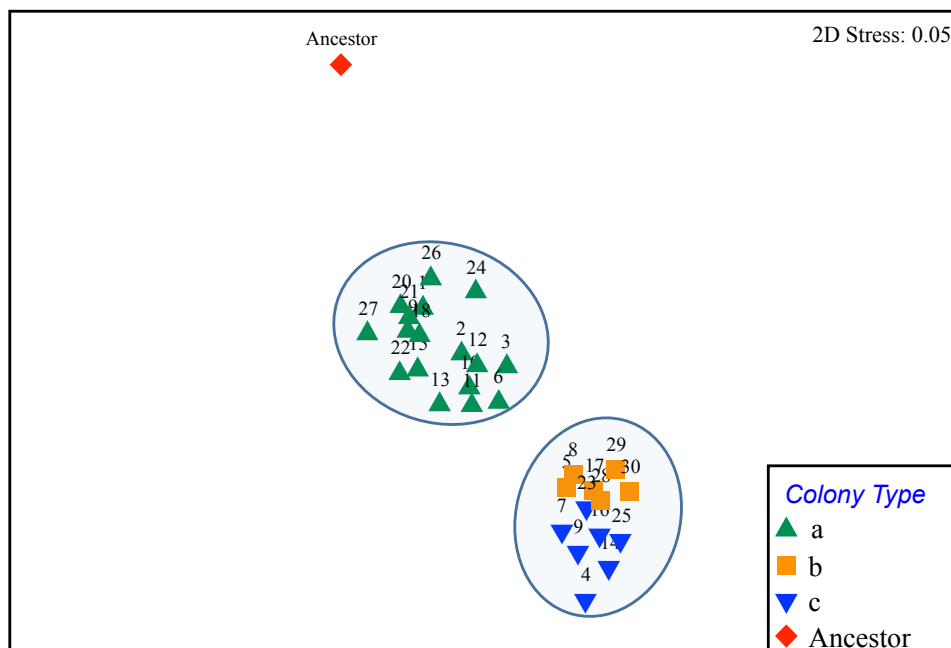
Figure 4.2. MDS plot for square-root normalized Biolog results using Euclidean distance. Distance was obtained from the values obtained in an end-point measurement (after 72 hours of incubation) of 30 randomly selected colonies. There are two main clusters suggesting two different metabolic patterns among the selected colonies.

We then used PCA to more closely observe the drivers of the Biolog substrate diversification. The analysis was performed using only the two observed "metabolic types", being metabolic type *a* (form "a") and the metabolic type *c* (form "c" used as a proxy for forms "b" and "c" together) (Fig. 4.3). The PCA revealed that most of the variation and metabolic distance between the two types is explained by the first axis (62.8%). At a "substrate-type" level, this distance involved carbohydrates, carboxylic acids, amino acids and a group of "other (unclassified) compounds".
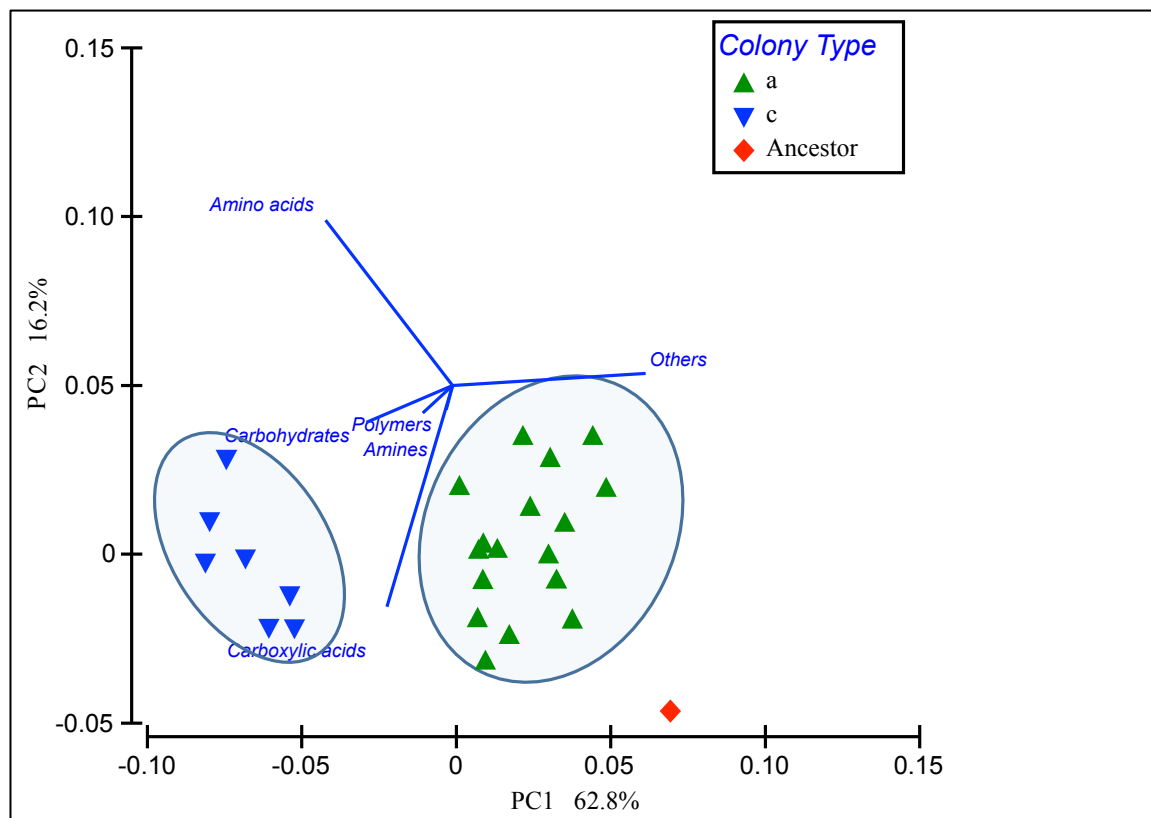
Figure 4.3. Principal component analysis with Biolog substrates classified by their nature, using morphology types *a* and *c*. Samples cluster in two different groups. The first axis explains most of the variation (62.8%), mainly triggered by carbohydrates, carboxylic acids amino acids and "other compounds" of the Biolog plate. The ancestor appears to have less metabolic distance to type *a*.

Metabolic preferences of the two selected metabolic types

Table 4.1 shows the average of the OD_{590} values obtained after 72 h in the Biolog plates for the most explanatory variables, listing only the substrates that showed

significant differences ($P<0.05$) between the two metabolic types. It also shows the response of the ancestor to these specific substrates. Clearly, type *c* appeared to be a significantly faster carbohydrate consumer than type *a* and the ancestor. After 72 h, the majority of carbohydrates had been consumed by this type to an $OD_{590}>1.0$, whereas type *a* showed an overall lower response in the utilization of these carbohydrates. According to the absorbance values (Table 4.1), type *c* grew faster on glucose as well as on glucose-related sugars (coupling to glycolysis).

Thus, monosaccharides such as N-acetyl-D-glucosamine, L-arabinose, D-fructose, L-fucose, D-galactose, D-mannose, L-psicose and β-methyl-d-glucoside were rapidly consumed by type *c*, as well as disaccharides such as α-d-lactose, maltose, lactulose, D-melibiose, D-trehalose and the sugar alcohol D-mannitol. Although LB broth does not contain glucose, the ancestral capability to metabolize glucose was maintained in the two selected types (although at lower degree than the other tested carbohydrates).

The enhanced carbohydrate utilization by type *c* might be the result of a more active metabolic cascade in the central metabolism. This is consistent with the measured faster response by type *c* (more than 50% faster, significant difference $P<0.05$) when grown on carboxylic acids related to specific sugars (glucose, galactose and lactose) such as D-galactonic acid lactone, D-galacturonic acid, D-gluconic acid, D-glucuronic acid, D,L-lactic acid, next to three carboxylic acids included in the central metabolic pathway, i.e. pyruvic acid methyl ester, α-ketoglutaric acid and succinic acid (all three part of the tricarboxylic acid cycle (TCA cycle)) (Table 4.1).

Table 4.1. OD $_{590}$ values obtained for the most discriminative (significantly different) substrates between the two "metabolic types" *a* and *c*. Significant difference: two-tailed *t* test (P<0.05). Absorbance values obtained by the ancestor are displayed in order to show a statistical comparison with the closest type (type *a*) (All values have been normalized as 0.00 = OD <0.2).

Substrates	Type *a*	± SEM	Type *c*	± SEM	Ancestor	± SEM
Carbohydrates						
N-acetyl-D-glucosamine	0.663	0.095	1.272	0.068	0.454	0.126
L-arabinose	0.433	0.042	1.130	0.088	0.000	0.000
D-fructose	0.403	0.042	1.041	0.081	0.341	0.093
L-fucose	0.893	0.055	1.256	0.141	0.266	0.041
D-galactose	0.588	0.027	1.350	0.148	0.000	0.000
D-glucose	0.350	0.045	0.920	0.099	0.479	0.045
α-D-lactose	0.505	0.070	1.248	0.066	0.000	0.000
Lactulose	0.985	0.055	1.404	0.072	0.000	0.000
Maltose	0.664	0.067	1.256	0.082	0.672	0.186
D-mannitol	0.866	0.081	1.307	0.142	0.543	0.083
D-mannose	0.513	0.054	1.145	0.098	0.431	0.153
D-melibiose	1.006	0.051	1.467	0.074	0.000	0.000
β-methyl-D-glucoside	0.357	0.039	0.885	0.080	0.000	0.000
D-psicose	0.000	0.000	0.219	0.011	0.000	0.000
L-rhamnose	1.244	0.062	1.520	0.038	0.281	0.000
D-sorbitol	1.024	0.070	1.485	0.028	0.381	0.016
D-trehalose	0.795	0.099	1.276	0.093	0.527	0.057
Carboxylic acids						
Pyruvic acid methyl ester	0.290	0.040	1.063	0.095	0.000	0.000
Acetic acid	0.463	0.017	0.000	0.000	0.000	0.000
D-galactonic acid lactone	0.677	0.047	1.490	0.125	0.389	0.157
D-galacturonic acid	0.281	0.047	1.428	0.081	0.376	0.239
D-gluconic acid	0.533	0.080	1.204	0.067	0.505	0.010
D-glucuronic acid	0.637	0.120	1.461	0.080	0.212	0.001
α-ketoglutaric acid	0.792	0.046	1.632	0.076	0.000	0.000
D,L-lactic acid	0.827	0.101	1.636	0.061	0.547	0.100
Propionic acid	0.538	0.034	0.680	0.022	0.000	0.000
D-saccharic acid	0.287	0.026	0.000	0.000	0.340	0.062
Succinic acid	0.425	0.046	1.084	0.076	0.000	0.000
Amino acids						
D-alanine	0.857	0.084	1.330	0.132	0.000	0.000
L-alanine	0.672	0.037	1.709	0.084	0.000	0.000
L-alanyl-glycine	0.439	0.041	1.600	0.110	0.201	0.058
L-asparagine	0.659	0.031	1.415	0.168	0.249	0.113
L-aspartic acid	0.596	0.068	1.871	0.091	0.233	0.096
L-glutamic acid	0.000	0.000	0.741	0.080	0.000	0.000
Glycyl-l-glutamic acid	0.000	0.000	0.304	0.046	0.000	0.000
L-proline	0.425	0.053	1.565	0.134	0.212	0.011
D-serine	0.446	0.067	1.404	0.035	0.343	0.065
L-serine	0.513	0.045	0.285	0.056	0.000	0.000
L-threonine	0.000	0.000	0.225	0.010	0.212	0.002
γ-aminobutyric acid	0.000	0.000	0.208	0.018	0.000	0.000
Amines						
Glucuronamide	0.435	0.057	1.113	0.130	0.000	0.000
L-alaninamide	0.000	0.000	0.221	0.018	0.000	0.000
Other compounds						
Bromosuccinic acid	0.387	0.021	0.514	0.046	0.000	0.000
Thymidine	0.898	0.039	1.169	0.067	0.293	0.071
Glycerol	0.770	0.022	1.520	0.048	0.232	0.022
D,L.α-glycerol phosphate	0.478	0.014	0.797	0.070	0.206	0.008
D-glucose-6-phosphate	0.251	0.011	1.017	0.064	0.366	0.154

In grey: Significantly different compared to type *a* (*P*<0.05); SEM: Standard error of the mean

Many glycolytic enzymes are constitutively expressed in *E. coli*, however those of the TCA cycle are inducible. Thus, the TCA cycle can respond differently to the presence of oxygen and carbon sources (Han , 2002). Hence, differences in the utilization of carbon sources between the two types (*a* and *c*) may have triggered the difference in response to the carboxylic acids of the TCA cycle.

We also found that the majority of amino acids were consumed faster by type *c* than by type *a* as well as the ancestor. *Escherichia coli* can utilize amino acids as structural components next to their use as carbon, nitrogen or energy sources. In particular D- and L-alanine, D-alanyl-glycine, L-asparagine, L-aspartic acid, L-proline, L-threonine γ-aminobutyric acid, D and L-serine have been found to serve as excellent carbon and nitrogen sources (Chan *et al*., 1981; Dover *et al*., 1972). These were indeed effectively consumed by type *c*. L-glutamic acid and glycyl-l-glutamic acid were consumed to a lower extent by type *c*. Glutamate is rarely used by *E. coli* K-12 as a carbon source, but it is a common nitrogen source. On the other hand, several mutants have been reported to grow on this amino acid as sole carbon source (Halpern *et al*., 1961; Marcus *et al*., 1969).

Concerning the amine group, there was an enhanced utilization of glucuronamide and L-alaninamide by type *c*. In the group of "other" compounds, we found an enhanced utilization of glycerol - and its metabolically related compound D,L,α-glycerol phosphate - by type *c* (as compared to the ancestor and type *a*). *E. coli* growing aerobically on glycerol slots the compound into the central metabolism as dihydroxyacetone phosphate (DHAP), a metabolite which can participate in both gluconeogenic and glycolytic processes (Martinez-Gómez *et al*., 2012), and so one can presume that this pathway had an enhanced rate in type *c*. Although not having a strong impact on the separation of the types in the PCA graph (observed when excluded; data not shown), the "other compounds" group revealed other substrates showing differences between types *a* and *c* related to glucose metabolism (thymidine, D-glucose-6-phosphate), which is consistent with the premise of type *c* cells being adaptive in the sense of a more active metabolism of carbohydrates.

Type *a,* on the other hand, revealed a metabolic profile which resembled that of the ancestor. Thus, there were similar differences between types *c* and *a* as between type *c* and the ancestor. However, some differences between type *a* and the ancestor were found to be significant (Table 4.1, in grey, *P*<0.05). Whereas type *a* was able to utilize several sugars and amino acids (even to a low degree), the ancestor did not utilize these or did so to a much lower degree. Thus, generally speaking, the ancestor revealed a more restricted metabolic potential than the evolved types.

Maximal growth rate in specific (LB related) substrates

To analyze the response to specific substrates of LB broth, we performed growth experiments with types *a* and *c*. LB lacks glucose (Hanko *et al*., 2000, 2004) and the two major components of LB, tryptone and yeast extract (YE), are derived from complex biological sources. These two components typically contain partially digested proteins (tryptone), vitamins (YE), next to a variety of trace carbohydrates and unknown substances (Hanko *et al*., 2004). Both components thus contain 7.7 and 17.5% of carbohydrates, respectively (BD, Franklin Lakes, New Jersey), at a final carbohydrate concentration of about 0.16%. Previous studies confirmed the presence in LB of considerable amounts of trehalose, D-galactose, L-arabinose, D-mannose, melibiose and L-fucose, among other compounds (Hanko 2000, Baev *et al*., 2006c). However, in *E. coli K-12* MG1655 cultures, steady-state growth ceased early (at an OD_{600} of approx. 0.3). This shift occurred in particular when the available carbohydrates were consumed, and the ensuing lowering of growth rate forced cell metabolism to switch to the use of amino acids as carbon source (Sezonov *et al*., 2007). The metabolic differences between types *a* and *c* can be thought of as indicating niche differentiation and (differential) evolutionary adaptation of these two types to the environment. We included various substrates in minimal medium in order to shed light on the different metabolic routes taken by types *a* and *c* (Table 4.2). Samples were taken at 4-h intervals, given the slow growth in the minimal medium, with just one substrate supporting growth as a carbon and energy source. Thus, maximal growth rates were determined (V_{Max}) using substrate concentrations that were convenient for estimating V_{Max}, i.e. the limiting maximal growth rate that is adopted by the microbial growth (Vasi *et al*., 1994).

The results (Table 4.2), confirmed the faster growth of type *c* than of type *a* as well as the ancestor on all tested carbohydrates (*P*<0.05). In contrast, the response to amino acids was not significantly different (L-arginine and lysine offer both carbon and nitrogen sources to *E. coli*), with the exception of type *c* growing with L-proline (0.4mM). L-proline is well utilized by many *E. coli* strains as an excellent carbon/nitrogen and energy source (Chen *et al*., 1991). In general, *E. coli* apparently has evolved to handle mixtures of amino acids and complex oligopeptides, rather than single amino acids, as energy sources (Chen *et al*., 1991). Therefore, our analysis was expected to not reveal a clear picture of the metabolic capabilities of *E. coli* consuming this type of compounds.

Table 4.2. *V Max* (h^{-1}) in M9 minimal medium with specific substrates as sole energy source.

Substrate	a (h^{-1}) ($\pm$ SEM)		c (h^{-1})	($\pm$SEM)	Ancestor (h^{-1})	($\pm$SEM)
Galactose 0.3% †	0.052	0.004	0.191*	0.002	0.032	0.006
Glucose 0.3%	0.004	0.002	0.162*	0.004	0.013	0.000
Fructose 0.3%	0.015	0.002	0.133*	0.005	0.023	0.010
Glycerol 0.3%	0.017	0.003	0.184*	0.002	0.021	0.004
Mannose 0.3%	0.022	0.002	0.172*	0.003	0.010	0.001
Trehalose 0.3%†	0.023	0.003	0.218*	0.016	0.034	0.004
L-arabinose (0.3%)†	0.090	0.011	0.130	0.009	0.000	0.000
L-arginine 0.6mM†	0.053	0.005	0.040	0.002	0.017	0.001
Lysine 0.4mM†	0.013	0.001	0.011	0.001	0.016	0.008
Acetate 0.03%	0.063*	0.003	0.032	0.004	0.022	0.008
Glycine 0.4mM†	0.043	0.005	0.066	0.000	0.081	0.011
Proline 0.4mM†	0.030	0.006	0.067*	0.002	0.010	0.002

†: Substrates present in LB broth (Hanko *et al*., 2004; Baev *et al*., 2006c)
*: Significant difference calculated between types *a* and *c*. P<0,05 (two-tailed *t* test) SEM: Standard error of the mean.

Overall, type *c* exhibited a more active metabolism on available substrates, possibly indicating a higher metabolic flux, than type *a* as well as the ancestor. One of the factors that may most strongly limit the growth rate is the import of substrates into

the cell ("metabolic bottleneck") (Chen *et al*., 1991). A highly active carbohydrate metabolism is often controlled by major regulatory systems like the carbohydrate phosphotransferase system (PTS), which catalyzes the transport/uptake of carbohydrates and their coupling to glycolysis. This PTS has been found to assure the optimal utilization of carbohydrates in complex environments (Kotrba *et al*., 2001). Since glucose is absent from LB broth, the type *c* cells may have developed a strategy based on alternative ways to get substrate for the glycolysis pathway and the central burning of carbon.

Interestingly, type *a* showed a significantly different response from that of type *c* when growing in minimal medium with acetate as the sole carbon and energy source, confirming the higher utilization of acetate that was already observed in the Biolog plates (Table 4.1). Acetic acid is a (released) by-product of the rapid aerobic growth of *E. coli* (Han *et al*., 1992). Its release is particularly associated with high rates of metabolism of glucose or glucose-related compounds. Acetate may actually establish adverse conditions to fast-growing cells. Production of acetate can be significant in batch fermentations, given the extended growth phase in such systems, which allows this compound to attain elevated concentrations. A metabolic switch has been observed when such adverse conditions occur in an environment in which oxygen and glucose decrease by 50 to 80%, with a concomitant initiation of acetate utilization (Kleman *et al*., 1994). In this scenario, acetate can rapidly accumulate and become an alternative carbon source, allowing the evolution of divergent types that are able to exploit the new niche, maintaining interactions with other members of the system. These interactions may sustain diversification. Helling *et al*. (1987) and Rosenzweig *et al*. (1994) already demonstrated the presence of stable heterogeneity involving acetate in defined medium cultures that had originated from a unique ancestor (Helling *et al*., 1987; Rosenzweig *et al*., 1994). In the system, the use of glucose as the carbon and limiting energy source effectively increased and, as a consequence, overproduction of acetate ensued. Then, a type able to specifically utilize acetate emerged. Thus, the catabolic pathways of glucose, acetate and glycerol may be interconnected in a multiple-type population, counteracting the emergence of a single clone that consumes these substrates together. Instead, three metabolically different types were successfully maintained in the population. Although a clear limiting substrate was not present in our experimental setup, we found here that primarily type *c* may have used alternative pathways for the available resources in LB,

feeding into glycolysis. Hence, the type *c* cells may have efficiently fed the central metabolic pathway, producing secondary metabolites like acetate and thus creating an available niche for the newly emerged type *a* which was able to utilize this substrate. The enhanced assimilation of acetate might be linked to an enhancement of acetyl-CoA synthetase (produced by the *acs* gene) (Lin *et al*., 2006); further studies will address this issue at the transcriptomic level in types *a* and *c*.

Dynamics of growth in LB broth

To observe the population dynamics of types *a* and *c* in the complete mosaic of resources offered by LB, growth experiments were performed for each type in separate in LB. Specific growth rates (μ) were determined by using the slope of the logarithm of growth in the exponential phase. Figure 4.4 shows data collected over 24 h in 96-well microplates containing 100uL of LB broth. The values of μ after 7 h were 0.22 h^{-1} (SD+0.019) for type *a* and 0.26 h^{-1} (SD +0.004) for type *c*. However, at the beginning of stationary phase (5 h later), when growth rates had decreased for both strains, type *c* had lowered its growth rate to 0.11 h^{-1} (SD +0.002) compared to 0.17 h^{-1} (SD+0.004) for type *a*. Thus, by the end of the measurement, there was a significant difference in the end-populations between types *a* and *c* ($P<0.001$). Viable (CFU) counts confirmed this, with CFU/ml end-values being $1x10^9$ CFU/ml for type *a* and $3.5x10^8$ CFU/ml for type *c*. These results suggest a faster exponential growth of type *c* over type *a* as a result of the faster consumption of the available primary resources. On the other hand, type *c* rapidly reduced its growth rate upon depletion of the available sources. Apparently, type *a* was not a fast "primary source" eater, however it finished its growth later, with a higher final cell count than type *c*.

Friesen *et al.* (2004) revealed the presence of stable heterogeneity in *E. coli* lines that evolved in mixtures of glucose and acetate. Diauxic behaviour (sequential metabolism of two energy sources resulting in a metabolic shift between these two substrates, producing two separate growth phases), in which the initial growth phase is correlated with the use of the preferred metabolite (e.g. glucose) and a subsequent growth phase with the use of the less preferred metabolite (e.g. acetate) was demonstrated (Tyerman *et al.*, 2005).
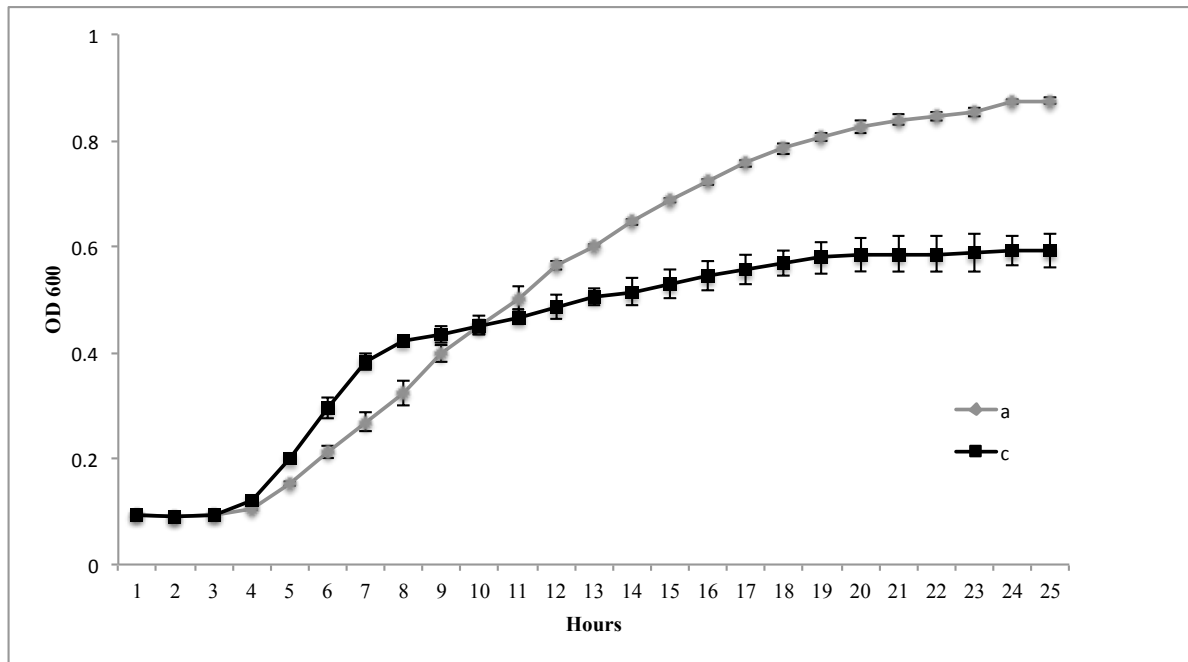
Figure 4.4. Optical density (OD$_{600}$) of types *a* and *c* growing separately in LB broth during 24h.

Our results obtained in LB broth did not demonstrate a diauxic shift, but instead demonstrated the emergence of new niches which were occupied by the evolved types in an environment without a defined limiting substrate. Although the scope of this work was limited to the description of the dynamics of the diversification and not the mechanisms maintaining it, the results suggest the presence of forces like frequency-dependent selection acting upon these types. Although sympatric diversification has been demonstrated before in experiments with a defined mixture of just a few substrates (Friesen *et al.*, 2004), showing splits of sympatrically growing linages at the genetic level (Herron *et al.*, 2013), the current study is - to our knowledge - the first one demonstrating such diversification in the presence of multiple substrates offered by LB without a pre-

established limiting substrate. The metabolic behavior that we found in type *a* correlated to a slower use of the available mosaic of resources of LB. Such a lower growth rate in the early growth stages may offer an opportunity (to a then lower number of cells) to easily consume the available resources, and since generation time is larger, the population can make use of the resources for a longer time. In the light of its acetate consumption, type *a* also was a superior by-product consumer. These combined characteristics, i.e. a lower growth rate (often linked to enhanced stress resistance (Notley *et al*., 1996)) next to an enhanced utilization of the by-product acetate, may confer on type *a* an unique survival capacity when together with type *c* under conditions that may be adverse to any other type. A future study will analyze the dynamics of the different metabolic types when growing together in LB and the presence of other "social" advantages between the members of the population.

Taken together, the results of this study confirm the emergence of heterogeneity and niche differentiation in an *E. coli K12* MC1000 population after its evolution in serial batch cultures over ~1000 generations. The habitat this population grew in represents a highly complex environment with a vast mixture of substrates providing a form of " structure" at a microscale. As a general rule for any habitat, the number of niches available for microbial niche occupants is thought to match the extent of environmental variation. Hence, given the presumed complexity in our habitat, even more subtle polymorphisms and other possible interactions might be assumed to occur. Here, we detected the clear emergence of two dominating metabolically-different types in the population, denoted "types" *a* and *c*, (metabolic type *c* represented two forms, denoted forms "b" and "c"). Types *a* and *c* both developed novel alternative strategies to deal with the substrates in LB broth, which lacks glucose. Type *c* was a fast consumer of carbohydrates alternative to glucose, with a metabolic slot into the glycolysis pathway, whereas type *a* was a slow grower on such carbohydrates. However, given the ability of type *a* to consume acetate, it might be able to better cope with an environment that is created in a later stage of development, when primary resources have been consumed and conditions have become rather adverse. Since acetate is produced by high-rate glucose transformation pathways, type *a* might be taking advantage of the production of this by-product, reincorporating it into its metabolism. The presence of interactions in the population correlates with the stable dynamic behavior in the system. This study

demonstrates the outcome of adaptation to complex environments commonly found in nature or in industrial settings involving the presence of multiple energy sources and the lack of a limiting substrate.

Acknowledgments

References

Adams J, Puskas-Rozsa S., Simlar J., Wilke CM. (1992). Adaptation and major chromosomal changes in populations of *Saccharomyces cerevisiae*. Curr Genet 22:13-19.

Baev MV, Baev D, Janeso A, Campbell JW. (2006c). Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols and organic acids with transcriptional microarrays. Appl Microbiol Biotechnol 71:310 -316.

Chan TT, Newman EB. 1981). Threonine as a carbon source for *Escherichia coli*. J Bacteriol. 145(3): 1150-1153.

Chen LM, Maloy S. (1991). Regulation of proline utilization in enteric bacteria: cloning and characterization of the Klebsiella put control region. J Bacteriol 173(2): 783-790.

Claassen, P.A.M., Kortstee, G.J.J., Oosterveld-van Vliet, W.M., van Neerven, A.R.W. (1986). Colonial heterogeneity of *Thiobacillus versutus*. J Bacteriol 168: 791 - 794.

Clarke, K., and Gorley, R. (2006). PRIMER v6, User manual/tutorial. PRIMER-E, Plymouth.

Dettman JR, Rodrigue N, Melnyk H, Wong A, Bailey SF, Kassen R. (2012). Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol Ecol 21:2058-2077.

Dover S, Halpern YS. (1972). Utilization of Aminobutyric Acid as the Sole Carbon and Nitrogen Source by *Escherichia coli* K-12 Mutants. J Bacteriol. 109(2): 835-843.

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(2): 245-260.

Hall AR, Colegrave N. (2007). How does resource supply affect evolutionary diversification? Proc R Soc B 274:73-78.

Halpern Y. S., Umbarger HE. (1961) . Utilization of L-Glutarnic and 2-Oxoglutaric Acid as Sole Sources of Carbon by *Escherichia coli*. J gen Microbiol. 26, 175-183

Han, K., H. C. Lim, and J. Hong. (1992). Acetic acid formation in *Escherichia coli* fermentation. Biotechnol Bioeng 39:663-671

Han, L. (2002). Physiology of *Escherichia coli* in Batch and Fed-batch Cultures with Special Emphasis on Amino Acid and Glucose Metabolism. Bioteknologi, Stockholm. ISBN 91-7283-276-2

Hanko VP, Rohrer J. (2000). Determination of Carbohydrates, Sugar Alcohols, and Glycols in Cell Cultures and Fermentation Broths Using High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection. Anal Biochem 283 (2): 192-199.

Hanko VP, Rohrer J. (2004). Determination of amino acids in cell culture and fermentation broth media using anion-exchange chromatography with integrated pulsed amperometric detection. Anal Biochem. 324:29-38

Hardin G. (1960). The Competitive Exclusion Principle. Science 131 (3409): 1292-1297

Helling RB, Vargas CN, Adams J. (1987). Evolution of *Escherichia coli* during growth in a constant environment. Genetics 116:349-358.

Herron MD., Doebeli M. (2013). Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. Plos Biol 11(2): e1001490. doi:10.1371/journal.pbio.1001490.

Jiménez DJ, Korenblum E, van Elsas JD. (2013). Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion . Appl Microbiol Biotechnol DOI 10.1007/s00253-013-5253-7

Kassen R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. Science. 15: 173-190

Kassen R., Rainey PB. (2004). The ecology and genetics of microbial diversity. Annu Rev Microbiol. 58:207-31

Kinnersley MA, Holben WE, Rosenzweig F. (2009).*E Unibus Plurum*: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. Plos Genet. 5(11):e1000713. doi: 10.1371/journal.pgen.1000713.

Kleman GL and Strohl WR. (1994). Acetate metabolism by *Escherichia coli* in high-cell-density fermentation. Appl Environ Microbiol. 60 (11):3952.

Korona R. (1996). Adaptation to Structurally Different Environments. P Roy Soc Lond B Bio. 263(1377) : 1665-1669

Kotrba P, Inui M, Yukawa H. (2001). Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. J Biosci Bioeng. 92(6):502-17.

Kurlandzka A, Rosenzweig RF, Adams J. (1991). Identification of adaptive changes in an evolving population of *Escherichia coli*: the role of changes with regulatory and highly pleiotrophic effects. Mol Biol Evol. 8:261-281.

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

Lin H., Castro N., Bennett G, San KY. (2006). Acetyl-CoA synthetase overexpression in *Escherichia coli* demonstrates more efficient acetate assimilation and lower acetate accumulation: a potential tool in metabolic engineering. Appl Microbiol Biotech. 71 (6):870-874.

MacLean RC., Dickson A, Bell G. (2005). Resource competition and adaptive radiation in a microbial microcosm. Ecol Lett. 8: 38-46

Maharjan RP, Seeto S, Ferenci T. (2007). Divergence and redundancy of transport and metabolic rate-yield strategies in a single *Escherichia coli* population. J Bacteriol.189:2350 -2358.

Marcus M, Halpern YS. (1969). Genetic analysis of the glutamate permease in *Escherichia coli* K-12. Bacteriol. 97:1118-1128.

Martínez-Gómez, K., Flores, N., Castañeda, H.M., Martínez-Batallar, G., Hernández-Chávez, G., Ramírez, O.T., Gosset G., Encarnación S., Bolivar F. (2012). New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb Cell Fact. 11: 46.

Notley, L., Ferenci, T. (1996). Induction of RpoS-dependent functions in glucose-limited continuous culture: what level of nutrient limitation induces the stationary phase of *Escherichia coli*?. J Bacteriol 178: 1465-1468

Ponciano, JM., La HJ., Joyce P., Forney LJ. (2009). Evolution of Diversity in Spatially Structured *Escherichia coli* Populations. Appl Environ Microb. 75 (19): 6047-6054.

Puentes-Téllez, P.E., Hansen, M.A., Sørensen, S.J., and van Elsas, J.D. (2013). Adaptation and heterogeneity of *Escherichia coli* MC1000 growing in complex environments. Appl Environ Microbiol. 79: 1008-1017.

Puentes-Téllez, P.E., Kovacs A.T., Kuipers O.P., van Elsas, J.D. (2014). Comparative genomics and transcriptomics analysis of experimentally evolved *Escherichia coli* MC1000 in complex environments. Environ Microbiol. 16(3):856-870.

Rainey P and Travisano M. (1998). Adaptive radiation in a heterogeneous environment. Nature 394, 69-72.

Reboud X., Bell G. (1997). Experimental evolution in Chlamydomonas. III. Evolution of specialist and generalist types in environments that vary in space and time. Heredity 78: 507-514

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.

Sezonov G., Joseleau-Petit D., D'Ari R. (2007). *Escherichia coli* Physiology in Luria Bertani Broth. J Bacteriol. 189 (23): 8746-8749.

Tyerman J., Havard, N., Saxer G., Travisano M., Doebeli1 M. (2005). Unparallel diversification in bacterial microcosms. Proc Biol Sci. 272 (1570): 1393-1398.

Vasi F. Travisano M, Lenski R.E. (1994) . Long term experimental evolution in *Escherichia coli* II. Changes in life-history traits during adaptation to a seasonal environment. Am Nat. 144 (3): 422-456.

Versalovic J, Koeuth T, Lupski JR. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. 19:6823- 6831.

Chapter 5

Differential Stress Resistance and Metabolic Traits Underlie Coexistence in a Sympatrically Evolved Bacterial Population

Pilar Eliana Puentes-Téllez, Jan Dirk van Elsas

Abstract

Following intermittent batch growth in LB for about 1,000 generations, differentially evolved forms were found in a population of *Escherichia coli* cells. Studies on this population revealed the emergence of polymorphisms, as evidenced by analysis of both whole genome sequences and transcription analysis. Here, we investigated the phenotypic differences and nature of such forms and found a remarkable interactive coexistence of forms that revealed the presence of (1) *trade-offs* conferring a raised tolerance to environmental stress, and (2) *trade-offs* involving enhanced capacity to utilize alternative carbohydrates like galactose. This confirms the existence of different ecological roles within the population i.e. niche partitioning, as a consequence of adaptive evolution. Remarkably, the two forms were shown to continue to coexist in an experiment that cycled them through periods of feast (plentiful growth substrates) and famine (growth-restrictive = stress conditions). The results revealed that when one of these environmental factors was high tuned, the equilibrium of the interactive coexistence was destroyed affecting stability of the coexisting pair.

138

Introduction

Many natural and artificial environments expose bacteria to fluctuating conditions, either in space or time. Bacterial populations can deal with these dynamics using their adaptive capacity, in other words they may evolve. Evolutionary theory posits that monomorphic populations in nature come about as a result of intense competition. Niche overlap between forms that emerge in sympatry was thus thought to generate only one winner type due to competitive exclusion (Gause, 1934). In contrast, stable coexistence of phenotypically-distinct bacterial populations in sympatric conditions, a phenomenon which was denoted niche partitioning, was found later (Rosenzweig *et al*., 1994). Thus isogenic populations of *Escherichia coli* in glucose-limited continuous culture have been shown to become polymorphic, as metabolically distinct populations emerged from an ancestral strain (Helling *et al*., 1987). The latter phenomenon was also found for *E. coli* cultures developing in complex growth medium (Puentes-Téllez *et al*., 2014b unpublished). In another experiment, replicate *E. coli* populations diversified into two coexisting metabolic types due to a diauxic shift in serial batch culture under a dual mix of substrates (glucose-acetate) (Friesen *et al*., 2004).

Bacterial populations may thus respond to the environment by diversifying according to available resources. Although experimental evidence in multi-resources and complex environments is still needed, information collected from simpler environments revealed that diversification occurs as a consequence of intense competition for commonly-used energy sources. Thus, the emergence of specialized forms able to use unexplored (vacant) niches or novel resources can generate divergent populations (Rainey and Travisano, 1998, MacLean *et al*., 2005). In complex environments, the diversification process is driven by the environment and by genomic changes occurring within populations (Puentes-Téllez *et al*., 2013a). In general, the most important mechanism maintaining such diversification is frequency-dependent selection (Friesen *et al*., 2004, Tyerman *et al*., 2005). Thus, nutrient mixtures can offer spatial "structure" with fine-scale variation (Kassen, 2002) and adaptation to this type of environments (and hence increased fitness) enhances the likelihood of diversification (Buckling *et al*., 2003,

Spencer *et al.*, 2008). Although diversification may be initially transient (Rainey and Travisano, 1998), the emergence of interactions between types increases the possibility of stable heterogeneous communities, and in complex environments the opportunity for the development of polymorphisms maintained by interactions between the components may be enhanced (Helling *et al.*, 1987, Rosenzweig *et al.*, 1994). The occurrence of polymorphisms and coexistence can be assessed using the specific phenotypic characteristics of different forms of the population. Direct experimental evidence for adaptive diversification under both sympatric and complex conditions is still largely lacking.

A population of *E. coli* cells growing in Luria-Bertani broth with sequential- batch cultures under constant oxygen (during ~1000 generations), was recently shown to diversify into multifold types. Two distinct metabolic types were identified after random selection of multiple colonies from the population. The initial studies on this population revealed that colony morphology correlated with distinct differential genetic and metabolic backgrounds (Puentes-Téllez *et al.,* 2013a, 2014a, 2014b unpublished). Here, we further assess the ecological nature of selected diverged types that apparently maintained a coexisting interaction and possible niche partitioning. The daily transfers (after 24h) exposed the population to alternating periods of 'feast and famine' (cycled short-exponential and a long-stationary phases). Thus, we made use of information obtained from these two different growth stages in which distinct physiological and metabolic events take place in response to transient environmental changes and assessed wheater coexistence equilibrium was lost when the environment favored one of the two types (Experimental setup; Fig. S5.1, Supplementary information). We obtained evidence for the contention that adaptive evolution differentially affects the survival properties of members of the coexisting populations and that these properties - together with specific regulatory *trade-offs* – may be beneficial and allow stable coexistence.

Materials and Methods

Strains

Isolated colonies obtained from an end-point populations of an experimental evolution setup with *Escherichia coli K12* MC1000 were characterized in terms of morphology and metabolic behavior (Puentes-Téllez *et al.,* 2013a, 2014b unpublished). We sorted the colonies in three different types: (1) small (*"a"* form), diameter <1 mm, pronounced center ("fried egg" shaped), (2) large/rough/irregular (*"b"* form), diameters >1mm and (3) large/smooth/regular (*"c"* form) with diameters > 1mm. The population had undergone evolution under aerobic conditions in sequential batch cultures at $37^{o}C$ during ~1,000 generations in Luria-Bertani (LB) medium. The ancestral strain was kept frozen at $-80^{o}C$ for further analysis and comparisons.

Assessment of differential survival capabilities

The three forms and the ancestor were exposed separately to oxidative (2.0mM H_2O_2), osmotic (1.5M NaCl), heat ($49^{o}C$) and acidity stresses (pH 3.0) during 4 h. Assessment of survival involved two stages: (1) preparation of the cells (production of the inoculum) and (2) challenge stage. For the initial stage, cells of each form were grown in triplicate systems until two different growth stages (incubation at $37^{o}C$; 200rpm) in LB medium, i.e. exponentially growing cells (after 5h, optical density (OD) 600nm $0.5\pm$ 0.05) and stationary-phase cells (after 18h of growth; $OD_{600} > 1.0$). When the desired OD was reached, samples were taken, serving as inocula for the challenge phase of the experiment. For the second stage, inocula were diluted into fresh LB broth to Log 4 CFU/ml. To reduce effects of the initial growth condition, each stress was applied after 20 min of acclimation ($37^{o}C$, 200 rpm), time 0. A sample was plated on LB just prior to stress application; these counts were taken as 100% of the population. Population size was then recorded every h by plating samples on LB medium to obtain the number of viable cells. Percentage survival was determined by dividing the number of CFU ml^{-1} during challenge by the total number of CFU ml^{-1} at time 0 and multiplying by 100.

Coexistence experiment

Two selected 'metabolic types' (*a* and *c*, representing morphologies *"a"* and *"c"* (the latter as a proxy of morphology *"b"),* Puentes-Téllez *et al*., 2014b unpublished) were competed against each other in a crossed experiment. To achieve this, marked strains were obtained by selecting spontaneous mutants to rifampicin (30μg/ml) and streptomycin (20μg/ml) for each form. The marked strains were analyzed by measuring their fitness after 24h of growth against their isogenic forms. The relative fitness (*w*) of each mutant relative to their isogenic form was calculated as the ratio of the Malthusian parameters (m= ln[(density at end of competition)/(density at time zero of competition)] (Lenski RE, *et al*., 1991). Only mutants with a fitness value equal to 1 ±0.05 were selected, meaning no difference between resistant and non-resistant types. Fitness was confirmed by measuring in competition each marked strain against its opposite marked form. Coexistence was assessed by growing mixes of the forms (having opposite markers), with daily transfers into fresh media (dilution 1:10,000) after 24 h and during 4 d (3 transfers). Growth took place in three different environments, i.e. LB broth, LB supplemented with galactose (0.3%) (LB+galactose) and LB supplemented with a non-lethal concentration of H_2O_2 (0.3mM) (LB+H_2O_2). Dynamics of the populations was evaluated by plating samples on LB agar with the respective antibiotic, obtaining the population density of each type at exponential (OD_{600} 0.5 ±0.05) and stationary phase (after 24 h) after each transfer. At the end of competition (d 4), averaged relative fitness was calculated from the ratio between the two types using their Malthusian parameters (ln (density at day 4/ density at time 0 (day 1)). After the last sample was taken, cultures were returned to incubation and allowed to continue growing without any transfers for three more days. After this time, samples were plated to assess the percentage of each type in the final population. Individual growth rates (μ) were obtained by growing each type in each environment separately; growth rate was calculated as the slope of the curve during exponential growth phase.

Specific growth rate in supplemented LB broth

Growth rate was evaluated in three different environments. For this, a sample of each type was grown in triplicates in LB broth, LB broth+galactose (0.3%) and LB broth+H_2O_2 (0.3mM). Growth took place in a microplate reader set at 37°C (VersaMax,

Molecular Devices Corp). OD_{600} was recorded every hour. Specific growth rate (μ) was calculated from the steepest slope of the curve against time $\mu = \ln(OD_1/OD_0)/t_1-t_0$.

Messenger RNA (mRNA) isolation

Growth curves of selected colony types were obtained by growing cells at low density in LB broth, LB broth+galactose (0.3%) and LB broth+H_2O_2 (0.3mM). For this, the optical density of two biological replicates per colony type was recorded at 600nm using a microplate reader. One mL of cells was harvested at mid-exponential phase (OD_{600} 0.5 $\pm$0.05) and at stationary phase (OD_{600} 1.1 $\pm$0.05), number of $CFUml^{-1}$ in the sample was obtained by plating on LB agar. Pellets, harvested by centrifugation (10.300g, 1 min) were immediately used for total RNA isolation with NucleoSpin RNA II isolation kit (Macherey – Nagel, Biokè, Leiden, the Netherlands). Total RNA concentrations and purity were assessed using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific). Total RNA samples were enriched for mRNA using the MICROBExpress™ Bacterial mRNA Enrichment Kit (Ambion, Austin, TX), which removes the 16S and 23S ribosomal RNAs from the total RNA population. Enriched mRNA populations were reverse-transcribed into cDNA with the Superscript III reverse transcriptase kit (Invitrogen, Carlsbad, USA).

Reverse transcription – quantitative (real-time) PCR

A total of nine genes were used for amplification with real-time PCR using cDNA as a template. For this, eight primer sets were designed using the default parameters of Clone Manager Suite (Sci-Ed Software). The sequences of the primer sets used for real-time PCR analysis are shown in Table S5.1 (Supplementary information) (Eurogentec, Maastricht, The Netherlands). All real-time PCR runs for each biological replicate were performed in duplicate. The reaction was performed using the SensiFAST™ SYBR Hi-ROX Kit (Bioline) on the ABI Prism 7300 Cycler (Applied Biosystems, Germany). The specificity of the amplification products was confirmed by melting-curve analysis and on 1.5% agarose gels. Standard curves were obtained using serial dilutions of genomic DNA containing a purified sequence (QIAquick PCR purification kit) of each target amplified by PCR (Liss 2002, Wong *et al.*, 2005). Dilutions ranged from 10^7 to 10^2 gene copy numbers/µl. Data normalization was performed taking into account the number of CFU per ml for each sampled timepoint (Copy number of RNA/cell) (Kanno *et al.*, 2006).

Statistical analysis of all data sets in order to observe differences between the forms was performed using two-tailed Students's *t* test.

Results and Discussion

Divergent survival capabilities under osmotic, temperature and acidity stresses

To assess whether the survival capabilities had diverged between forms in the population, we exposed the three selected forms ("*a*", "*b*" and "*c*", representing different colony morphologies), next to the ancestor, to osmotic, high temperature and acidity stresses using cells from two different growth stages (exponential growth and stationary phase). Fig. 5.1 shows the changes in viability of the cells (as percentage survival) during the first 4 h of challenge.

We observed remarkable differences between the data obtained with cells from both growth stages. The response to osmotic and high temperature conditions of exponential-phase cells was quite diverse (Fig 5.1a). Under both stresses, the ancestor appeared as the better survivor, maintaining the population stable during the first 2 h of challenge. This resistance persisted until the end of the heat challenge. In the osmotic stress treatment, after 3 h, the ancestral cells revealed a decreasing trend. A second striking observation was that exponential-phase cells of forms "*b*" and "*c*" behaved similarly under both osmotic and heat stress. Interestingly, these forms were more sensitive to osmotic than to heat stress, revealing greater decreases under the former stress condition. Exponential-phase cells of form "*a*" revealed a particular behavior under osmotic and heat stress challenges. Whereas the ancestor seems to have better survival abilities in both conditions, type "*a*" shows a decrease to around 2% of the initial population after the first 2 h under osmotic conditions, however after this point, survival became slightly higher, with about 10% of the initial cell numbers surviving at the end of the challenge. There is an apparent recovering capacity of this form being able to deal in a better way with this challenging condition. On the other hand, form "*a*" showed a dramatic decay without any recovering as from the first h of the heat challenge, and after 2 h only 1% of the population was still viable.

For all forms (including the ancestor), stationary-phase cells revealed similar survival in the osmotic and heat stress challenges (Fig. 5.1b). However, whereas osmotic stress reduced the population sizes to less than 10% after the first h, heat stress did so to a limited extent (only >50% of loss was observed up to the third h of challenge). The similar response patterns found upon osmotic and heat challenges may have been related to a similar activation of the key proteins that act upon starvation for energy (or carbon) sources. During stationary phase, bacterial cells produce a large set of 'starvation proteins', which assist in enhance the cross-protection against stressors such as osmotic tension, heat, oxidative stress and the pressure of antibiotics (Groat *et al.,* 1986). The main responsible for this stress-tolerant state in bacteria is the *rpoS* gene, of which the product allows the cells to survive the actual stresses as well as additional stresses that are not yet present (Hengge-Aronis, 2002, Dragosits *et al*., 2013).
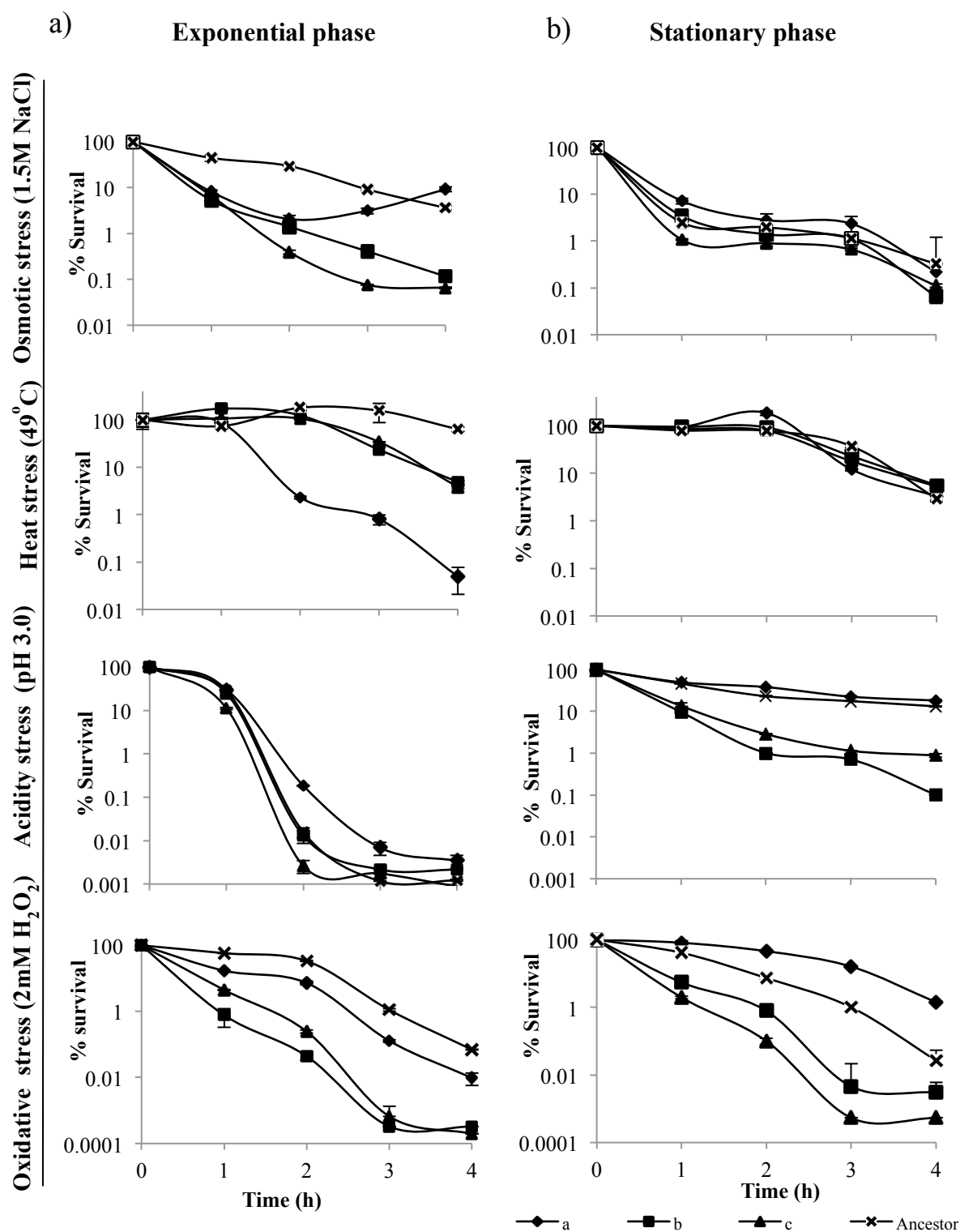
Figure 5.1. Survival test. Forms *"a", "b"* and *"c "* and the ancestor were challenged to four different stress conditions added to LB. Osmotic stress (1.5M NaCl), heat stress (49°C), oxidative stress (2mM H_2O_2) and pH stress (pH 3.0). Challenge took place during 4h and after pre-adaptation (5 h and 18 h – Exponential-phase cells and stationary phase cells respectively). Logarithmic plots show the percentage of loss relative to the initial (100%) population. All results are reported as the mean + SD of three replicates.

Moreover, all exponential-phase forms revealed a similar response to acidity (pH 3.0), being sensitive (Fig. 5.1a). Concerning stationary-phase cells, a tendency to acid tolerance was found in form *"a"* and the ancestor (Fig. 5.1b), whereas the population sizes of forms *"b"* and *"c"*, decreased dramatically after the first h of challenge. A remarkable activation of a "pH resistance response" in "form *a*"-like cells was observed in our previous work, with several genes involved in the maintenance of pH homeostasis (*hdeB, hdeA, hdeD, gadA, gadE*) being induced in late exponential phase (Puentes-Téllez *et al*., 2014a). This suggested that form *"a"* cells had an enhanced capacity to resist acid conditions and thus occupied this specific niche. In addition, our previous work also found that forms *"b"* and *"c"* had similar metabolic behavior, including the propensity to efficiently breakdown substrates under aerobic conditions. This active metabolism generates acetate, which can become available for reincorporation into the central metabolism. The fast and active metabolism in forms *"b* and *"c"* suggested a rapid production of acetate. Form *"a"* has been previously found to be able to consume the acetate that is likely released by the active metabolism of forms *"b"* and *"c"* (Puentes-Téllez *et al*., 2014b unpublished). Under these conditions, acetate might accumulate and lower the pH (Wolfe *et al*., 2005). This suggests an evolutionary pressure favoring form *"a"*, which kept the ancestral capability to cope with pH stress conditions, an ability that was apparently lost by forms *"b"* and *"c"*.

We thus observed fairly similar declines of all types under the three stress conditions (osmotic, heat and acidity stresses), with major differences being seen between two groups, i.e. (1) form *"a"* and the ancestor versus (2) forms *"b"* and *"c"*. Apparently, the evolved genomic/regulatory systems of forms *"b"* and *"c"* did not include important strategies to withstand stress, whereas the ancestral capacity to deal with stress was kept by form *"a"*.

Oxidative stress resistance in forms *"a", "b"* and *"c"*

We then focused on the response of form *"a"* (versus *"b"* and *"c"*) to oxidative stress, which had shown a striking behavior under this stress condition (Fig 5.1). First of all, the similarity between the response of the ancestor and form *"a"* was also seen in the

challenge with exponential-phase cells. Forms "*b*" and "*c*" were the more sensitive forms in both growth phases. In contrast, the behavior of the ancestor cells and form "*a*" differed remarkably during stationary phase. Whereas the ancestor's population declined rapidly after the first h of oxidative stress (with a loss of around 60% of the population), form "*a*" revealed a reduction of only about 20%. After three h, the ancestor was practically eliminated, whereas form "*a*" still revealed a population of about 20% of the initial number.

This striking difference in response of form "*a*" confirmed it had evolved to an improved ability to deal with oxidative stress conditions. Often, such adaptive responses are due to changes at the regulatory level, being generally controlled by master regulators. Under a variety of stress conditions, bacteria may convert to a transient mutator state, where they are prone to genetic change important for adaptive evolution (Foster, 2007). We (Puentes-Tellez *et al*. 2013a) found a suite of non-consistent (not occurring in all forms) mutations. Concerning the three forms in the population, form "*a*" carried a raised number of mutations (67), whereas forms "*b*" and "*c*" had 47 and 45, respectively. Table S5.2 lists the non-consistent mutations present in the three forms. Interestingly, when the mutations between forms "*b*" and "*c*" are compared, we observed only few differences between them, i.e. form "*b*" has an unique mutation in the *yhhZ* gene (a protein of unknown function, Rudd, 2000), whereas form "*c*" had two mutations in different loci of gene *rfc* (encoding a protein involved in the production of lipopolysaccharides in the outer membrane, Lukomski *et al*., 1996) and a mutation in *yfiF* (encoding a protein of unknown function, Rudd, 2000). The impact of these differences should be further evaluated. However, among the genetic commonalities between forms "*b*" and "*c*", there is a mutation (not present in form "*a*") in the key major stress regulator *rpoA* (the gene encoding the α subunit of RNA polymerase). The mutation is located in codon 294 and leads to a change of the codon from asparagine to histidine in RpoA. The mutation is located at the OxyR protein contact site in the C-terminal region of RpoA. Interestingly, OxyR is a positive regulator of hydrogen peroxide-inducible genes in *E. coli*, and such activation has not been found in mutants containing a C-terminal-truncated α subunit (Tao *et al*., 1993, 1995). Thus, a changed efficiency of OxyR protein synthesis might be involved in the observed lowered resistance to oxidative stress in forms "*b*" and "*c*". When functional annotation was performed on the list of "unique" genes with mutations

in form *"a"* using DAVID v. 6.7 (Huang *et al*., 2009), we observed two main clusters of GO terms, genes involved in membrane processes and genes involved in DNA binding processes; these processes are likely involved in the oxidative resistance observed in this form. The lack of the *rpoA* mutation in form *"a"* likely yielded the oxidative resistance advantage in combination with the resulting genomic background after the evolution process.

Dynamics of coexistence between metabolic types *a* and *c*

In previous work, we singled out two metabolic types with different metabolic capabilities among the selected forms (*"a"*, *"b"* and *"c"*): "metabolic types" *a* and *c* (Puentes-Téllez *et al*., 2014b unpublished). Whereas type *c* had evolved to a fast carbohydrate consumption ability, type *a* a slow grower on LB substrates, had developed the ability to consume acetate (which is likely released by the active metabolism of type *c*). The assessment of growth of both types growing separately in LB broth showed faster exponential growth (μ=0.26) of type *c* and a subsequent earlier stationary phase compared to the slow grower (type *a*; μ=0.22). Type *a* revealed a later exponential phase and a later stationary phase (Puentes-Téllez *et al*., 2014b unpublished).

On the basis of the current data, we now understand that tolerance to stress is another major factor that differentiates types *a* and *c*. We thus decided to test the hypothesis that these two forms are co-selected in a batch culture cycling regimen, in which conditions stimulating rapid growth on LB are interspersed with those offering stress to the community. Growth was assessed over four transfers using mutants of each type having different markers and starting with equal densities. Fig 5.2a shows the dynamics of the mixed populations, as CFU/ml (including samples taken at two growth phases: Exponential-phase (5h) and stationary phase (24h)). After each transfer, after 5 h, type *c* revealed a consistent advantage over type *a* (P<0.05). However, although lower in population size, type *a* was continuously maintained in the population, and veered back in relative abundance after the stationary phase. This group behavior, with a significant difference between the two types, remained until the last day of coexistence (P<0.001) with a final ratio of 20:1. Thus there was no consistent ecological disadvantage for either

of the two types in LB broth and a balanced equilibrium appeared to exist between them, which supports the idea of coexistence through the cycling regimen.
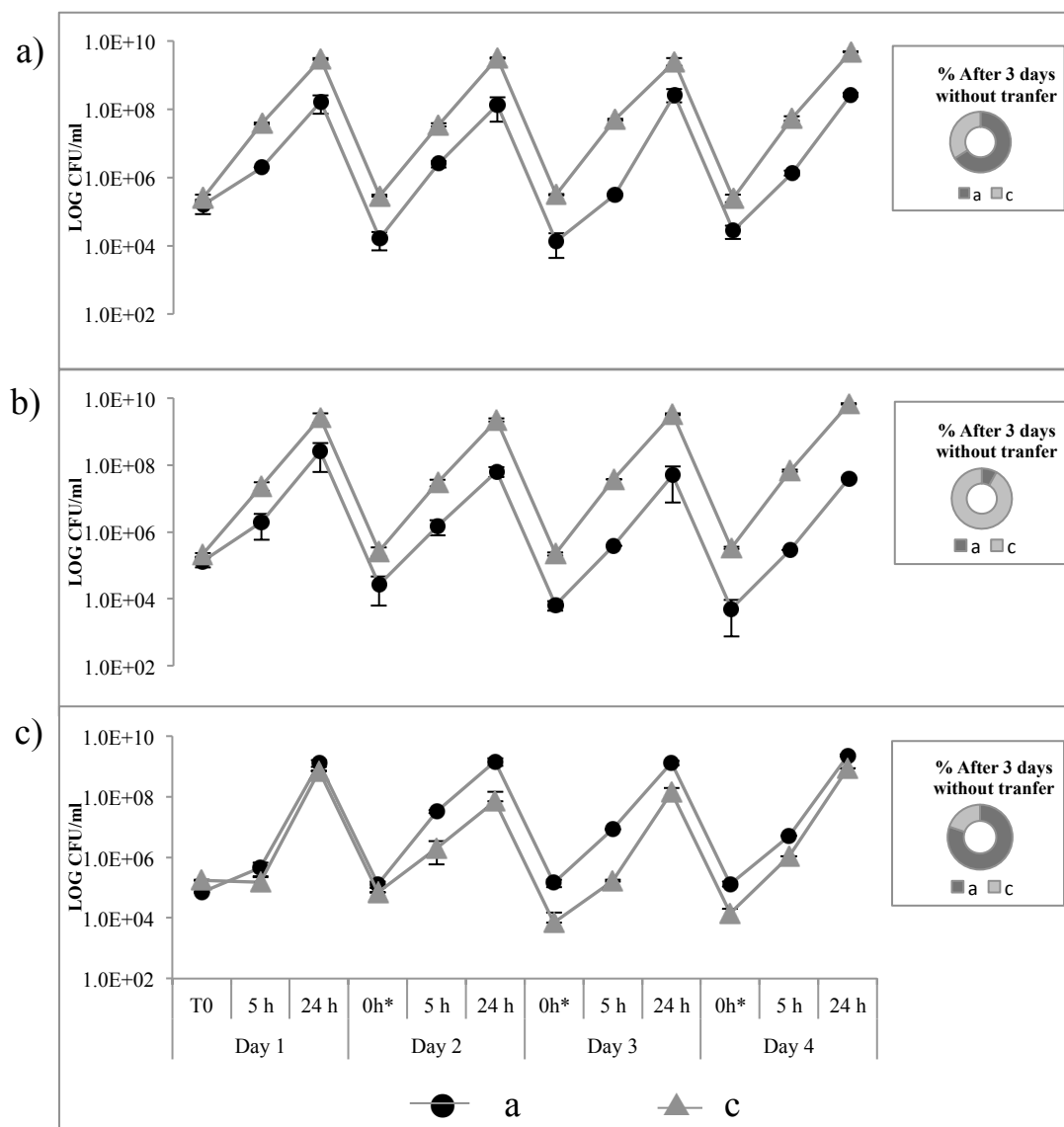
Figure 5.2. Coexistence experiment. Types *a* and *c* growing together in (a) LB broth, LB broth supplemented with (b) galactose (0.3%) and (c) H_2O_2 (0.3mM). * Predicted CFU/ml at 0h, calculated from dilution (1:10000). Error bars show standard deviation. In boxes: Percentage of each type following day 4 and after 3 days without any transfer.

Then, we aimed to observe the dynamics of the population in relation to the specific properties and stability of the coexisting pair. The stability of coexistence was tested when growing both metabolic types in "LB+galactose" (0.3%) or "LB+H_2O_2" (0.3mM) as 'inducer environments'. The results show an clear difference between these two environments. Whereas in LB+galactose, type *c* was a clear winner in terms of

population size, ending up at day 4 with a very significant difference in population size when compared to type *a* (P<0.001) (Fig. 5.2b), type *a* virtually took over the population growing in LB+H_2O_2 (P<0.05 at day4) (Fig. 5.2c). Even though there was a great influence of the introduced environmental factor, LB supported the background maintenance of each type. When growing separately in LB+galactose, a significant advantage of type *c* (μ=0.310 $\pm$SD 0.009) over type *a* (μ=0.234 $\pm$SD 0.005) (P<0.001) was found, whereas in LB+H_2O_2 there was an advantage of type *a* (μ=0.245 $\pm$SD 0.005) compared to type *c* (μ=0.185 $\pm$SD 0.014) (P<0.05). Table S5.3 shows the individual growth rates ($\pm$ SD) obtained during growth in LB supplemented with galactose and H_2O_2. These results confirmed the metabolic advantage of type *c* to deal with carbohydrates and the metabolic advantage of type *a* to deal with stress.

We then sought to observe if a longer stationary phase (and a posterior death phase) would have a selective effect on the types present in the population. In LB broth and LB+H_2O_2, type *a* almost took over the population (66 and 80.2% respectively) after 3 d of culturing without transfer. Exhausted medium presumably gives the opportunity to dominate the population to the form with raised stress resistance. Interestingly, in LB+galactose, type *a* did not persist well, on the contrary, 91.9% of the population consisted of type *c* cells. A detrimental end-effect of the galactose added to LB on type *a* was apparent from the beginning of the experiment. As type *c* has a higher-affinity for the uptake of carbohydrates, the high concentration of galactose present in the medium may have been selective for type *c* and consequently detrimental to *a*. In addition, after 5h of growth at day 1 (initial pH in all transfers adjusted to 6.9) the pH in the LB broth had dropped to 6.7 $\pm$0.03 and in LB+H_2O_2 to 6.72 $\pm$0.01. In LB+galactose, the pH had dropped even further, to 6.20 $\pm$0.02. Apparently, the accumulation of intolerable levels of acidity (possibly acetic acid) and consequent lowered pH in the galactose-containing medium was not allowing type *a* to use this carbohydrate as an optimal substrate, then the concentration of galactose added with each subsequent transfer maintains a negative effect on type *a*. In other words, if in environments with surplus carbohydrates as well as in one with stress factors the tolerance level of the more sensitive type is surpassed, equilibrium and coexistence is lost, and the form with the highest intrinsic rate of growth displaces the other.

Relative fitness comparison during coexistence in ´inducer environments´

Relative fitness (*w*) between the forms was calculated at the end of the coexistence experiment (day 4) using the ratio of their Malthusian parameters. Figure S5.2, shows the fitness of type *a* relative to type *c*. The null hypothesis is that the mean fitness of the evolved forms is one. The results confirmed the metabolic advantage of type *c* in LB and LB supplemented with galactose (*w*=0.78 ± 0.003 and *w*=0.51 ± 0.062) and that of type *a* when grown under stress conditions (*w*=1.20 ± 0.027). The two environments are clearly differentially selective, our results suggesting differential niche occupation and the presence of internal *trade-offs* within the population.

Differential transcript levels of central metabolism during exponential growth phase

We sought to observe the putative differences in transcription patterns between the "metabolic types" *a* and *c*. Thus, the activities of specific genes and pathways were tested during growth in LB broth and in the two inducer environments. For this, eight primer sets were used to target cDNA generated from purified mRNA obtained from exponential- and stationary-phase cells. We found that the data based on copy numbers per ng of mRNA were comparable to those analyzed as copy number of mRNA/cell. Thus, genes related to carbohydrate metabolism and aerobic respiration (*galE, treB, ptsH, cyoA, acs*) were tested (Fig. 5.3a). There were striking differences between the expression levels in the three environments. As part of the gal operon, *galE* catalyzes the inter-conversion of UDP-galactose and UDP-glucose during galactose catabolism. A consistent point mutation was found previously in the evolved forms *"a", "b"* and *"c"* of this population, reverting a mutated *galE* gene in the ancestral strain (Puentes-Téllez *et al*., 2013a). Moreover, a point mutation in the galactose repressor *galR* was observed, which likely allowed galactose metabolism (Puentes-Téllez *et al*., 2013a). The metabolic advantage of type *c* over type *a* when using galactose was confirmed by the increased mRNA copies in both LB and galactose supplemented medium (*P*<0.05 and *P*<0.01) during the exponential growth phase. Moreover, in this phase there was higher mRNA copy number of gene *treB* in the population of type *c* growing in LB (*P*<0.01). The product of gene *treB* is involved in the transport of trehalose, moving trehalose 6-phosphate into the cytoplasm (Horlacher *et al*., 1997). It belongs to the

phosphotransferases system group (PTS), a group of enzymes involved in sugar transportation (Klein *et al*., 1995). Interestingly, there is no difference in the inducer (galactose) environment suggesting an equal preference from type *c* relative to type *a* for trehalose.

Another PTS, the *ptsH* gene product, is involved in non-sugar-specific uptake and transport in bacteria, resulting in phosphorylation of a number of carbohydrates in *Escherichia coli (*Kornberg*,* 1986*;* Anderson *et al.,* 1971, Postma *et al.,* 1993*).* Our results revealed a significant difference in the number of *ptsH* transcripts between the two types during exponential phase in the three tested environments. Type *c* shows a higher number of transcripts compared to type *a* (*P*<0.05). On the other hand, being an indicator of higher respiration and faster metabolism (Nakamura *et al*., 1990), the raised levels of *cyoA* transcripts (compared to type *a*) confirmed the raised activity of type *c* during exponential phase (*P*<0.05). The early activation of carbohydrate-related genes confirmed the fine tuning of type *c* with respect to rapid carbon metabolism and respiration. On another note, the transcript levels of the *acs* gene in cells from exponential and even stationary phase were significantly raised in type *a* as compared to type *c* in the environments were type *c* has a growth advantage (LB and LB+galactose). The *acs* gene produces acetyl CoA synthetase (ACS), bringing acetate into the central metabolism (Brown *et al*. 1977). Thus, the consumption of acetate was confirmed by a higher expression of *acs* gene in type *a* and not in type *c*. Overexpression of *acs* gene has been previously demonstrated in polymorphic populations growing in simpler environments (Kinnersley *et al*., 2009).

Types *a* and *c* did not reveal significant differences in the transcript levels of the carbohydrate-related genes when grown under H_2O_2 stress. This could be linked to the high impact that the stress factor introduces in their general metabolic activities; as an exception, there was a significant difference in the number of transcripts related to *ptsH* (*P*<0.05), which suggested a persistent initial response of type *c* under this stress. This coincides nicely with the persistent response observed in the first day of the coexistence experiment in this environment (Fig. 5.2c).

a)
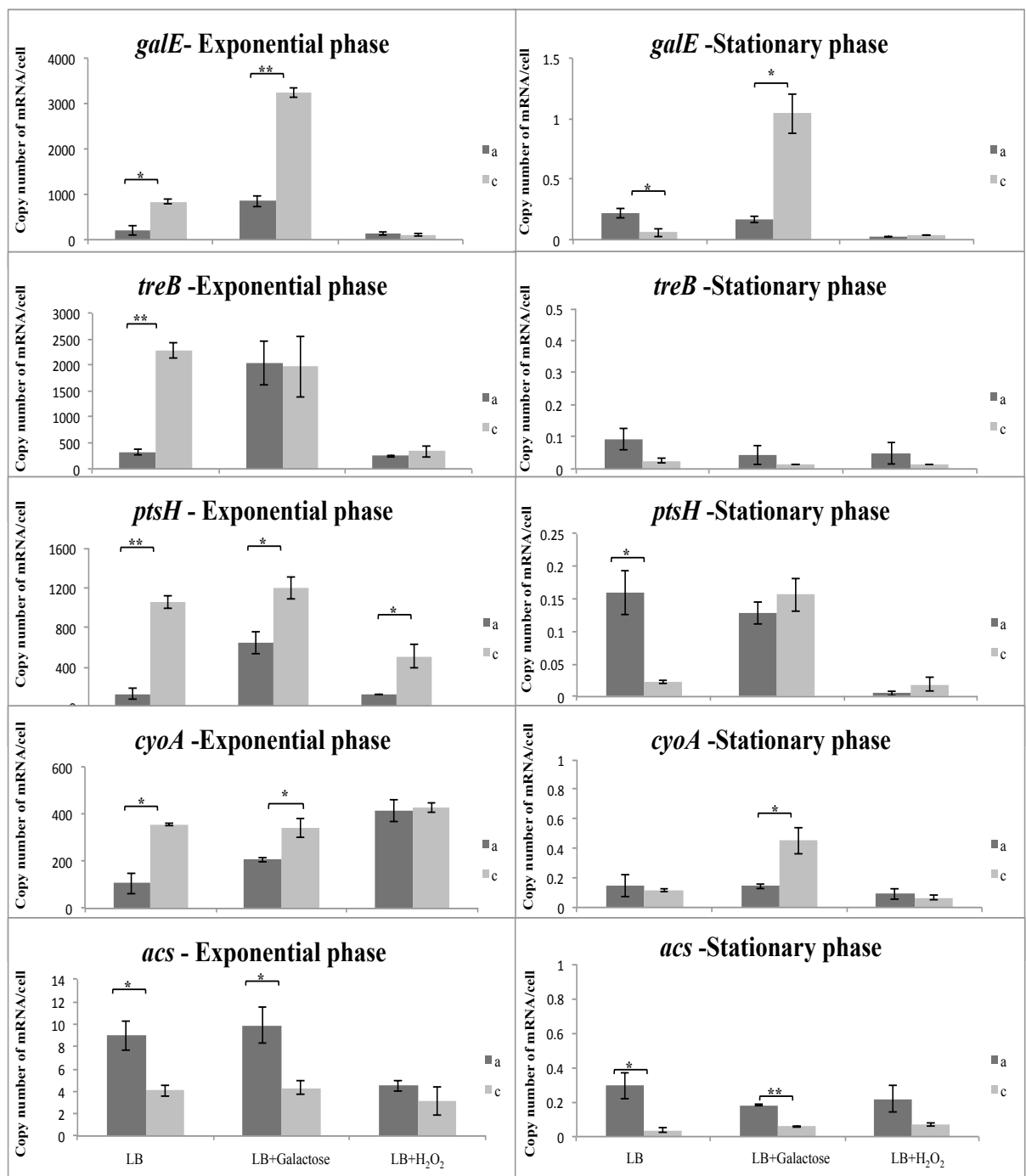
b)
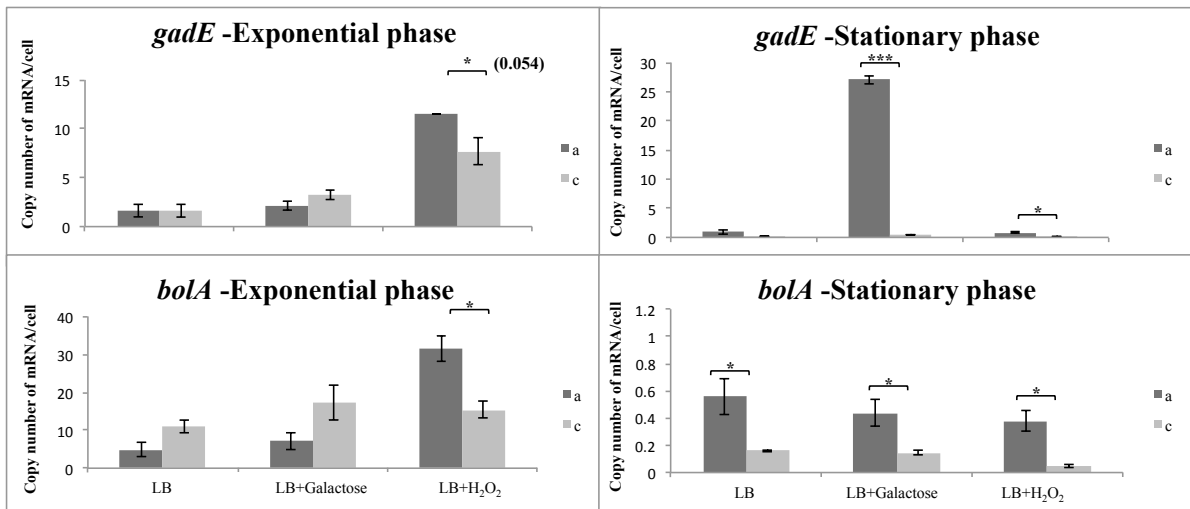
Figure 5.3. Distinct transcriptional levels of central metabolism and stress-related genes.
a) Copy number of mRNA/cell of five selected central metabolism related genes. mRNA from cells in two different growth stages and grown under three different environments (LB broth and LB supplemented with 0.3% of galactose or with 0.3mM of H2O2) was used in the Real-time PCR reaction. b) Results obtained in two stress-response related genes (*gadE* and *bolA*). Two-tailed t-test. Three different cut-off P values were established in order to observe with more detail differences between the types (P<0.05 (*), P<0.01 (**), P<0.001 (***)).

Distinct transcriptional levels of central metabolism genes during stationary phase

Stationary-phase cells revealed a lower activity of the selected genes compared to cells in exponential-phase. Specifically, *galE* transcript levels were significantly higher in type *c* in LB with galactose than in the other two environments (P<0.05) (Fig. 5.3a). This might be due not only to the preference of this type for galactose but also to the poor performance of type *a* in LB+galactose. The rapid utilization of galactose by type *c* was confirmed by the high levels of aerobic respiration (measured by *cyoA*) observed during growth with galactose. On another notice, *galE* was apparently activated in the later stage of growth of type *a* in LB broth which is confirmed by the higher levels of *ptsH* in this type during stationary phase (in both LB and LB+galactose). The consumption of trehalose at this point of growth was low for both types *a* and *c*, which corresponds to the early consumption of this sugar during LB growth (Baev *et al*., 2006c). While having a faster response during exponential-phase (relative to type *a*), type *c* apparently lowers carbohydrate transport at stationary phase when carbohydrates are mostly depleted.

Differences when dealing with stress (*gadE* and *bolA* genes)

When cells approach the stationary phase, they start facing several adversities, triggering the activation of specific and general stress response genes. As an RpoS-dependent gene, *gadE* is mainly active at stationary phase, although low levels of RpoS have also been found in exponential growth phase (Dong *et al.*, 2009), implying the activation of *rpoS*-dependent genes. The *gadE* gene product is a regulator of several genes required in the maintenance of pH homeostasis (Hommais *et al.*, 2004). We observed a lowered expression of this gene during exponential growth phase and no differences between the two types in LB and/or LB+galactose (Fig. 5.3b). However, higher and differential levels (*P* value =0.054) were observed during early growth in LB+H_2O_2, suggesting an early negative effect of this environment on both types, but a better response of type *a* to this condition, which was maintained until stationary phase. A remarkable increase of the *gadE* mRNA levels was observed in stationary-phase type *a* cells in LB+galactose, possibly linked to the crossed protection response during adverse conditions. During stationary phase, the levels of environmental stress increase and many changes at physiological and genetic level can occur, our results suggest that type *a* has potentially developed a growth advantage in stationary phase (GASP phenotype) (Zambrano *et al.*, 1993).

On the other hand, the *bolA* gene product has been found to respond to several types of stress (osmotic, oxidative, heat, carbon starvation and acid stress), partially independent from *rpoS*. Although there were no significant differences between the types in LB broth and LB+galactose, our results show *bolA* mRNA levels to be higher during the exponential growth phase, surpassing the levels in stationary phase cells. This response agrees with previous findings, in which *bolA* was induced upon adverse conditions in early growth (Santos *et al.*, 1999). A higher level (P<0.05) was observed during growth with H_2O_2 and in all environments in stationary phase. Although lower transcript levels of this gene were observed during the last growth phase, significant differences between types were observed, meaning a differential capacity to deal with stress at this stage of growth.

Altogether, the results reported here suggest the presence of *trade-offs* in both types involving an activation of carbohydrate uptake and consumption by one type as well as an active survival stress response in the other one. These two distinct properties are beneficial to either of the two forms and contribute to maintain a "balanced" coexistence within the population, in which events of substrate abundance are temporally interspersed with those of stress.

Suggestive effects of differential arcA gene mutations

Previous work on the selected population found a key differential genetic change after ~1000 generations of evolution under aerobic conditions *(*Puentes-Téllez *et al., 2013a)*. Type *c* colonies carry a mutation in the *arcA* gene (A76V, G->A), whereas type *a* did not. As a major regulator (controlling the activity of about 168 genes), and making part of a two-component regulatory system, *arcA* is involved in the repression of aerobic metabolism (aerobic respiration, TCA cycle genes) upon anaerobiosis. Under aerobic conditions, *arcA* remains inactive, allowing the normal functioning of aerobic respiration genes (Iuchi *et al*., 1996). Active *arcA* has also been found to down-regulate and repress *rpoS* in exponential-phase cells (Mika *et al*., 2005, Hengge, 2008). We found a correlation of the effect on *rpoS* related genes with the non-mutant (inactive *arcA*) type *a*, i.e. the transcription levels of genes that are positively regulated by *arcA*, such as those involved in pH homeostasis, *gadE, gadA, hdeA* (data not shown), are significantly higher in this metabolic type. There might be a correlation with the surprisingly increased levels of *arcA* in type *c* and the repression of *rpoS* genes in this type (Fig. 5.4).
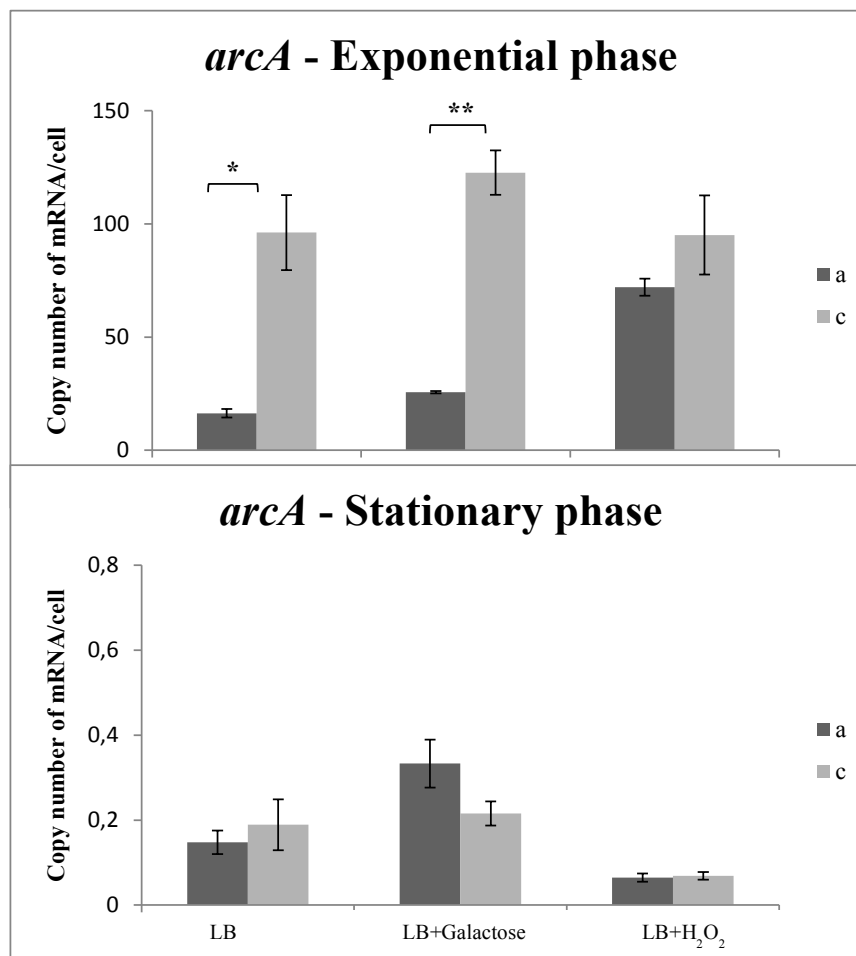
Figure 5.4. *arcA* transcriptional assessment. Transcription levels of *arcA* gene.

In addition, deletions in *arcA* may be related with hydrogen peroxide sensitive strains (Wong *et al.* 2007, Gonidakis *et al.*, 2011) and also allow to survive under prolonged starvation (Nystrom *et al.*, Sevcík, *et al.*, 2001). There were remarkable differences in gene expression by type *c* during growth in LB and LB+galactose during exponential-phase, and in contrast the expression levels in stationary phase remained equally low (Fig. 5.4). Moreover, the *cyoA* gene transcription data suggested that the effect of this mutation might have an influence in the respiratory pathways taken by type *c*; a non-mutated *arcA* would be a negative regulator of this gene. Moreover, high levels of *arcA* transcription would repress aerobic respiration. Overall, there is a putative effect of this gene on the resulting evolved phenotypes. A direct and indirect effect of *arcA* has been recently found on >350 genes, however the full extent of the ArcA regulon remains unclear, preventing a comprehensive understanding of its physiological role (Park *et al.*, 2013).

Conclusions

We unraveled the phenotypic differences among emergent types of a population of *Escherichia coli* that has evolved during ~1,000 generations by discriminating the observation of phenotypic traits in two different cell-growth physiological states. By using this approach, we observed that adaptive evolution had a consequence in the survival properties of members of the resulting population, and the scope of this adaptation influenced the survival capabilities of each metabolic type. Distinctive responses to stress were observed, which corresponded to a raised ability of one of the types to deal with stress from an early stage of growth. Oxidative stress resistance seems to be particularly favored in this form, which might be related to genomic changes affecting specific regulatory characteristics. In addition, a mutation found at a specific contact site of *rpoA* suggests negative consequences on the upregulation of hydrogen peroxide-inducible genes in the sensitive forms.

Coexistence between these forms was observed during growth in LB and the stability of coexistence was lost when selective pressure favors one of the types. The majority of differences at the regulatory level between the types was observed in the early stage of growth and these are suggestive of niche partitioning and specific internal beneficial *trade-offs* with differing roles, maintaining stability in the coexisting population.

Acknowledgments

References

Anderson B, Weigel N, Kundig W, Roseman S. (1971) Purification and properties of a phosphocarrier protein (hpr) of the phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*. J Biol Chem 246 (22): 7023-7033.

Baev MV, Baev D, Janeso A, Campbell JW. (2006c) Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols and organic acids with transcriptional microarrays. Appl. Microbiol. Biotechnol. 71:310 –316.

Brown TD, Jones-Mortimer MC, Kornberg HL. (1977) The enzymatic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. J. Gen. Microbiol. 102:327–336.

Buckling A, Wills MA, Colegrave N. (2003) Adaptation limits diversification of experimental bacterial populations. Science 302(5653):2107-9.

Dong T, Schellhorn HE. (2009) Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933 BMC Genomics 10:349.

Dragosits M, Mozhayskiy V, Quinones-Soto S, Park J, Tagkopoulos I. (2013) Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in *Escherichia coli*. Mol Syst Biol. 9:643.

Foster PL. (2007). Stress-Induced Mutagenesis in Bacteria. Crit Rev Biochem Mol Biol: 42(5) 373–397.

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(2): 245–260

Gause GF. (1934) The Struggle for Existence. Williams and Wilkins, Baltimore.

Gonidakis S, Finkel SE, Longo VD. (2011) Lifespan extension and paraquat resistance in a ubiquinone-deficient *Escherichia coli* mutant depend on transcription factors ArcA and TdcA. Aging (Albany NY) 3(3): 291-303.

Groat RG, Schultz JE, Zychlinsky E, Bockman A., Matin A. (1986) Starvation proteins in *Escherichia coli*: kinetics of synthesis and role in starvation survival. J. Bacteriol. 168(2):486.

Helling, R, Vargas C, Adams J. (1987) Evolution of *Escherichia coli* during growth in a constant environment. Genetics. 116(3): 349–358.

Hengge R. (2008) The Two-Component Network and the General Stress Sigma Factor RpoS (σS) in *Escherichia coli*. Adv Exp Med Biol.;631:40-53.

Hommais F, Krin E, Coppée JY, Lacroix C, Yeramian E, Danchin A, Bertin P. (2004) GadE (YhiE): a novel activator involved in the response to acid environment in *Escherichia coli*. Microbiology. 150(Pt 1):61-72.

Horlacher R, Boss W. (1997) Characterization of TreR, the Major Regulator of the *Escherichia coli* Trehalose System. J. Biol. Chem., 272:13026-13032.

Huang DW, Sherman BT, Lempicki RA. (2009) Systematic and integrative analysis of

large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44 –57.

Iuchi S., Weiner L. (1996) Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments. J Biochem. 120(6):1055-63.

Kanno J., Aisaki K., Igarashi K., Nakatsu N., Ono A., Kodama Y., Nagao T. (2006) "Per cell" normalization method for mRNA measurement by quantitative PCR and microarrays. BMC Genomics, 7:64

Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity. Science 15: 173–190

Klein W, Horlacher R, Boos W. (1995) Molecular analysis of *treB* encoding the *Escherichia coli* enzyme II specific for trehalose. J Bacteriol. 177(14):4043-52

Kornberg H. (1986) The roles of HPr and FPr in the utilization of fructose by *Escherichia coli*. FEBS Lett. 194(1):12-5.

Lenski RE, Rose MR, Simpsom SC, Tadler SC. (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138(6):1315–1341.

Liss, B. (2002) Improved quantitative real-time RT–PCR for expression profiling of individual cells. Nucleic Acids Res., 30, E89.

Lukomski, S Hull R, Hull S. (1996) Identification of the O antigen polymerase (*rfc*) gene in *Escherichia coli* O4 by insertional mutagenesis using a nonpolar chloramphenicol resistance cassette. J Bacteriol. 178(1): 240–247.

MacLean RC., Dickson A, Bell G. (2005) Resource competition and adaptive radiation in a microbial microcosm. *Ecol Lett* 8: 38–46

Kinnersley M, Holben W, Rosenzweig F. (2009) *E Unibus Plurum*: Genomic Analysis of an Experimentally Evolved Polymorphism in *Escherichia coli*. Plos Genetics. 5 (11): e1000713

Mika F, Hengge R. (2005) A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of σS(RpoS) in E. coli. Genes Dev. 19(22):2770-81.

Nakamura H, Yamato I, Anraku Y, Lemieux L, Gennis RB. (1990) Expression of *cyoA* and *cyoB* demonstrates that the CO-binding heme component of the *Escherichia coli* cytochrome o complex is in subunit I. J Biol Chem. 265(19):11193-7.

Nystrom, T., Larsson, C. & Gustafsson, L. (1996) Bacterial defense against aging: role of the *Escherichia coli* ArcA regulator in gene expression, readjusted energy flux and survival during stasis. EMBO J 15 (13):3219-3228.

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. doi:10.1371/journal.pgen.1003839.

Postma PW., Lengeler JW. (1993) Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. 57(3):543.

Puentes-Téllez, P.E., Hansen, M.A., Sørensen, S.J., and van Elsas, J.D. (2013a) Adaptation and heterogeneity of *Escherichia coli* MC1000 growing in complex environments. Appl Environ Microbiol 79: 1008–1017.

Puentes-Téllez, P.E., Kovacs A.T., Kuipers O.P., van Elsas, J.D. (2014a) Comparative genomics and transcriptomics analysis of experimentally evolved *Escherichia coli* MC1000 in complex environments. Environ Microbiol. 16(3):856–870.

Puentes-Téllez, P.E., van Elsas, J.D. (2014b). Sympatric metabolic diversification of experimentally evolved *Escherichia coli* in a heterogeneous environment. *In Submission.*

Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature 32: 69–72.

Hengge-Aronis R. (2002) Signal Transduction and Regulatory Mechanisms Involved in Control of the σS (RpoS) Subunit of RNA Polymerase. Microbiol Mol Biol Rev. 66(3): 373–395.

Rosenzweig RF, Sharp RR, Treves DS, Adams J. (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics 137(4): 903–917.

Rudd, KE (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. Nucleic Acids Res 28:60-4.

Santos J., Freire P., Vicente M., Arraiano CM. (1999) The stationary-phase morphogene bolA from *Escherichia coli* is induced by stress during early stages of growth. Mol Microbiol. 32(4):789-98.

Sevcík M, Sebková A, Volf J, Rychlík I. (2001) Transcription of *arcA* and *rpoS* during growth of *Salmonella typhimurium* under aerobic and microaerobic conditions. Microbiology 147(3): 701-708.

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(5): 1585-1589.

Tao, K., Fujita N, Ishihama A. (1993) Involvement of the RNA polymerase a subunit C-terminal region in cooperative interaction and transcriptional activation with OxyR protein. Mol. Microbiol. 7:859–864.

Tao K, Zou C, Fujita N, Ishihama A. (1995) Mapping of the OxyR protein contact site in the C-Terminal region of RNA polymerase a Subunit. J Bacteriol 177(23):6740-4.

Tyerman J, Havard N, Saxer G, Travisano M, Doebeli M. (2005) Unparallel diversification in bacterial microcosms. Proc Biol Sci. 7;272(1570):1393-8.

Wolfe AJ. The Acetate Switch (2005) Microbiol. Mol. Biol. Rev. 69(1):12.

Wong ML, Medrano JF. (2005) Real-time PCR for mRNA quantitation. Biotechniques. 39(1):75-85.

Wong SM, Alugupalli KR, Ram S, Akerley BJ. (2007) The ArcA regulon and oxidative stress resistance in *Haemophilus influenzae*. Mol Microbiol. 64(5):1375-90.

Zambrano MM, Siegele DA, Almiron M, Tormo A & Kolter R. (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science* 259: 1757-1760.

Chapter 6

General Discussion, Conclusions

and Future Perspectives

The complexity of many habitats offers a wide range of variation in conditions for living micro-organisms. Understanding the bacterial response to these settings is crucial for our understanding of the breadth of their adaptability (Elena *et al.*, 2003). In the last decades, light has been shed on the plethora of bacterial abilities to cope with the environment. A main observation coming from manipulated studies is that the environmental tolerance boundaries provide borders to the adaptive capacity. Studies of evolutionary trajectories and outcomes addressed the dynamics of adaptation and evolution in specific environments and setups. Most such experiments on microbial evolution are conceptually simple and compare populations adapting to uniform and "basic" environments. Hence, the effects of environmental heterogeneity, an important factor in the evolution of populations, have not been frequently addressed (Cooper *et al.*, 2010). Thus, in the present work, we hypothesized that the simplicity of most experimental setups do not match the complexity of natural or even man-made environments. Therefore, studies on bacterial adaptation and its dynamics in complex conditions are in high demand (Dettman *et al.*, 2012).

In complex or heterogeneous environments, alterations of conditions could occur as related to the factors space (spatially structure and unstructured) and time (temporal and constant). Variations in local conditions, like uneven (access to) resources, pH, oxygen tension or sudden disturbances, turn the environment into an uneven and heterogeneous place for bacterial growth. These varying conditions create empty niches and their positioning and breadth may evolve, establishing environmental variation that drives bacterial selection (Kassen *et al.*, 2002). In addition, an often disregarded source of heterogeneity is the fine-scale-structure creating niche availability and caused by the presence of nutrient mixtures in the system.

Access to whole genomes has provided a comprehensive view of genomic (and inferred phenotypic) changes that underlie evolution (Dettman *et al.*, 2009). The long-sought link between genotype and phenotype may thus be becoming more accessible. However, even though mechanistic and integrative links between genomic data can now be associated with ecologically relevant phenotypes, the complexity of local conditions is

often of such magnitude that an elucidation of the total picture is very difficult (Autumn *et al*., 2002; Dalziel *et al*., 2009).

In the light of the foregoing, this study attempted to elucidate the adaptation and diversification of *Escherichia coli K12* MC1000*,* an organism that is extensively used in large-scale fermentation processes, in a complex environment involving sequential (serial)-batch cultures with multiple energy sources. The propagation scheme ensured that the organism experienced a "feast and famine" environment with an abundance of initial resources that are exhausted as the population grows. Thus, the experimental setup encompassed exposure of the organism to two constantly cycled phases of growth: a short exponential growth phase and a long stationary phase under three different/independent oxygen regimes. These conditions mimic the high degree of environmental pressures and gradients that organisms can face under industrial settings. Thus, the importance of this work from an industrial point of view lies in assessing if maintenance of optimal performance (high metabolic capacity of the population and resistance to the global and local environmental cues) might be feasible, taking into account that industrial populations are influenced by a vast number of biotic and abiotic factors.

Although the observation of the shape of fitness landscapes and peaks can reveal the trajectory of natural selection (Colegrave *et al*., 2005), the scope of our work was not to fully evaluate the trajectory taken by *Escherichia coli* in its adaptive process, but instead to "zoom in" (as a common "landmark") into the end result after ~1,000 generations. The experimental setup involved the incorporation of complexity in the medium used and the inclusion of constant as well as temporal variation of the factor oxygen. Thus, the propagation regimes included serial transfers in aerobic and anaerobic environments, next to cycling transfers between these.

The overall theme of this study was the use of a multiple-energy-source background without the establishment of a clear limiting resource. This is to our knowledge, the first report assessing adaptation using Luria-Bertani broth (LB). Adaptation and heterogeneity under these complex conditions were assessed in an integrative way (genome sequencing and transcriptomic and phenotypic studies) to elucidate the outcome of evolution under these conditions and to assess the main drivers

and the scope of adaptation and diversification. In this setup, I simulated "some" of the factors that affect bacterial growth, most importantly factors such as supply or scarcity of oxygen as also found in artificial (industrial) systems, in which bacteria can be exposed to short growth phases and long stress states with a vast mix of substrates.

The main driver of adaptation in complex environments: its extent and specificity

In experiments designed to test the effects of environmental variability on the adaptation of replicate populations of *Escherichia coli*, similar trends in terms of specialization of newly-emerged forms came up. In a range of experiments, it was shown that selection tends to lead to the evolution of specialists in so-called "constant" environments, whereas it tends to lead to the evolution of generalists in temporally-variable environments (Bennett *et al*., 1992, Leroi *et al*., 1994, Hughes *et al*., 2007). In **chapters 2** and **3,** I examine the genomic and phenotypic outcomes of evolution in the established complex experimental setup. I was interested in understanding whether heterogeneous resource environments (complexity) would influence the repeatability of evolutionary routes. Moreover, I observed the extent of adaptation under these conditions. The data obtained agree with results that were previously obtained from experiments in simpler environments. Bluntly speaking, the environment appeared to be the main driving force of adaptation and enhanced fitness. In broad terms, the fluctuating environment gave rise to the evolution of generalists whereas the constant environments yielded more specialists, trading enhanced fitness in their evolving environment for loosing capacities in alternative ones. . The emerged forms invariably revealed increases of fitness, but the level/scope of this fitness showed different extents across the environmental regimes. If adaptation had occurred only in response to shared aspects of the environment (like the medium or the temperature), then the magnitude of fitness gains was hypothesized to be indistinguishable among the treatment groups. In our results, both constant environments gave rise to fitter forms. However, the globally structured environment (constant/static condition) revealed an outstanding increase. Thus, both constant and perhaps more stressful environments incited the emergence of more specialized, positive, adaptive responses than the fluctuating environment, which incited more generalist behavior. The number of mutations and positive (active) gene expression changes correlated with an enhanced capability to deal with environmental pressures, since the populations from

environments with higher numbers of changes were the fittest. The resulting higher fitness of the evolved forms in these environments can be due to the presence of specific fitness peaks present in these populations. A constant environment may imply that these populations evolved on a more rugged (*i.e.* multiple peaks, constantly changing in height) fitness landscape (Colegrave *et al.*, 2005). Thus, given sufficiently large population sizes and high mutation rates, many cells in the *E. coli* population could concurrently acquire beneficial mutations of varying adaptive (fitness) values. These lineages are thought to compete with the ancestral cells as well as between themselves for fixation ('clonal interference'), the latter adding a constraint on adaptation (Jayaraman, 2011).

Functional annotation of the phenotypic (gene expression) changes revealed the expected impact of the constant aerobic and anaerobic environments since the vast majority of changes in the evolved forms were related to oxygen availability or depletion. Specifically, in aerobiosis, a more active regulation of processes related to cellular respiration/ central metabolism had occurred, whereas in the anaerobic treatment the changes in terms of altered (upregulated) gene expression were related to carbohydrate-active transport systems and acid-regulated pathways. However, a more striking response occurred under both regimes. More specific and specialized changes in accordance with the oxygen availability included a more 'alleviating' response, with the activation of several stress response systems. On the other hand, responses in the fluctuating treatment were phenotypically more polymorphic and less specific.

The results obtained in this study revealed that being a generalist in complex conditions brings along a cost, since all fitness increases in the fluctuating environment were the lowest across the board. Thus, the oxygen variation applied seems to have a stronger (unfavorable) influence on fitness, which is an undesirable characteristic when oxygen gradients are found in industrial setups. Possibly the factor time in our experimental setup (daily transfers into a new type of environment) included in this treatment did not allow enough "space" to more beneficial mutations to be fixed. The scale of temporal variation was perhaps not sufficiently broad or heavily enough weighted towards transitions and increased fitness (Leroi *et al.*, 1994). These striking results incited us to assess the fitness in the populations as-a-whole, as a useful measure of the costs or benefits of being generalists (Buckling *et al*, 2007). We found increases

of, on average, 18% in mean fitness values when all members of the *E. coli* population were competing against the ancestor. This suggested the occurrence of possible interactive strategies in these populations, which allowed them to cope with the transitions of the environment. The results suggest that early fixation of beneficial changes may have occurred, giving to these populations particular advantages (although more limited when relative to the other regimes) when compared to the ancestral strain, and also the establishment of possible interactive strategies.

Effect of the environment on the patterns of divergence

The results presented here describe diversification based on two different aspects of the cell, *i.e.* genotype (genetic differences) and phenotype (fitness as measured by growth in competition, gene expression levels, metabolic and survival capacity). Profound differences between the evolved forms and the ancestor were found in all aspects. Moreover, differences between the evolved forms were also detected, at the inter-populational level as well as within the populations. Very few studies have addressed phenotypic diversification as related to fitness values (Cooper *et al*., 2010, Puentes-Téllez *et al*., 2013a). Here, I obtained data that emphasize the importance of analyzing fitness outcomes by considering *first* any phenotypic (or perhaps genetic) characteristic in the end-point populations from an evolution experiment. From the results obtained, I suggest that averaged results from an initial "colony-pooling" method, aliquots of mixed end-point populations or randomly selected isolates, which are methods used by many researchers, implicitly assume homogeneity in the population and works to the exclusion of diversification biasing the results. Also, statistical variations in the averaged fitness results obtained from randomly selected- isolated colonies should be carefully observed. In cases in which phenotypic (or genetic) characteristics are apparently similar, the assumption of divergence between colonies during fitness analysis (following prior isolation of colonies) is clearly more accurate.

The data on the individual fitness gains obtained by the evolved *E. coli K12* MC1000 forms were not consistent with the findings of Cooper *et al* (2010), who found that fitness variation was greater in populations of fluctuating than in those of constant environments. In our study, most of the variation in fitness between populations of the

same environment was observed in the static condition, which agrees with theories on adaptive radiation in structured environments with higher environmental heterogeneity (Korona *et al*, 1994, Rainey & Travisano, 1998, Kassen *et al*., 2004). It is worth to mention, that ecological specialization - like in this case - and the occurrence of fitness *trade-offs* across environments has been found to be a necessary but insufficient condition for the maintenance of diversity in a heterogeneous environment (Kassen, 2004).

As mentioned above, genomic and phenotypic diversification was rampant at the inter- and intra-populational level. Compared to other studies, our study revealed an extreme number of genomic and phenotypic changes in the evolved *E. coli* forms, and we attribute this to the complexity of our experimental setting, involving a complex medium containing a great variety of energy sources.

The work provided evidence for the contention that the multiple niches that are likely present in a complex medium like LB generate "spatial structure" heterogeneity, resulting in a plethora of distinct microenvironments. Moreover, such heterogeneity and the emergence of niches can be exacerbated by the action of the bacteria themselves, so that theoretically the number of niches can even increase. The availability of increased numbers of niches in complex environments has in fact been demonstrated in systems under different scenarios (Korona *et al*., 1994, Rainey & Travisano, 1998, Traverse *et al*., 2013). If the environment is heterogeneous, variants with increased mutation rates may tend to be selected, given the increased probability of mutations being selectable by any locally emerging niche. Moreover, it is known that mutation rates can increase genome-wide under certain conditions, such that both non-adaptive and adaptive mutations are enhanced ("hypermutation"). Such elevated mutation rates are helpful to organisms to adapt to sudden and unforeseen threats. Some populations in adaptive evolution experiments have been shown to become dominated by such hypermutators (Barrick *et al*., 2013). However, hypermutability can lead to the enhanced generation of deleterious mutations, which offsets the adaptive value of mutation and therefore becomes disadvantageous (Jayaraman, 2011). Conversely, in constant environments, as the organism becomes maximally adapted, mutation rates were hypothesized to decrease because of the costs associated with deleterious mutations (Baquero *et al*., 2004).

This work highlighted the fact that the scope of diversification in complex environments depends greatly on the local ecological conditions. For example, the *E. coli* populations in the structured (anaerobic) environment with the higher fitness variation between them had the most remarkable genomic and gene expression differences at the inter-populational level. On the other hand, the genomic and gene expression changes in the *E. coli* forms in the temporal (fluctuating) environment showed a striking result when compared between populations, as there was relatively low genetic variation among these forms but, in contrast, high phenotypic (gene expression) variation. It is clear that organisms are selected on the basis of phenotype, that is, the resultant of the expressed genes. The phenotypic variation might have set on early in time, even on the basis of just a few genomic changes. Most of the changes found in the fluctuating environment were related to the maintenance of an internal steady state, which might actually become costly to the cells, and so it might become more difficult to reach high fitness peaks.

In our *E. coli K12* MC1000 evolution study, intra-population heterogeneity was widely observed with respect to gene expression, and the metabolic and genomic characteristics of the forms that emerged in the constant environments. The plethora of different forms found was a strong indicator of diversifying selection driving the process. Unique mutations and upregulation events were observed, which were assumed to be the result of genetic drift (or experimental noise) either followed by selection or not. The role of selection is backed up by multiple niche availability offered by the multiplicity of substrates in the LB medium. The phenotypic effects of the diversity were different between the differing genomic and environmental backgrounds. We found, for example, several consistently mutated genes across all populations (like *yqhC*, *cytR*, *dnaA*, *nhaR*), which did not exert a comparative effect on the gene expression response. These results are consistent with the reported general difficulties of relating genotype to phenotype in a straightforward manner, which might be due to epistatic interactions or other confounding factors.

Impact of the growth medium on the parallel adaptive responses by *Escherichia coli K12* MC1000

The occurrence of diverse parallel genomic and phenotypic responses in independently evolving lines of *E. coli K12* MC1000 growing in sequential-batch cultures was a major striking finding of this work. Given the uncertainty of occurrence of future mutational events and their fixation, the likelihood of parallel genetic changes is thought to be small (Nakatsu *et al*, 1998). However, a certain degree of predictability in the evolutionary outcome of our emerging *E. coli* population was observed. I thus observed how selection acted on the genetic changes and shaped the outcome of evolution in parallel. Thus, selection clearly acted on what was in "need" by the growing *E. coli* population in terms of growth substrate or condition. With the results presented in this thesis, I can affirm that the apparent evolution observed after 1,000 generations of *E, coli* was most likely the result of a combination of chance (stochasticity = neutrality = drift) and need (selection).

Even though the link between the *E. coli* genotypes and phenotypes is still far from being resolved, in **chapter 3** I used the genomic and phenotypic datasets that were obtained to elucidate whether there was any connection between the selective processes that had led to the emergence of the evolved forms. Parallel and convergent changes across lineages have been proposed as hallmarks of adaptive evolution (Cooper *et al*, 2003, Woods *et al*, 2006) and thus attention to the genes and phenotypic traits that served as targets of selection is warranted (Cooper *et al*, 2008). We considered the parallel genetic and consequent gene expression data found across all evolved populations to be driven by the common environmental factor, the LB medium. This complex medium apparently had a key role in the adaptive responses seen, acting as a strong driving force that shaped the determinant phenotypic outcomes like fitness.

Parallel changes across the emerged forms were observed in important genes (major regulons). These are of special interest because of their major impact on metabolism and behavior. Global regulatory or response systems can cause sweeping changes in gene expression and cellular metabolism, the responses being mainly controlled by master regulators (Foster, 2007). As mentioned before, consistent genomic

changes were found in global regulators across all selected forms in all populations: GalR, NhaR, YqhC, DnaA, CytR. The most striking response with a clear phenotype observed across all populations was a reversion of a mutation of GalR present in the ancestor strain, which reactivated the use of galactose in the evolved forms. The occurrence of this reactivation across all environments suggested that selection acted strongly and in parallel to occupy this vacant niche (growth on galactose, reported to be found in LB (Hanko *et al*., 2004). The strength of this selection was remarkable, because the main applied environmental force of each treatment (differential oxygen availability) did not do away with it. Not all parallel changes, however, could be linked to the same phenotypic outcomes, which was possibly due to differential epistatic interactions. The genomic backgrounds of the members of the different populations and also the environmental force (oxygen availability) could have influenced the expression or not of the parallel changes.

Specific metabolic strategies as the main adaptive response of *Escherichia coli K12* MC1000 to growth in sequential-batch LB medium are related to the use of alternative carbohydrates

"Adaptive" or "stress-induced" mutations occur in bacteria exposed to growth-limiting environments, and such mutations appear to be formed in response to the environment (Hersh *et al*., 2004). They have the potential to speed up evolution in the face of adverse conditions, in case they confer a growth advantage relieving stress (Foster, 1999, Gonzalez, 2008). Stress-induced mutations may relate to pathways different from those observed in rapidly proliferating cells (Gonzalez, 2008). In chemostat cultures, bacteria are faced with perpetual resource limitation and hence transport systems for the limiting resource are the likely targets of selection. The genes that encode those systems are thus candidate loci for change (Elena *et al*., 2003).

LB medium contains a vast mixture of components, but it appears to provide a limited environment for optimal growth. As discussed by Nikaido (2009) in his text "The limitations of LB medium", this relates to the fact that, although full of other components, LB medium provides only a scant amount of carbohydrates (recent formulations lack glucose) next to small amounts of other utilizable carbon sources (amino acids). The few

carbohydrates present in LB are consumed early in growth, after which a switch to the consumption of palatable amino acids takes place (Sezonov *et al*., 2007). Contrary to what was previously thought, these characteristics make LB a stressful environment which can severely limit growth. In addition, long stationary phases occurred in the populations, in between the transfers (performed every 24h), which offered a very challenging environment in the three environments. Under these conditions, starvation for essential cell components may have taken its toll, whereas mutations involving changes in the general stress response (*rpoS*) may offer growth advantages to the relevant mutants, known as GASP (Growth advantage in stationary phase) mutants (Zambrano and Kolter, 1996).

In **chapters 2** and **4**, I found that during evolution of *E. coli* K12 MC1000 in sequential-batch LB medium, key adaptive responses occurred. The main response was related to the activation of alternative pathways converging into glycolysis. Thus, selection for enhanced metabolic efficiency likely dominated, because mutations were found that affected central metabolism. Each of these mutations appeared to expand the metabolic potential of the end-population forms and enhanced competitiveness of these as compared to earlier mutants. As previously mentioned, we found a key adaptive response related to the selection for galactose derepression that had a positive impact on fitness. Specifically, the mutational events consisted of non-synonymous changes in the repressor of the gal operon (*galR* gene) and a reversion (or "repair") event of a detrimental change affecting galactose metabolism that was already present in the ancestor (*galE* gene). The activation of this alternative pathway feeding into glycolysis appeared to exert a concomitant effect on galactose-related disaccharides present in the medium (D-lactose, D-meliobiose, and lactulose). The consistent *galR* mutation and metabolic upshift observed in the Biolog results provide ammunition for the hypothesis that these are at the basis of the consistently enhanced fitness in all evolved forms. Thus, *E. coli K12* MC1000 cells evolving in LB medium that underwent a switch to derepression of the galactose operon are amenable to selection by the force exerted by the carbohydrates present in the medium. These results are remarkable, because data gathered until now have concluded that, in general, organisms develop complex adaptive traits by globally fine-tuning gene expression rather than by locally restructuring pathways that are involved in specific traits (Hindré *et al*., 2012).

The specific genomic changes found in the galactose operon and their direct effects on fitness can be directly tested by manipulating these mutations in the evolved forms and by assessing the fitness effects of this manipulation in the selecting environment. The typical approach of genetically manipulating the ancestor (Elena *et al.*, 2003) would not work in this case (or would not be easily achievable) because of the vast number of genetic changes found in our evolved forms. Manipulation of specific genes in the mutants, however, should be carefully assessed because potential epistatic interactions with other genes could be at the basis of the fitness increase.

On the other hand, systems for use of alternative sugars present in LB, that might provide an entrance into glycolysis, were upshifted in the evolved forms. The upshift of sugar utilization systems such as for N-acetyl-D-glucosamine, D-fructose, l-fucose, D-mannose, D-mannitol and β-methyl-d-glucoside, maltose, D-trehalose (coupled to glycolysis) was widely observed. Moreover, a consistent upshift in L-arabinose utilization (which provides an entrance to the pentose phosphate pathway) was observed. We conclude that the fitness increase across all populations is closely related to a switch in the metabolism of alternative carbohydrates. This most likely occurred as a strategy to cope with the scarcity of primary carbohydrates, still providing entrance to the central metabolism (glucose metabolism), as a strategy to make use of "whatever is available" in the environment.

Scope of adaptation of *Escherichia coli K12* MC1000 in sympatric conditions and the benefits of diversification

In **chapters 4** and **5**, I examined an *E. coli K12* MC1000 population that was found to have remarkably differential genomic and phenotypic characteristics. In a first examination of the phenotypic differences among members of the population (**chapter 4),** two dominating metabolically-different types, denoted types *a* and *c* were detected. The set of colonies used in the analysis revealed that colony morphology had a correlation with metabolic behavior. The observed types had developed novel strategies to deal with the substrates in LB broth. Whereas type *c* was a fast consumer of carbohydrates alternative to glucose, with a metabolic slot into the glycolysis pathway, type *a* was a slow grower on such carbohydrates (although still better than the ancestor). However, an

enhanced ability of type *a* to consume acetate was detected. These results suggested an interactive relationship between the metabolic types, in which the first form (utilizer of the available primary resources) interacts with the second type, which was a better secondary-metabolite consumer (acetate, likely present in later stages of growth as a by-product of the active metabolism of the first type) (See Figure 6.1). Future work should assess acetate production and consumption rates, which would give further information on the dynamics of this specific interaction.
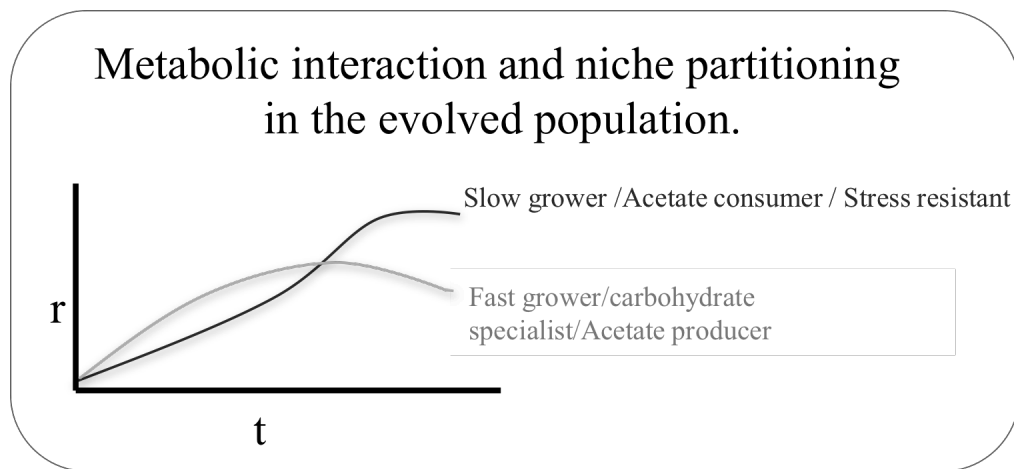
Figure 6.1. Metabolic interactions between types *a* and *c* and their ecological roles. r: Growth; t: time.

An examination of the literature revealed that heterogeneities with similar outcomes in sympatric conditions were already described by Helling *et al.* (1987) and Rosenzweig *et al.* (1994) and later on by Friesen *et al*. (2004). The difference of our findings with these studies lies in the complexity of the medium used, next to the experimental setup. To our knowledge, ours is the first study that empirically demonstrates phenotypic adaptive diversification at an intra-populational level in highly complex conditions. In previous studies, the evolution of cross-feeding polymorphisms, scavenging acetate or developing a diauxic switch between glucose and acetate were demonstrated under conditions of limiting substrate in rather simple environments. The results from the present work are striking, because, even though there is not one established limiting resource in LB medium, the outcome of evolution was comparable to that in environments where one substrate was used as the limiting one. The evolutive

strategies in LB worked such that the metabolism was readjusted to the use of alternative ways (*i.e.* use of galactose and other alternative sugars) to feed into glycolysis. In other words, replacing glucose by other sugars that feed into the central metabolism. The result of getting into the central metabolism cascade was the same, *i.e.* an overproduction of acetate; the whole process is consistent with the opening of a new niche that is occupied by new forms. The data suggest that the scope of evolution and adaptation is large enough to accommodate the metabolic repertoire of the cells, as necessary genomic changes occurred that activated unused alternative pathways to cope with the new conditions. In a recent review, Barrick and Lenski (2013) discuss the strength and scope of selection, defining it as an important 'deal', which, when it is less stringent, sustains more genetic diversity, allowing more opportunities for populations to become optimized by exploring alternative paths. Niches that might be already present in the environment as well as niches created by genomic and phenotypic readjustment can thus be occupied by a plethora of forms maintaining interactive relationships and diversification.

In the light of the results mentioned above, I further investigated the possibility of the presence of a dynamic and stable coexistence (**chapter 5**). I also analyzed the ecological nature and further phenotypic characteristics of each type that would add to a beneficial interaction between them and would confirm niche partitioning in the population. Since the experimental setup involved a propagation scheme encompassing clear phases of growth (short-exponential) and stationary phase, by focusing on information obtained from these two different growth stages, I could analyze and observe the emergence of distinct physiological and metabolic events. This served as evidence for the beneficial use of different physiological states of the cells. Given the fact that persistent behavior of one type (type *a*) in the face of adverse conditions (created in later stages of growth) had been observed (chapter 4), in **chapter 5** I challenged the selected types with several stress factors, in an attempt to find observable differential phenotypic behaviors. From the results, I obtained evidence for the contention that adaptive evolution can affect the survival properties of members of the population differently. A raised ability of type *a* to deal with stress when in an early stage of growth was observed. Such a response was likely related to so-called cross-protection, which could be related to the emergence of a GASP phenotype (Zinser and Kolter, 2004). In particular, oxidative stress resistance appeared to be favored. A mutation found in the other type, type *c,* was likely

related to the negative effect when confronted with oxidative stress. The mutation was found in a major regulator, the product of the *rpoA* gene, which confirmed the importance and impact of mutational events occurring in this type of gene. The results also confirmed that *trade-offs* occurred among the forms. Type *c* had "traded in" its ability to efficiently cope with stressful conditions for a better metabolic capacity, whereas type *a* had done the opposite. The lowered growth rates observed in type *a* for most tested substrates are likely related to a more active stringent response involving the activation of biosynthetic processes which are costly for the cells. It is clear that the environment created with the propagation scheme (short exponential and long stationary phases) had an impact on the physiological state of the sub-populations. The new physiological characteristics and prior history of cells had a great influence on the response to stress comprising a large amount of overlap between the stress responses, with multiple layers of complexity. In addition, I found a suggestive correlation with a non-mutated *arcA* gene (major regulator involved in oxygen transitions, inactive during aerobic growth but repressor of *rpoS* when active) in type *a* and a higher expression of *rpoS* related genes. The "normal" functioning of an inactive *arcA* gene might confer a raised ability to cope with stress. The *arcA gene* showed higher expression levels in the *arcA*-mutated type *c,* implying an active negative effect of this regulator on *rpoS* related genes.

Altogether, the results presented in this work indicate that the study of adaptation of *E. coli K12* MC1000 in sequential-batch cultures under stress-provoking and complex conditions has the potential to expand our understanding of genome stability and organismal adaptability. In continuous cultures in "simpler" (one-substrate-limited) environments, fewer and more "classical" mutations might become evident and perhaps be lost before they prove to be useful. Stress-induced mutagenesis may have been at the basis of the changes observed in LB medium, potentially accelerating or broadening the scope of evolution. Thus stress-induced mutagenesis is key to understanding the scope of evolution and adaption to natural environments, in particular since stress is a common state for microbes in nature. Even though I focused on only one *E. coli* population from the experimental evolution setup to observe diversification in detail, the results obtained are suggestive of similar outcomes in the remaining populations. In future work, an assessment of diversification such as the one performed with the selected population would be very interesting, in order to observe similarities or differences in the outcome of

evolution in the other environments. Also, even though the use of a common " landmark", *i.e.* the end populations after ~1000 generations, was sufficient to address our scientific questions, also the study of fitness landscapes, in which the speed of genomic and phenotypic changes are considered, would be very valuable to understand the dynamics of evolution.

Stable coexistence

In Chapter 5, I tested the stability of coexistence of selected evolved forms of *E. coli K12* MC1000, by growing the metabolically-interactive types in the environment they were selected in, as well as in environments that would possibly favor each of the types. I demonstrated an equilibrated coexistence in the selecting LB medium over time and an unbalanced response in the other environments. In these environments, with surplus carbohydrates as well as with stress factors applied, the tolerance level of the more sensitive type was apparently surpassed and equilibrium and coexistence were lost. This demonstrated that the scope of adaptive evolution in complex environments like LB medium, can reach up to causing niche partitioning involving not only metabolic *trade-offs* but also differential survival capabilities in the divergent forms which are able to maintain a balanced coexistence (Figure 6.2).

"Niche construction" (Barrack *et al*., 2013) can occur through the action of organisms themselves, whereas the available initial niches can also lead to coexistence. Thus, this study demonstrated that the presence of interactions between the different coexisting types, in combination with the already postulated ecological specialization and spatial heterogeneity maintained diversity which otherwise would be lost (Kassen, 2004). Models of selection in heterogeneous environments have suggested that, under the right conditions, diversity and coexistence can be stably maintained through negative frequency-dependent selection, the fitness of a genotype being higher when rare than when common.

I can shortly summarize that complex heterogeneous environments provide more niches— and therefore are more likely to cause diversity—than simpler environments with fewer niches. But the abiotic environment (oxygen availability and energy sources

availability) is not the sole determinant of niche space. The biotic environment is also relevant, with new ecological opportunities being continually created through the growth and activities of organisms. The complexity of natural and industrial settings has the potential to trigger the establishment of adapted coexisting forms.
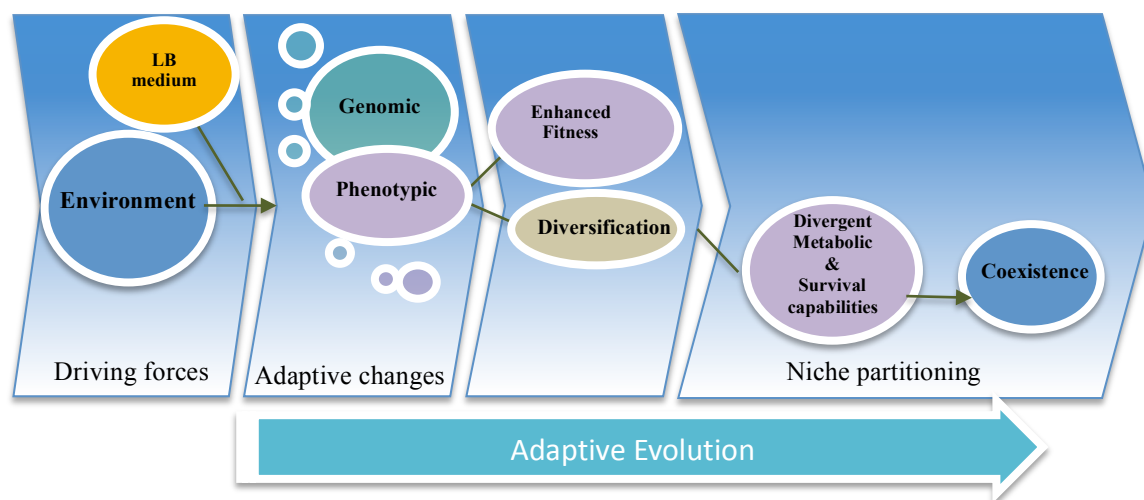
Figure 6.2. The scope of adaptation under a complex-heterogeneous environment (LB medium)

General conclusions:

In the light of the results obtained in the present work, the following general conclusions were achieved:

- Constant aerobic and anaerobic sequential-batch LB cultures represent a stressful environment to *Escherichia coli K12* MC1000. The study of adaptation of bacteria in such stressful and complex conditions has the potential to expand our understanding of genome and organismal adaptability to the challenges of the environment.

- The number of genetic and gene expression changes found after evolution of *E. coli K12* MC1000 in LB medium correlates with the complexity of the environment and with the ability to deal with the environment (fitness). This suggests an overall positive influence of such changes (*e.g.* activation of pathways) to deal with the challenging environmental cues.

- Preliminary phenotypic or genomic sorting is a useful approach when assessing adaptation from experimental evolution's populations because the presence of sub-populations and diversification can be effectively unrevealed.

- Oxygen availability was the main driving force for adaptation of *E. coli K12* MC1000 in LB medium and more specialized and adapted forms where obtained when this availability is constant and accompanied by the pressure of a long (stressful) stationary phase.

- The nature and multiplicity of the substrates in the LB medium formed a remarkable driving force for adaptation and diversification.

- Phenotypic innovations depend on the relationship between genomic changes and global and fine-scaled ecological conditions.

- Parallel responses in independent populations can occur in similar ecological settings (*i.e.* the same medium), but their presence does not avert diversification.

- Genetic and phenotypic adaptive changes in *Escherichia coli K12* MC1000 growing under a stressful and complex environment (LB medium) with multiple substrates (and lacking glucose) involve the strategic use of alternative metabolic pathways feeding into glycolysis. Thus, in a response likely driven by the medium, cells probably restructured pathways that were involved in the use of particular driving nutrients. In particular, genomic changes across all populations involved reactivation of the *gal* operon (mutation in *galR* and *galE*), thus conferring an outstanding enhanced capacity to utilize galactose and galactose-related carbohydrates. This concomitant response was amenable to selection by the medium. The resulting phenotypes were fitter in the stressful environment offered by LB without a limiting substrate.

- The physiological state and prior history of cells had a great influence on the response to stress. A focus on different physiological states can elucidate more specific traits and serve as a tool to identify the basis of diversification among the emerged sub-populations.

- Evolution of *Escherichia coli K12* MC1000 in sequential-batch LB cultures can incite niche partitioning as triggered by the multiplicity of substrates (providing a form of " structure" at a microscale) and possibly supported by interactions within the population. A stable and interactive coexistence of two main diverged forms with specific *trade-offs* was observed. One type is a more active carbohydrate

consumer (alternative pathways feeding into glycolysis). A second type (slow grower) acetate consumer (by- product from an active metabolism), has a raised tolerance to environmental stress.

- The scope of adaptive evolution in *E. coli* involves not only diversification resulting in enhanced (metabolic) fitness, but also has an impact on the survival capabilities of evolved forms in the face of ecological hardship. These properties may combine into an equilibrated coexistence. These characteristics demonstrate the potential of populations to maintain optimal performances (high metabolic capacity of the population and resistance to the global and local environmental cues) under complex and heterogeneous environments.

Future perspectives

There is a growing realization that bacterial populations that were previously believed to be isogenic are actually not so, in terms of selected or even unselected mutations. The selected mutations may become detectable only in cases where the experimenter is able to establish the conditions that select such forms, and thus often go unnoticed. Yet, they may be important as potential founders of future selected forms. These contentions demand a reconsideration of experimental settings and methodologies. Luckily, the new high-throughput technologies of genomics and transcriptomics allow us to make quick progress in this experimental field, achieving data in a relatively rapid manner. For example, the direct sequencing of mixed communities (*i.e.,* metagenomics) and subsequent annotation has already generated fantastic information of the functions and specialization of members of the populations (Herron *et al*., 2013). On the other hand, difficulties still lie in the manner in which we are able to detect differences in emerged forms at the level of phenotype. A phenotype which is selected under a particular (ephemeral) condition in a sympatric system may go unnoticed due to its very nature, however it may have had an impact, temporarily selecting a particular type. On the other hand, persistently selecting conditions will be easier to establish and thus the resulting adapted forms can be detected. Given the fact that what we call "phenotype" is actually the result of the complex interplay of the products of gene expression in the cell (for *E. coli*, this amounts to up to 4,500 gene products, which are not all expressed at the

same time), one is faced with a dazzling puzzle. It is thus likely that other particular forms in our sequential-batch LB cultures, under the three oxygen regimes, were selected, but that we simply have had no way to detect their selection. On the positive side, in this study we did come across several emerged forms that indeed could be shown to be the resultants of selective processes, and we did show that some of these could co-exist in a rather stable manner throughout the sequential-batch cycling conditions (one oxygen regime). So, what's next? We now understand that, across our 19 selected forms, a plethora of genomic changes (mutations) had taken place, of which some (but only these) could be shown to be selectable and selected. What about all the other ones? It is very likely that most of these are the resultant of genetic drift, escaping any selective force. However, some of the other ones may have been selected, may have had an impact in the population but gone unnoticed by us. Here, I feel, there is a great challenge to future work: attempt to explain what other forms may still have been selected, in terms of what selective force acted upon them and what this means for the total population in terms of interactions and overall fitness.

The positive side of current-day experimental evolution work lies in our expanding possibilities to assess the sequences of whole genomes. Genome editing technologies, including means to assess genetic exchange (horizontal gene transfer), confer powerful tools to infer the types and consistencies of genome changes in evolving populations, leading to overall adaptation. Here, considerations should be on a better distinction of randomness (changes due to drift) versus directedness (changes on which selective force has acted). In the end, one would need to provide proof-of-concept of selective events by directed mutagenesis studies. In this case, the introduction of genetic changes in reference or ancestral genomes is not efficient, as other changes may have occurred in the evolved forms as well. Thus it becomes necessary to introduce directed mutations (*i.e.* reversion or removal of a mutation) in the evolved forms in order to consider any differential behavior between mutated and parental form in the face of local (selective) conditions. This would also allow to disentangle putative epistatic interactions in the parental form.

We can safely state that the sheer amount of data obtained from the genomic and phenotypic analyses and all results presented here open a great window-of-opportunity to

further investigate the effect of long sequential-batch evolution on the nature of *E. coli* diversity and coexisting types in this or any other experimental setting. Some of the remaining open questions are:

-Since the occurrence of prominent parallel responses was an outstanding outcome of *E. coli* evolution triggered by the medium in all conditions, was diversification and niche partitioning in the other populations (*i.e.* from the fluctuating and anaerobic environments) equally influenced by the effects of the medium?

-How does diversification of *E. coli* affect the speed of adaptation?

-How does evolution act upon these populations after several more thousands of generations?

Using molecular approaches in bacterial cultures in simpler environments, we have already learned that selection appears to favor genomic and phenotypic changes in the forms that emerge after growth. We have also seen that successive adaptive mutations can shift the metabolic capacities of the cell and thus redefine its niche. These changes may lead to coexistence of diverged adapted forms within one medium (sympatric coexistence). Even though very similar basic rules govern the adaptation of bacteria like *E. coli* to complex environments, there are differences, and these lie in the sheer effect of the growth medium. I thus believe there is still much to be discovered about evolution in complex model systems. Bluntly, in spite of the fact that simpler (one limiting substrate, one or few environmental pressures) systems can give a rapid understanding of the mechanisms and paths of evolution, the complexity of most systems, natural and industrial ones, corroborates the inclusion of complex environmental pressures. Coexistence in such complex media is still rather poorly understood, in spite of the fact that I elucidated the basis of coexistence for one such coexisting pair. It is likely that the complex media hide many more still cryptic forms that also find their ***raison d'etre*** in the medium and local conditions. Well-concocted laboratory experiments (mimicking such environments) should be used to address such coexistences and these should possibly be taken to higher allowed levels of complexity. The extension of the approach to such systems would be of benefit to our knowledge of bacterial adaptation to complex systems and will undoubtedly lead to important advances with respect to the identification of

specific mutations that contribute to (1) enhanced activity and/or efficiency in utilization of available resources, and (2) enhanced resistance to environmental pressure, including antibiotic and hygiene treatments. Future studies should reveal the extent to which ecological interactions and differences in evolvability in genetically and phenotypically diverse populations affect the adaptive outcomes, which can benefit our knowledge in both industrial and medical settings.

References

Autumn K, Ryan MJ, Wake D. (2002). Integrating Historical and Mechanistic Biology Enhances the Study of Adaptation. The Quarterly Review of Biology:77 (4) :383-408

Baquero MR, Nilsson A, Turrientes M, Sandvang D, Galán JC, Martínez JL, Frimodt-Møller N,

Baquero F, Andersson D. Polymorphic Mutation Frequencies in *Escherichia coli*: Emergence of Weak Mutators in Clinical Isolates . (2004). J Bacteriol. 186(16): 5538-5542.

Barrick JE, Lenski R. (2013). Genome dynamics during experimental evolution. Nature Reviews Genetics 14, 827-839.

Buckling A, Brockhurst, M, Travisano M, Rainey P. (2007). Experimental adaptation to high and low quality environments under different scales of temporal variation. J Evol Biol. 20(1):296-300.

Colegrave N, Bukling A. (2005). Microbial experiments on adaptive landscapes. Bioessays. 11:1167-73.

Cooper T, Remold S, Lenski R, Schneider D. (2008). Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. PLoS Genetics. 4 (2) e351

Cooper T, Rozen, D, Lenski R. (2003).   Parallel and convergent changes across lineages are hallmarks of adaptive evolution. Proc Natl Acad Sci USA 100 (3):1072-1077, doi: 10.1073/pnas.0334340100

Cooper TF, Lenski R. (2010). Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMC Evol Biol. 13;10:11.

Dalziel A, Rogers S, Schulte P. (2009). Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. Molecular Ecology: 18, 4997-5017

Dettman JR, Rodrigue N, Melnyk H, Wong A, Bailey SF, Kassen R. (2012). Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol. Ecol. 21:2058-2077.

Elena S., Lenski R. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews, Genetics 4: 457-69.

Foster PL. (1999). Mechanisms of stationary phase mutation: a decade of adaptive mutation. Annu Rev Genet. 33:57-88

Foster PL. (2007). Stress-induced mutagenesis in bacteria. Crit Rev Biochem Mol Biol. 42(5):373-97.

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(2): 245-260

Gonzalez C, Hadany L, Ponder R, Price M, Hastings PJ, Rosenberg S. (2008). Mutability and Importance of a Hypermutable Cell Subpopulation that Produces Stress-Induced Mutants in *Escherichia coli*. PLoS Genet. 4(10): e1000208.

Hanko, V. P., & Rohrer, J. S. (2004). Determination of amino acids in cell culture and fermentation broth media using anion-exchange chromatography with integrated pulsed amperometric detection. Analytical Biochemistry, 324(1), 29–38. doi:10.1016/j.ab.2003.09.028

Helling, R, Vargas C, Adams J. (1987). Evolution of *Escherichia coli* during growth in a constant environment. Genetics. 116(3): 349-358.

Herron MD, Doebeli M (2013) Parallel evolutionary dynamics of adaptive diversification in
Escherichia coli. PLoS Biol 11: e1001490 doi:10.1371/journal.pbio.1001490.

Hersh M, Ponder R, Hastings PJ., Rosenberg S. (2004).Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. Research in Microbiology 155:352-359

Hughes BS, Cullum AJ, Bennett AF. (2007). Evolutionary adaptation to environmental pH in experimental lineages of *Escherichia coli*. Evolution 61(7):1725-34.

Jayaraman R. (2011). Hypermutation and stress adaptation in bacteria. J Genet.:90(2):383-91.

Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity. Science 15: 173-190

Kassen R, Rainey P. (2004). The ecology and genetics of microbial diversity. Annu. Rev. Microbiol. 58:207-31

Kolter R, Zambrano MM. (1996). GASPing for Life in Stationary Phase. Cell 86: 181-184

Kolter R, Zinser ER. (2004). *Escherichia coli* evolution during stationary phase. Res Microbiol. 155(5):328-36

Korona R, Nakatsu CH, Forney LJ, Lenski RE. (1994). Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. Proc Natl Acad Sci USA 91:9037-9041.

Korona, R., Nakatsu, C. H., Forney, L. J. & Lenski, R. E. (1994). Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. Proc. Natl Acad. Sci. USA91, 9037-9041

Leroi, A. M., Lenski R, and Bennett AF. (1994). Evolutionary adaptation to temperature. III. Adaptation of *Escherichia coli* to a temporally varying environment. Evolution 48:1222-1229.

Nakatsu CH, Korona R, Lenski R, De Bruijn F, Marsh T, Forney L. (1998). Parallel and Divergent Genotypic Evolution in Experimental Populations of *Ralstonia sp*. J Bacteriol. 180(17):4325-31.

Nikaido H. (2009). "The limitations of LB medium". Blog: Small things to consider. American Society for Microbiology. http://schaechter.asmblog.org/schaechter/2009/11/the-limitations-of-lb-medium.html

Rainey P, Travisano M. (1998). Adaptive radiation in a heterogeneous environment. Nature. 1998 394(6688):69-72.

Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. Nature 394, 69-72 (1998).

Rosenzweig RF, Sharp RR, Treves DS, Adams J. (1994). Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics 137(4): 903-917.

Saxer G, Doebeli1 M, Travisano M.(2009). Spatial structure leads to ecological breakdown and loss of diversity. Proc. R. Soc. B. 276, 2065-2070

Sezonov G, Joseleau-Petit D, D'Ari R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. J Bacteriol.189(23):8746-9.

Traverse, C. C., Mayo-Smith, L. M., Poltak, S. R, Cooper, V. S. (2013). Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. Proc. Natl Acad. Sci. USA 110, E250-E259.

Woods R, Schneider D, Winkworth C, Riley M, Lenski R (2006). Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci U S A. 103(24):9107-12.

Supplementary Information

Figure S1.1. Propagation scheme of the experimental evolution using LB medium with *Escherichia coli K12* MC1000 (~1000 generations)
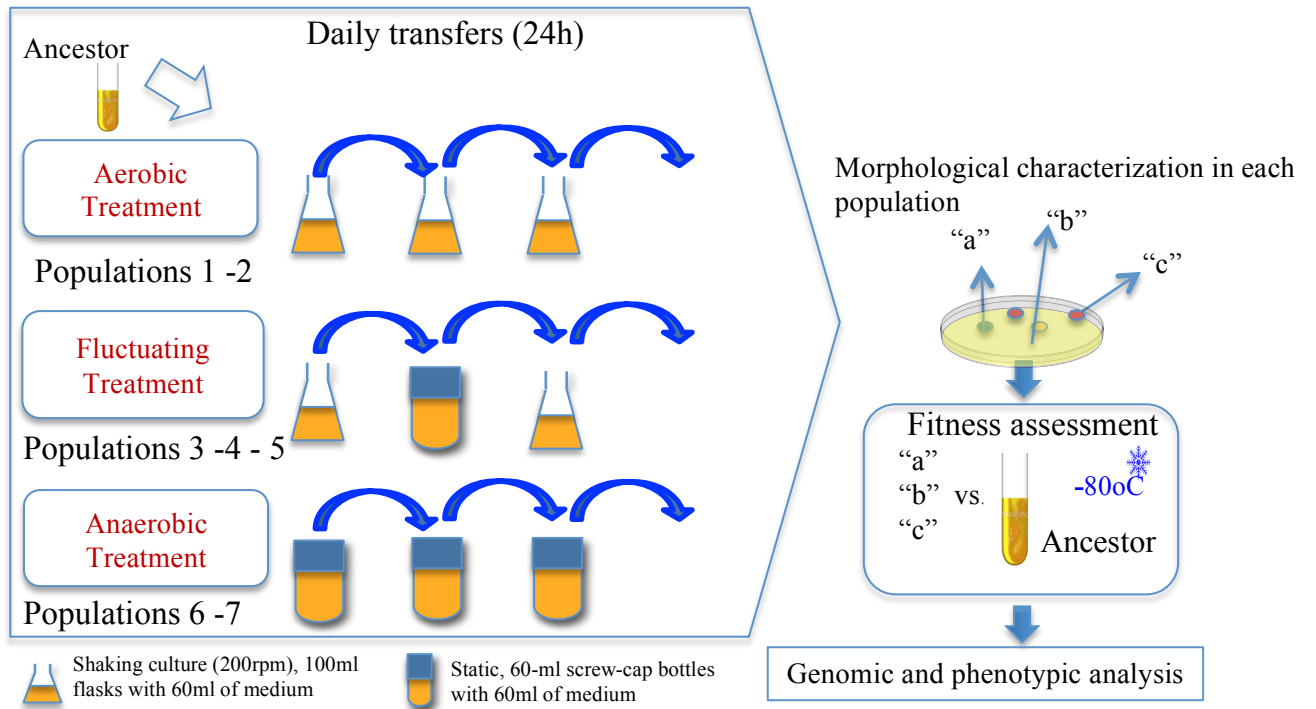
Figure S1.2. Carbohydrate and amino acid composition analysed by chromatography.
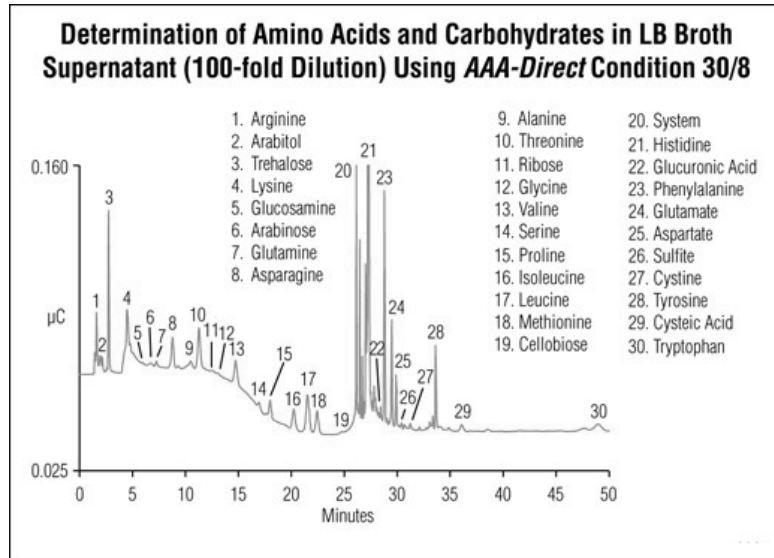
Source : Hanko *et al*., 2004. After 24 h of culture, high traces of 2,3-butanediol, erythritol, mannitol, galactose and glucose are detected. (Hanko *et al*., 2004, 2013).

Figure S2.1. Growth in M9 minimal medium with 0.4% of D- galactose as sole carbon source. The graph shows the OD600 values obtained by population 2 forms and the ancestor (growing separately) under oxygen constant conditions (treatment A) during 24h.
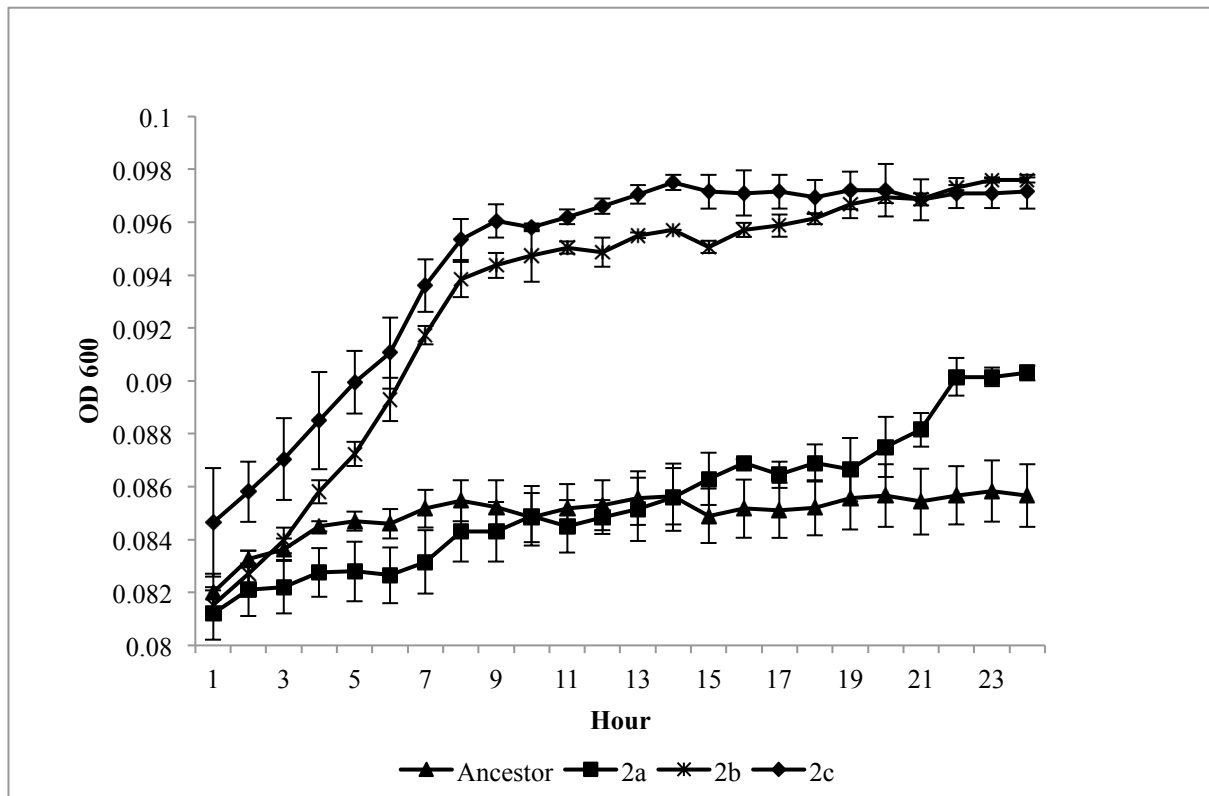
Table S2.1. Changes in intergenic regions consistent to all colony types

Intergenic region	Change in nucleotide	Position in the genome	Gene position
cra/mraZ	C→A	89089	+57/-545
	G→C	89103	+71/-531
	A→G	89143	+111/-491
	G→T	89145	+113/-489
	G→A	89239	+207/-395
	C→A	89241	+209/-393
	A→G	89263	+231/-371
	T→C	89316	+284/-318
	A→G	89321	+289/-313
	T→G	89471	+439/-163
	A→C	89570	+538/-64
purE/lpxH	G→A	552384	-61/+57
csgC/ymdA	C→A	1104604	+88/-33
treA/dhaM	G→A	1246725	-126/+194
ftnB/yecJ	C→A	1985523	+71/+8
argZ/argY	T→A	2816213	-56/+7
iap/ygbF	C→T	2876099	+459/+492
queF/ygdH	C→T	2924314	+96/-16
speC/yqgA	+G	3107198	-21/-377
yhbX/leuU	G→A	3319870	-235/+224
yigG/rarD	+C	4001264	-48/+47
glyV/glyX	2 bp→TT	4390491	+33/-3
insD/yjgZ	T→C	4497684	+161/-1599

Table S2.2 Unique mutations found in each colony type. Small colonies (a) from treatment A (populations 1 and 2) and form 7c from treatment C have the highest number of unique mutations

1a	leuA, proB, ispA, ybbC/selU , entF, hscC, rhtA, mntR, dacC/deoR, potH, elfG, flgF, ycgB/dadA, trpA, insH/ynaJ, hrpA, aldA, mokB/trg, narZ, flxA/dicC, sppA, yoaE, yegL/ibsA, yehK/yehL, yeiS, yfcM, trxC, kgtP, yfjR, lhgO, hycA/hypA, fucO, xerD, hybE, plsC, insD/yqiH, argG, mlaF/yrbG, tusD, yhfS, xylB, rfaG, ibpB, rnpA, mdtL, rbsA, wzzE, rrsE, metA/aceB, dcuA, yjeM, aidB, ulaE, yjhQ, yjjU.
1b	yafE, rrsG/clpB, yqjA.
1c	ilvD
2a	ftsI, lpxC/secM, pdhR/ace, yafJ, ykfA, yahK, ddlA/iraP, ybeF, gsiA, ycaN, ycaL, serX, flgI, cob, adhE, hipB/lsrK, dtpA/gstA, anmK/slyB, lhr, ydiI, btuE, ydiY, xthA, araF/ftnB, yedL/fliE, asnW/yeeO, yees, yegP/yegQ, yehY, yohF/yohP, cdd, yfaQ/yfaT, yfdE, yfeW, yfeY, yfhG/glrK, ffh/ypjD, hycD, prfB, yhaC/garK, agaV/agaS, yhbP/yhbQ, rlmE, sspA, trkA, pabA, yhfT, glpG, livJ, yhjG, bcsB, yiaD, yifB, rnr, insI, hsdS.
2b	yhhZ
2c	None -
3a	rpmB
3b	None -
3b	acnA, tnaA/tnaB.
4a	metQ, ydiM, recG.
4b	cutC, yfjL/yfjM.
4c	None -
5a	ycjM, ttdA
5b	yaeJ, yaeF, prpD.
5c	malX, asnT/yeeJ, ppiC/rep, gltP/yjcO.
6b	hha, yfgO, nanR/dcuD, ibpA/yidQ.
6c	modA, msbA, ydiJ, cytR.
7b	mmuM/afuC, cynX, ydhB, [qseB]–[qseC].
7c	aroP/pdhR, betT/yahA, cof, ftsK, ycdU/serX, ynaK/insH, preA, mglA, bcr, nuoH, yfcA, rnlB/yfjP, bglA, hybB, gltF, arsB, rhtC, nrfC, psd, yjhI, yjjI.

Table S3.1. List of primers used for RT-qPCR testing. Design was performed using Clone Manager Suite (Sci-Ed Software).

rpoA	5' GTGGAGCGTATTGCCTACA 3'
	5' CACTTCAGGCTGACGTACA 3'
gadC	5' AATACCGGTGGCGGTAACAA 3'
	5' TGAAGTCAGCAGACCGACAA 3'
gatB	5' TTGCGACCTCTACGATGG 3'
	5' TGCCGTGAACTAACGGAA 3'
hdeA	5' AACTCCTGGACCTGTGAA 3'
	5' ACAAGCCTGAACGATAGC 3'
ompF	5' TTCTAACGGCGACGGTGTTG 3'
	5' AGTTCGCTGCCAGGTAGATG 3'
ompW	5' AAGGTGCTGGTGGTACGTTA 3'
	5' TCAGTGTTGGTGGCAGATGA 3'
galE	5' ACAATGTCAACGGCACTCTG 3'
	5' ATCGGTGAGGATCTGTTCCA 3'
aceF	5' ACGTAAGCTGGATGTGAAGA 3'
	5' TGAATACCGGAACAACCAGA 3'

Table S3.2. List of non-synonymous and intragenic mutations (by treatment and form). Position in the genome, mutation and gene are indicated. Consistent mutations to all forms are included.

Aerobic treatment																	
Population 1									Population 2								
1a			1b			1c			2a			2b			2c		
Position	Mutation	Gene	Position	Mutation	Gene	Position	Mutation	Gene	Position	Mutation	Gene	Position	Mutation	Gene	Position	Mutation	Gene
82790	G→T	leuA	232316	T→A	yafE	236356	T→A	dnaQ	91984	G→A	ftsI	236356	T→A	dnaQ	236356	T→A	dnaQ
236356	T→A	dnaQ	236356	T→A	dnaQ	280505	G→A	yagA	244907	C→T	yafJ	280505	G→A	yagA	280505	G→A	yagA
259892	A→T	proB	280505	G→A	yagA	480800	A→T	acrB	267466	C→T	ykfA	480800	A→T	acrB	480800	A→T	acrB
280505	G→A	yagA	480800	A→T	acrB	920009	G→C	macB	342276	C→T	yahK	603166	T→A	mscM	603166	T→A	mscM
439531	A→T	ispA	920009	G→C	macB	993388	A→T	ssuC	483561	+CC	acrB	730733	T→G	rhsC	730733	T→G	rhsC
480800	A→T	acrB	993388	A→T	ssuC	1018787	A→G	ompA	659914	T→C	ybeF	742366	T→A	ybgI	742366	T→A	ybgI
614770	A→T	entF	1018787	A→G	ompA	1049254	C→T	insA	867645	+C	gsiA	798556	G→T	pgl	798556	G→T	pgl
681598	A→G	hscC	1049254	C→T	insA	1391717	T→C	mppA	948053	A→G	ycaN	985802	T→C	ompF	985802	T→C	ompF
848859	Δ1 bp	rhtA	1391717	T→C	mppA	1605781	A→C	tam	959885	C→T	ycaL	992502	A→C	ssuB	992502	A→C	ssuB
852767	T→C	mntR	1605781	A→C	tam	1736925	G→C	ydhB	1096803	Δ1 bp	serX	993388	A→T	ssuC	993388	A→T	ssuC
896033	T→A	potH	2006011	G→C	yedD	2715835	G→A	yfiF	1135744	+A	flgI	994013	A→T	ssuC	994013	A→T	ssuC
993388	A→T	ssuC	2104934	G→A	rfc	3078108	A→G	tktA	1179448	A→G	cobB	1004231	T→A	pyrD	1004231	T→A	pyrD
1001958	C→T	elfG	2104945	C→A	rfc	3233453	T→A	rlmG	1295532	C→T	adhE	1018787	A→G	ompA	1018787	A→G	ompA
1018787	A→G	ompA	2715835	G→A	yfiF	3362177	A→T	yhcD	1728396	Δ1 bp	lhr	1143882	T→A	yceQ	1143882	T→A	yceQ
1133155	T→A	flgF	3078108	A→G	tktA	3676960	T→G	rpoA	1763625	G→A	ydiI	1388056	A→C	mpaA	1388056	A→C	mpaA
1314895	A→T	trpA	3233453	T→A	rlmG	3676960	T→A	yhjH	1791874	+C	btuE	1391717	T→C	mppA	1391717	T→C	mppA
1391717	T→C	mppA	3245929	G→T	yqjA	3953327	C→T	ilvD	1804066	+C	ydiY	2249286	C→A	nfo	2104942	T→A	rfc
1484484	T→C	hrpA	3362177	A→T	yhcD	4031240	G→A	trkH	1831033	+G	xthA	3043313	T→G	ygfF	2104945	C→A	rfc
1486943	A→C	aldA	3438172	T→G	rpoA	4278367	T→A	yihQ	1975505	T→C	flhC	3063383	T→A	scpC	2249286	C→A	nfo
1537655	A→C	narZ	3676960	T→A	yhjH	4278367	T→A	yjcE	2006011	G→C	yedD	3078108	A→G	tktA	2715835	G→A	yfiF
1846896	Δ1 bp	sppA	4031240	G→A	trkH	4356892	A→T	cadB	2074483	T→C	yeeS	3233453	T→A	rlmG	3043313	T→G	ygfF
1898465	A→T	yoaE	4278367	T→A	yjcE	4390539	G→A	glyX	2215948	G→A	yehY	3438172	T→G	rpoA	3063383	T→A	scpC
2104934	G→A	rfc	4356892	A→T	cadB	4638103	G→A	arcA	2230064	C→T	cdd	3580884	T→C	yhhZ	3078108	A→G	tktA
2104945	C→A	rfc	4390539	G→A	glyX				2486602	Δ1 bp	yfdE	4031240	G→A	trkH	3233453	T→A	rlmG
2231644	T→C	yeiS	4638103	G→A	arcA				2547079	+G	yfeW	4066842	T→A	yihQ	3438172	T→G	rpoA
2442536	A→T	yfcM							2548958	G→A	yfeY	4273042	A→T	yjcB	4031240	G→A	trkH
2715835	G→A	yfiF							2715835	G→A	yfiF	4390539	G→A	glyX	4066842	T→A	yihQ
2716848	A→G	trxC							2844951	C→T	hycD	4638103	G→A	arcA	4273042	A→T	yjcB
2723744	C→T	kgtP							3033568	T→C	prfB				4390539	G→A	glyX
2768360	A→C	yfjR							3325328	A→G	rlmE				4638103	G→A	arcA
2788740	A→C	lhgO							3375426	T→C	sspA						
2930651	T→C	fucO							3435648	T→C	trkA						
3037475	A→T	xerD							3488511	A→G	pabA						
3078108	A→G	tktA							3504317	G→A	yhfT						
3138778	T→A	hybE							3559264	G→T	glpG						
3160995	G→T	plsC							3597017	T→C	livJ						
3233453	T→A	rlmG							3675164	G→A	yhjG						
3317163	C→A	argG							3689961	T→C	bcsB						
3438172	T→G	rpoA							3714907	A→G	yiaD						
3473520	G→T	tusD							3946828	+G	yifB						
3503348	G→T	yhfS							4406478	G→C	rnr						
3726129	Δ1 bp	xylB							4506224	Δ1 bp	insl						
3804930	A→T	rfaG							4578883	+C	hsdS						
3864646	G→A	ibpB															
3882843	C→A	rnpA															
3889642	C→T	mdtL															
3932240	A→C	rbsA															
3968059	A→T	wzzE															
4031240	G→A	trkH															
4066842	T→A	yihQ															
4207699	G→A	rrsE															
4357410	G→T	cadB															
4364206	A→T	dcuA															
4382226	A→T	yjeM															
4390539	G→A	glyX															
4413285	C→T	aidB															
4421335	T→C	ulaE															
4531461	A→T	yjhQ															
4610726	T→C	yjjU															
4638103	G→A	arcA															

Fluctuating treatment

3a Position	Mutation	Gene	3b Position	Mutation	Gene	3c Position	Mutation	Gene	4a Position	Mutation	Gene	4b Position	Mutation	Gene	4c Position	Mutation	Gene	5a Position	Mutation	Gene	5b Position	Mutation	Gene	5c Position	Mutation	Gene
2006011	G→C	yedD	2715835	G→A	yfiF	1335503	C→G	acnA	220388	C→T	metQ	1956957	C→G	cutC	1975679	A→G	flhC	1369371	C→A	ycjM	215040	Δ10 bp	yaeJ	1697605	T→C	malX
2104942	T→A	rfc	3350639	T→C	arcB	1736925	G→C	ydhB	1736925	G→C	ydhB	1975679	A→G	flhC	2715835	G→A	yfiF	1790503	G→T	nlpC	216913	A→G	yaeF	2006011	G→C	yedD
2104945	C→A	rfc	3438172	T→G	rpoA	1856637	A→C	ydjJ	1770249	A→T	ydiM	2104934	G→A	rfc	3438147	T→G	rpoA	2281802	C→A	yejK	350974	C→T	prpD	2715835	G→A	yfiF
2281802	C→A	yejK	4032285	T→G	trkH	2281802	C→A	yejK	2006011	G→C	yedD	2281802	C→A	yejK				2751288	A→C	recN	1736925	G→C	ydhB	2751288	A→C	recN
2715835	G→A	yfiF				2715835	G→A	yfiF	2104945	C→A	rfc	2715835	G→A	yfiF				3204772	T→C	ttdA	2715835	G→A	yfiF	3820669	6 bp x 2	spoT
3438172	T→G	rpoA				3350639	T→C	arcB	2281802	C→A	yejK	3438147	T→G	rpoA				3438172	T→G	rpoA	2751288	A→C	recN	4031240	G→A	trkH
3809590	G→C	rpmB				3438172	T→G	rpoA	2715835	G→A	yfiF	4280355	+A	yjcF				4031240	G→A	trkH	3438172	T→G	rpoA	4280355	+A	yjcF
4280355	+A	yjcF				4032285	T→G	trkH	3438147	T→G	rpoA							4280355	+A	yjcF	4031240	G→A	trkH			
									3825205	T→C	recG										4280355	+A	yjcF			

Anaerobic treatment

Population 6 — 6b Position	Mutation	Gene	6c Position	Mutation	Gene	Population 7 — 7b Position	Mutation	Gene	7c Position	Mutation	Gene
479519	2 bp→AA	hha	794385	A→G	modA	14838	A→G	dnaJ	14838	A→G	dnaJ
1790503	G→T	nlpC	966250	Δ3 bp	msbA	198766	G→A	bamA	198766	G→A	bamA
2613712	G→C	yfgO	1764260	A→G	ydiJ	359726	Δ3 bp	cynX	415474	C→T	sbcB
2715835	G→A	yfiF	1856637	A→C	ydjJ	415474	C→T	sbcB	447041	T→C	cyoD
2751288	A→C	recN	2006011	G→C	yedD	447041	T→C	cyoD	467423	A→G	cof
3820669	6 bp x 2	spoT	2104945	C→A	rfc	487347	T→C	kefA	487347	T→C	kefA
3820674	+CAGGAT	spoT	2281802	C→A	yejK	588178	T→C	nfrA	588178	T→C	nfrA
4280355	+A	yjcF	2751288	A→C	recN	686556	T→G	gltI	686556	T→G	gltI
4463866	A→T	treB	3438172	T→G	rpoA	721060	Δ1 bp	kdpD	721060	Δ1 bp	kdpD
			4122338	C→G	cytR	744719	A→G	ybgL	744719	A→G	ybgL
			4280355	+A	yjcF	969476	Δ1 bp	ycaQ	934786	G→A	ftsK
						991312	T→C	pepN	969476	Δ1 bp	ycaQ
						1044998	A→G	gfcE	991312	T→C	pepN
						1146959	T→C	plsX	1044998	A→G	gfcE
						1251255	A→G	dhaR	1146959	T→C	plsX
						1564870	C→T	dosC	1251255	A→G	dhaR
						1643549	+G	relE	1564870	C→T	dosC
						1736925	G→C	ydhB	1643549	+G	relE
						1856637	A→C	ydjJ	1925265	Δ1 bp	ptrB
						1925265	Δ1 bp	ptrB	1955508	T→C	torY
						1955508	T→C	torY	2016554	Δ1 bp	fliK
						2006011	G→C	yedD	2113205	T→C	wcaM
						2016554	Δ1 bp	fliK	2148375	Δ1 bp	yegI
						2113205	T→C	wcaM	2233466	T→C	preA
						2148375	Δ1 bp	yegI	2237056	G→T	mglA
						2310395	T→C	ompC	2276603	+T	bcr
						2528570	T→C	zipA	2310395	T→C	ompC
						2600039	Δ1 bp	hyfB	2394791	T→C	nuoH
						2602740	+T	hyfC	2443441	T→C	yfcA
						2603023	T→C	hyfD	2528570	T→C	zipA
						2665602	Δ1 bp	hcaT	2600039	Δ1 bp	hyfB
						2715835	G→A	yfiF	2602740	+T	hyfC
						2758995	T→C	yfjJ	2603023	T→C	hyfD
						2855365	T→A	mutS	2665602	Δ1 bp	hcaT
						2887937	T→G	cysI	2715835	G→A	yfiF
						2918979	T→C	gudP	2758995	T→C	yfjJ
						3099061	+C	yggN	2855365	T→A	mutS
						3112732	C→T	yghJ	2887937	T→G	cysI
						3148431	T→C	yghA	2918979	T→C	gudP
						3152326	+C	yqhC	3042605	A→G	bglA
						3175524	T→C	nudF	3099061	+C	yggN
						3242082	G→A	uxaC	3112732	C→T	yghJ
						3277527	Δ1 bp	kbaZ	3142080	A→G	hybB
						3339562	T→C	kdsD	3148431	T→C	yghA
						3344972	T→C	ptsN	3152326	+C	yqhC
						3604604	A→G	zntA	3175524	T→C	nudF
						3611314	T→C	acpT	3242082	G→A	uxaC
						3642991	T→C	prlC	3277527	Δ1 bp	kbaZ
						3805538	A→C	rfaQ	3339562	G→A	kdsD
						3822769	A→G	trmH	3344972	T→C	ptsN
						4077456	T→C	yiiE	3359859	T→C	gltF
						4280355	+A	yjcF	3604604	A→G	zntA
						4359158	T→C	cadC	3604971	G→A	zntA
						4409860	A→G	yjfK	3611314	T→C	acpT
						4463866	A→T	treB	3642991	T→C	prlC
						4542617	Δ1 bp	fimC	3647223	G→A	arsB
									3805538	A→C	rfaQ
									3822769	A→G	trmH
									4005839	T→C	rhtC
									4077456	T→C	yiiE
									4287949	T→G	nrfC
									4359158	T→C	cadC
									4387747	A→G	psd
									4409860	A→G	yjfK
									4463866	A→T	treB
									4523316	T→C	yjhI
									4542617	Δ1 bp	fimC
									4614450	T→G	yjjI

List of consistent mutations to all forms

Position	Mutation	Gene
19111	**C→T**	**nhaR**
89754	T→C	mraZ
477265	G→C	ylaB
790911	G→A	galE
904622	G→A	amiD
989107	G→A	pncB
1000027	T→A	elfC
1016487	C→A	ycbZ
1099293	C→T	ycdY
1122779	C→T	grxB
1176874	G→A	lolE
1296981	+A	adhE
1426111	C→T	insH
1441480	C→T	ydbH
1556221	G→T	ddpD
1608100	T→G	uxaB
1608464	C→T	uxaB
1854992	T→A	ydjI
1973248	G→A	cheA
2001300	A→C	fliC
2001486	C→A	fliC
2174013	T→A	gatZ
2248262	C→T	yeiH
2594417	C→T	ypfJ
2709307	+C	nadB
2754998	G→A	intA
2790231	G→A	gabD
2798457	G→A	ygaM
2839260	G→A	ascB
2861795	C→T	ygbL
2903847	C→T	ygcG
2936103	G→A	fucK
2975171	C→T	galR
3153053	G→A	yqhC
3317043	G→A	argG
3365966	G→A	yhcG
3397951	A→T	mreC
3466790	C→T	chiA
3472447	T→C	rpsL
3526394	C→G	yrfF
3627944	C→T	yhiL
3694932	G→A	bcsE
3759253	G→A	selA
3874696	G→A	yidA
3881701	G→A	dnaA
3910097	G→A	glmS
3929092	C→T	ravA
3986138	C→T	hemX
3994932	T→A	xerC
3995114	C→A	xerC
4122424	C→T	cytR
4138453	C→T	ptsA
4192019	T→G	thiE
4224722	G→A	metH
4514150	C→T	fecA

Table S3.3. Supplementary table for Figure 3.2. Genes located at the position encoded by the number in the PCA graph

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
aidB	folE	araC	aceE	amiA	arsB	ansB	argI	allS	cysW	dcuC	aspA	aspC	adhE
cbpA	grxA	lacI	aceF	argT	bcsG	araD	dcd	aroF	fadB	dhaM	atpF	dmsB	efp
crcB	hokD	melR	ahpC	bglF	ccmA	carA	fadR	arsC	fimI	fdnG	atpG	flhD	fimH
dps	ibpB	ubiC	bamD	csgB	citD	cdd	ivY	cirA	hdeB	fdnH	crp	glyS	hfq
erpA	ihfA	uidR	bhsA	ddpC	csiE	cusR	katP	glyQ	hspQ	fdnI	cyoA	lon	iraP
gst	ipxC	yadI	cspB	fdoG	cynS	cysJ	lacZ	hisC	pstS	glpA	cyoC	pdxK	pepD
hyaA	marA	ybeL	cspG	fieF	espD	der	purU	hlyA	recN	glpB	cyoD	pepE	rplD
iscA	mokP	ybiA	dusB	fliG	exoD	eamB	trpS	insI	spy	glpC	cyoE	rfbA	rplL
malX	relB	ydgD	galK	gloB	exoO	fdrA	ulaE	leuS	yaaX	glpD	frdA	rfbX	rplQ
metC	relE	yeeZ	hscA	nikC	fhuA	glf	yagJ	lgt	yaiY	glpF	ftnA	ydjY	rpmF
mtlD	rof	yfcH	hscB	pinH	lomK	hemL	yahJ	mltC	yceA	glpK	gnd	yeiT	rpsF
osmE	secG	yfeC	iscR	potG	murA	hyaC	yahK	modB	ycfJ	glpQ	ispF	ygfJ	rpsJ
purA	uspB	yghB	iscS	recQ	nemA	kdgR	ybiO	nudF	ygiB	glpT	maeB	yggL	rpsO
slp	uspD	yhfY	manZ	rfaS	paaE	lamB	ydfZ	pdhR	yqaE	hypD	nmpC	yjiM	tdcC
sufA	uspE	yjfI	nlpI	rhaD	paaI	lolA	ydjX	pgaC		malK	nuoE	ynfF	tdcD
sufD	yaeH	yqiB	oppC	ribF	phnL	moaE	yjhR	pheA		malM	nuoG	ynfG	tdcE
ubiA	ybfA	ytfH	oppD	rimK	pncA	msyB	yjiH	phnD		moaA	nuoM	yoaB	uspA
uxaB	ybhQ		proP	rsxA	ptsG	nrfA	ykgK	rmhA		nagB	rplK	yqeA	yhaR
ybaT	ydcH		psd	tauC	rhsB	oppA	ylaC	serB		nanA	wbbI		ynjE
ycaD	yefM		rnlA	yadD	rluB	racR	yqhC	stpA		nanB	ybaL		
yedP	ygaW		serA	yceF	rpiR	rtcA		truC		nanC	yfeW		
yegP	ygiT		thrB	ycfH	rpsK	tfaR		wbbJ		nanK	yffB		
yehZ	yhbS		thrC	ydgC	rusA	trpB		wcaF		nanT	ygeY		
yfcZ	yiiT		tolA	yffM	solA	ycfM		yaaA		napA	ymfR		
ygaU			ybiS	ygaZ	ubiD	ydiT		ybiH		napB			
ygiW			yeeD	yhhK	yagE	yfaH		ycbR		napC			
ynaJ			yeeR	ykgM	ybiR	yggG		ydjI		narG			
			ynaE	ylcB	ycbU	yhhX		yebO		narH			
			yrbE		ycjV	yjdP		yedZ		narI			
					ydcC	yjfK		yeiH		narK			
					ydcX			yfcG		narZ			
					ydfB			yghW		nirB			
					yebR			yhdJ		nirD			
					yebT			yiaO		yhcH			
					yfbN			yigL		yhcI			
					yggU			yiiX		yhcJ			
					yhiL			yjeO		yjiA			
					yibI			yjhH		yjiX			
					yicJ					yjiY			
					yihF					yjjI			
					yjfL								
					ykfH								
					ykgB								
					ynaI								

Figure S5.1. Experimental setup assessing the phenotypic nature, coexistence and niche differentiation of the emerged forms in the population after experimental evolution.
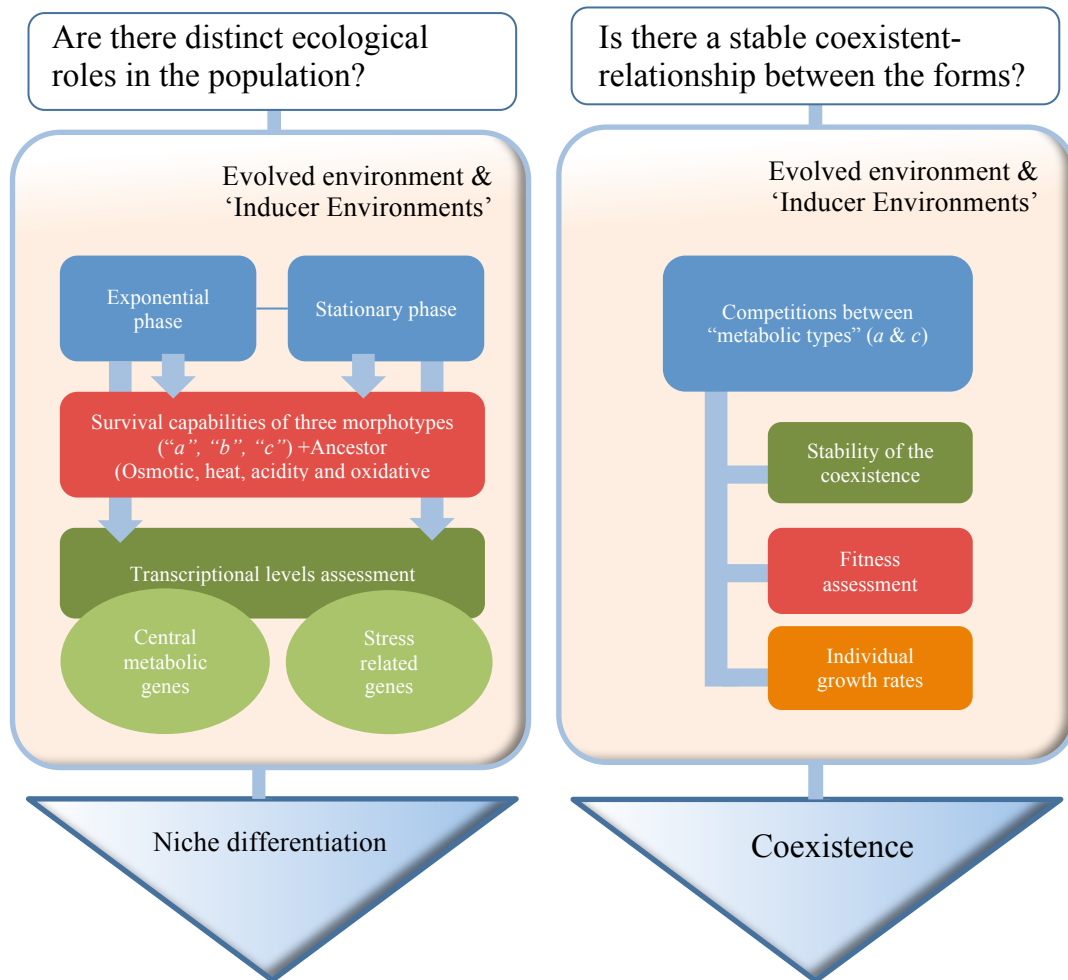
Figure S5.2. Relative fitness (*w*) of metabolic type *a* relative to type *c*. Fitness assessment after competition experiment under three different environments. Competition took place during 4 days with consecutive transfers into fresh medium with or without supplements.
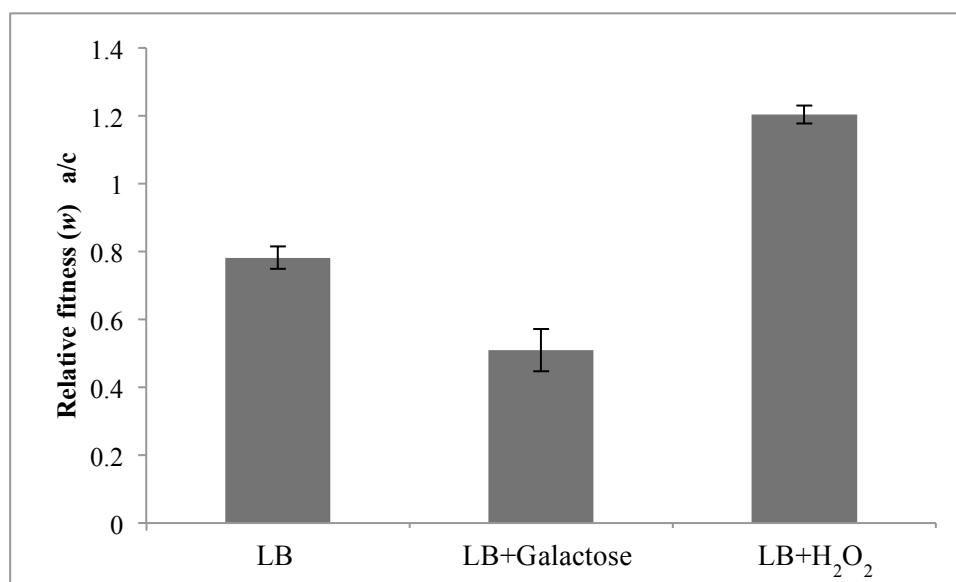

Table S5.1. Primers used for real-time PCR reactions

Oligo name	Sequence (5' → 3')
galE-a	CCT-GCA-CGA-TCA-CGC-TAT
galE-b	GCT-AAT-CAG-GCG-CAG-AGT
ptsH-a	TAC-CAT-TAC-CGC-TCC-GAA
ptsH-b	CTG-CGG-AGA-TAG-TCA-CAA
acs-a	GAG-CAT-GTG-GTG-GTA-CTG-AA
acs-b	TAG-AAC-CGG-AGG-TGT-AGA-GA
gadE-a	CTT-TTA-CAG-GGC-TTT-TGG
gadE-b	AAT-AAG-ATG-TGA-TAC-CCA-GG
bolA-a	GAT-ACG-TGA-GCG-GAT-AGA-AG
bolA-b	CGC-ATG-AAC-GGT-AGT-AGA-GA
cyoA-a	CTG-GAA-GTA-CCG-TGC-GAG-CAA
cyoA-b	GCT-AGG-CTC-AAG-AGC-GTG-AGT
treB-a	CGA-CGG-TGA-GCC-ACT-GTA-TT
treB-b	TTA-TCA-ACC-TGC-GCC-TGT-CC
arcA-a	CCG-TGC-TGA-ACT-GCT-GAA-GA
arcA-b	CCG-TGA-ATG-GTG-GCG-ATG-AT
gcvH-a	TTG-TTG-ACC-TGC-CGG-AAG-TG
gcvH-b	CCT-GCA-TAC-GGT-TCG-CTG-TT

Table S5.2. Non-consistent mutations in the population. In grey: consistent genes between the 3 types.

Non-consistent mutations in the population								
a			b			c		
Position	Mutation	Gene	Position	Mutation	Gene	Position	Mutation	Gene
91984	G→A	ftsI	236356	T→A	dnaQ	236356	T→A	dnaQ
244907	C→T	yafJ	280505	G→A	yagA	280505	G→A	yagA
267466	C→T	ykfA	480800	A→T	acrB	480800	A→T	acrB
342276	C→T	yahK	603166	T→A	mscM	603166	T→A	mscM
483561	+CC	acrB	730733	T→G	rhsC	730733	T→G	rhsC
659914	T→C	ybeF	742366	T→A	ybgI	742366	T→A	ybgI
867645	+C	gsiA	798556	G→T	pgl	798556	G→T	pgl
948053	A→G	ycaN	985802	T→C	ompF	985802	T→C	ompF
959885	C→T	ycaL	992502	A→C	ssuB	992502	A→C	ssuB
1096803	Δ1 bp	serX	993388	A→T	ssuC	993388	A→T	ssuC
1135744	+A	flgI	994013	A→T	ssuC	994013	A→T	ssuC
1179448	A→G	cobB	1004231	T→A	pyrD	1004231	T→A	pyrD
1295532	C→T	adhE	1018787	A→G	ompA	1018787	A→G	ompA
1728396	Δ1 bp	lhr	1143882	T→A	yceQ	1143882	T→A	yceQ
1763625	G→A	ydiI	1388056	A→C	mpaA	1388056	A→C	mpaA
1791874	+C	btuE	1391717	T→C	mppA	1391717	T→C	mppA
1804066	+C	ydiY	2249286	C→A	nfo	2104942	T→A	rfc
1831033	+G	xthA	3043313	T→G	ygfF	2104945	C→A	rfc
1975505	T→C	flhC	3063383	T→A	scpC	2249286	C→A	nfo
2006011	G→C	yedD	3078108	A→G	tktA	2715835	G→A	yfiF
2074483	T→C	yeeS	3233453	T→A	rlmG	3043313	T→G	ygfF
2215948	G→A	yehY	3438172	T→G	rpoA	3063383	T→A	scpC
2230064	C→T	cdd	3580884	T→C	yhhZ	3078108	A→G	tktA
2486602	Δ1 bp	yfdE	4031240	G→A	trkH	3233453	T→A	rlmG
2547079	+G	yfeW	4066842	T→A	yihQ	3438172	T→G	rpoA
2548958	G→A	yfeY	4273042	A→T	yjcB	4031240	G→A	trkH
2715835	G→A	yfiF	4390539	G→A	glyX	4066842	T→A	yihQ
2844951	C→T	hycD	4638103	G→A	arcA	4273042	A→T	yjcB
3033568	T→C	prfB				4390539	G→A	glyX
3325328	A→G	rlmE				4638103	G→A	arcA
3375426	T→C	sspA						
3435648	T→C	trkA						
3488511	A→G	pabA						
3504317	G→A	yhfT						
3559264	G→T	glpG						
3597017	T→C	livJ						
3675164	G→A	yhjG						
3689961	T→C	bcsB						
3714907	A→G	yiaD						
3946828	+G	yifB						
4406478	G→C	rnr						
4506224	Δ1 bp	insI						
4578883	+C	hsdS						

Table S5.3. Maximal specific growth (μ) in LB broth, LB broth supplemented with galactose 0.3% and LB broth supplemented with non-lethal concentration of H_2O_2 (0.3mM).

Medium	a (h^{-1})	± SD	c (h^{-1})	±SD	Ancestor (h^{-1})	±SD
LB broth	0.222	0.019	0.256	0.004	0.263	0.008
LB + galactose	0.234	0.005	0.310	0.009	0.261	0.006
LB + H_2O_2	0.245	0.005	0.185	0.014	0.224	0.002

Summaries

Summary

An inherent characteristic of natural as well as applied industrial environments is complexity. Scientific studies in the last decades have revealed the very flexible genetic and phenotypic capacities of microorganisms to cope with the conditions in such complexity. The vast majority of these studies have focused on physiological responses as well as evolutionary trajectories in specific environments, and it has generally been found that the environment was a main driver of the outcomes of adaptation in a selective manner. However, most experiments on microbial evolution have been conceptually simple, as they compare populations adapting to rather uniform and 'basic' environments that are defined by specific limiting substrates. Hence, the effects of environmental heterogeneity and complexity, important factors that shape the evolution of populations, have not been frequently addressed. Therefore, the present work addressed bacterial evolution in a complex environment mimicking an industrial setting. We hypothesized that previous experiments have overlooked adaptation/evolution in complexity, as the simplicity of most experimental setups do not match such complexity. This thesis thus focused on investigating the establishment of divergence within an *Escherichia coli* population evolving under complex (medium) conditions. The emphasis was on unraveling the level and direction of diversification in respect of the genetic and physiological changes that the organism underwent, which allowed it to either acquire superior fitness (survival of the fittest) or occupy a different niche (niche differentiation).

In the experiments, the long-term (~1000 generations) adaptive responses of *E. coli K12* MC1000 in Luria-Bertani (LB) broth under three conditions (aerobic, fluctuating and anaerobic conditions), mimicking local conditions of systems of industrial fermentations, without a clearly defined limiting substrate were evaluated. A set of evolved forms, selected on the basis of colony morphology, was thus obtained and subjected to the analyses comparative to the ancestor. By using whole-genome sequencing (NGS), transcriptomics (whole-genome microarrays) and phenotypic microarrays, in chapter 2 and 3, a complete overview of the genomic and phenotypic changes that had occurred over experimental time was achieved. It was found that several genetic solutions (adaptive paths) had led to adaptation. The evolved forms showed enhanced fitness in competition experiments against the ancestral strain. Moreover, the level of adaptation was greatly influenced by the main environmental force (availability of oxygen).

In further work, a positive relationship between upregulation of gene expression and the number of mutations was observed, suggesting that a number of metabolic pathways were activated. The number and nature of the genetic and gene expression changes correlated with a better ability to deal with the environmental traits. Consistent reproducibility of evolutionary paths acting on genuine targets of selection was then observed, as parallel genetic and phenotypic responses were found across all populations. This suggested a response that was triggered by the LB medium as a key driving factor. The specific response was related to the gal operon (*galR* -a repressor of the gal operon- and *galE* -involved in the metabolism of galactose). The mutated forms were shown to exhibit enhanced growth on galactose as well as galactose-containing disaccharides. From a global perspective, such phenotypic innovations were hypothesized to illustrate the tight

relationship between genomic changes and ecological conditions (i.e the environment seen from a broad as well as a more local perspective).

Although unique genomic and phenotypic changes were observed in the evolved forms, suggesting habitat-specific selection, considerable heterogeneity was also found between populations across replicates of the same treatments and even at the intra-populational level. This was suggestive of genetic drift. Thus, genomic as well as phenotypic polymorphisms (of which some were related with mutations in regulatory genes) were observed in the parallel cultures. In the light of the differential phenotypic outcomes found, it was hypothesized that the parallel responses were affected by differing genomic backgrounds. A relationship between the relatively high number of genomic changes found in all the systems and the multiplicity of resources and complexity present in the medium may be presumed. Analysis of the polymorphisms in one evolved population (aerobic regime) revealed clear and distinct metabolic profiles in selected forms, with an interactive relationship. Thus, in chapter 4 a large number of colonies obtained from the selected end-population were metabolically characterized. This revealed the existence of two main metabolic types and a correlation of type with colony morphology. Whereas one type was able to rapidly consume carbohydrates, the other one was a slow grower and able to use a by-product that was presumably excreted as a result of the active metabolic action of the first type. This property offers the possibility to the second type to maintain the population under conditions of depleted resources.

Chapter 5 investigated the emergence of additional specific phenotypic traits (enhanced stress resistance properties and metabolic capabilities). Thus, key distinct capabilities were found to co-exist in the population, leading to the concept of a physiologically dichotomous population. First, alternative strategies feeding into the central metabolic pathway were apparently adopted by one type to be able to deal with the scarcity of limiting carbon sources (i.e. glucose). Second, another type revealed enhanced resistance or tolerance to stress conditions. The remarkable interactive and stable coexistence of these forms revealed the presence, in the two forms, of *trade-offs* of metabolic pace versus the raised tolerance to environmental stress. The experimental conditions (complex medium, temporal cycling of conditions between growth-supportive and stressing) apparently drove the coexistence of the two types found. This provides a clear example of niche partitioning in a combined sense, rather than just nutritional, resulting from divergent adaptive evolution. Thus, environment (defined by oxygen and carbon/energy source availability) was not the sole determinant of niche space but also the niche opportunities created through the growth and activities of the organisms. The complexity of industrial (next to natural) settings has the potential to trigger the establishment of adapted and coexisting forms.

Nederlandse samenvatting

Complexiteit is inherent aan zowel natuurlijke als industriele omgevingen ('habitats'). In de laatste decaden heeft wetenschappelijk onderzoek aangetoond dat microorganismen zeer flexibel zijn, zowel fenotypisch als genetisch, in hun capaciteit om onder complexe omstandigheden te gedijen. Een groot deel van de tot dusverre verrichte studies beschouwden de fysiologische en evolutionaire reacties in specifieke habitats, en algemeen werd habitat gezien als belangrijkste 'driver' van aanpassingsprocessen. Echter, de meeste studies van de microbiele evolutie waren "simpel" van opzet gezien het feit dat zij adaptatie in populaties in uniforme media beschouwden, waarin enkele substraten limiterend waren. Dit leidde ertoe dat effecten van heterogeniteit en complexiteit, wat toch sleutelfactoren zijn in de evolutie van populaties, niet zijn meegenomen. De huidige studie richtte zich derhalve op de bacteriele evolutie in een complexe habitat, als model voor een industrieel systeem, met als vooronderstelling dat de versimpeling in de habitats gebruikt in voorgaande studies het vinden van de habitat-relevante evolutionaire vormen heeft bemoeilijkt. Het proefschrift beschrijft het ontstaan van divergentie in een *Escherichia coli* populatie die zich ontwikkelt onder complexe medium condities. In het onderzoek lag de nadruk op de analyse van de mate van diversificatie, en de richting daarvan, met betrekking tot verhoging van fitness ("survival of the fittest") of de bezetting van een verschillende niche (nichedifferentiatie).

In de experimenten zijn de lange-termijn-aanpassingen (~1000 generaties) van *E. coli K12* MC1000 in Luria-Bertani (LB) bouillon (zonder voorondersteld limiterend substraat), onder drie condities (aeroob, wisselend en anaeroob; waarbij condities van industriele processen zijn nagebootst) geevalueerd. De fitness van een aantal geevolueerde vormen, geselecteerd op basis van hun verschillende koloniemorfologie, ten opzichte van het oudertype, werd daarna bepaald. Alle vormen vertoonden een sterk verhoogde fitness. Vervolgens werden de basenvolgordes van hun genomen, hun transcriptoom en hun metabole vermogen (aan de hand van fenotype microarrays) bepaald, resulterend in een uitgebreid overzicht van alle genetische en fenotypische veranderingen in de verschillende vormen (hoofdstukken 2 en 3). Verschillende genetische wegen werden zo blootgelegd, die allen tot fitnessverhoging leidden. Tenslotte werd het niveau van aanpassing sterk gemoduleerd door de belangrijkste variabele in het experiment, i.e. het zuurstofniveau.

In verder werk werd een positieve relatie gevonden tussen de mate van verhoging van genexpressie en het aantal mutaties, hetgeen suggereerde dat een aantal metabole routes (in plaats van een enkele) geactiveerd waren. De aard van deze genetische en genexpressie mutaties correleerde met het verbeterde vermogen om onder de locale condities te bestaan. Voorts werd, gezien het feit dat parallelle genetische en fenotypische responses werden vastgesteld in replicacultures, een consistente reproduceerbaarheid van de evolutionaire paden (waarvan de drivers op realistische selecteerbare functies acteerden) gevonden. Dit suggereerde een response die getriggerd werd door het groeimedium LB. Meer specifiek was de response gelieerd aan het gal operon (*galR*, repressor van het gal operon, en *galE*, betrokken bij het metabolisme van galactose). De gemuteerde vormen vertoonden verhoogde groei op galactose alsook op

galactose-bevattende disacchariden. Dergelijke fenotypische innovaties zijn een afspiegeling van de directe relatie tussen genoomveranderingen en ecologische omgevingscondities (de omgeving beschouwd vanuit een globaal alsook locaal perspectief).

Hoewel er unieke genomische en fenotypische veranderingen warden geconstateerd in de geevolueerde vormen (hetgeen habitat-specifieke selectie suggereert), werd er tevens een behoorlijke mate van heterogeniteit gevonden tussen de replicapopulaties van dezelfde behandeling, en zelfs binnen de populaties. Dit gaf genetische drift aan. Specifiek weren genomische (en fenotypische) polymorfismen gevonden, waarvan sommigen in regulatiegenen, in de replicaculturen. In lijn met de verschillende fenotypen, werd als hypothese genomen dat de parallel responses voortkwamen uit verschillende genomische achtergronden. Het bestaan van een mogelijk verband tussen het relatief hoge aantal genomische veranderingen dat gevonden werd in alle systemen en de veelheid aan verschillende substraten , alsmede de complexiteit, in het medium, werd gepostuleerd. Een analyse van de polymorfismen in een geevolueerde populatie (aeroob regime) liet duidelijk verschillende metabole profielen zien, met interactie tussen verschillende vormen. In hoofdstuk 4 wordt de metabole karakterisering van een groot aantal vormen van de geselecteerde populatie beschreven. Twee metabole hoofdtypen werden gevonden, waarbij er een correlatie tussen metabool type en koloniemorfologie bestond. Het eerste type vertoonde een hoge capaciteit om koolhydraten te gebruiken, terwijl het tweede type langzaam groeide en een bijproduct gebruikte dat met grote waarschijnlijkheid door het eerste type uitgescheiden werd. Deze eigenschap geeft het tweede type de capaciteit om voort te bestaan onder condities van uitputting van het hoofdsubstraat in de cultuur.

Hoofdstuk 5 beschrijft het onderzoek naar het optreden van additionele specifieke eigenschappen (verhoogde stressresistentie en metabole eigenschappen). De co-existentie van vormen met verschillen in sleuteleigenschappen binnen een populatie werd aldus aangetoond, hetgeen resulteerde in het concept van een fysiologisch dichotome populatie. Allereerst werden door een sleuteltype alternatieve strategieen - die voeding geven aan het centrale metabolisme - aangewend om beter met substraatschaarste (i.e. glucose) om te gaan. Ten tweede liet een ander geevolueerd type een verhoging van stresstolerantie of – resistentie zien. De co-existentie van de twee typen was opmerkelijk, en wees op *trade-offs* in beide typen in relatie tot metabool vermogen en stresstolerantie. De gebruikte experimentele omstandigheden (complex medium, wisseling van condities tussen groeibevorderend en stress) waren naar alle waarschijnlijkheid debet aan deze coexistentie, hetgeen een duidelijk voorbeeld is van "gecombineerde" nicheverdeling, i.e. naar voedings- alsook stressniveau. Derhalve was 'habitat' (gedefinieerd naar zuurstof – en substraatbeschikbaarheid) niet de enige factor die de niche bepaalde, maar speelden zich-vormende niches door groei en activiteit van de microorganismen zelf ook een grote rol. De complexiteit van industriele, naast natuurlijke, habitats kan, waar het microbiele productieprocessen betreft, derhalve resulteren in de vorming van aangepaste en coexisterende vormen.

Resumen

Una característica inherente de los ambientes naturales e industriales es la complejidad. En las ultimas décadas, estudios científicos han revelado las capacidades genéticas y fenotípicas de los microorganismos al enfrentarse a dichas condiciones. La gran mayoría de estos estudios se ha enfocado en las respuestas fisiológicas así como en la trayectoria evolutiva en ambientes específicos, y se ha concluido que el ambiente es un factor selectivo y determinante en el resultado evolutivo. Sin embargo, la mayoría de los experimentos sobre evolución microbiana han sido conceptualmente simples ya que comparan las poblaciones en su adaptación a ambientes uniformes y 'básicos' definidos por substratos limitantes y específicos. Por lo tanto, los efectos de la heterogeneidad ambiental y la complejidad, factores importantes que moldean la evolución de poblaciones, no han sido frecuentemente estudiados. Por consiguiente, el presente trabajo estudió la evolución bacteriana en un ambiente complejo, simulando un entorno industrial. Otros experimentos han pasado por alto la adaptación / evolución durante condiciones complejas, así, la sencillez de la mayoría de los montajes experimentales no han coincidido con dicha complejidad. Esta tesis se enfoca entonces en investigar el establecimiento de la divergencia ocurrida dentro de una población de *Escherichia coli* evolucionando en condiciones complejas (medio de cultivo). El énfasis estuvo en revelar el nivel y la dirección de la diversificación con respecto a los cambios genéticos y fisiológicos que el organismo sufrió, los cuales le permitieron adquirir eficacia biológica (supervivencia del más apto, *fitness*) u ocupar un nicho distinto (diferenciación de nicho).

En los experimentos, las respuestas adaptativas a largo-plazo (~1000 generaciones) de *E. coli K12* MC1000 en caldo Luria-Bertani (LB), en tres condiciones (aeróbica, fluctuante y anaeróbica), simulando condiciones locales de sistemas industriales sin un substrato claramente definido, fueron evaluadas. Un set de formas evolucionadas fue seleccionado con base en la morfología de colonias y sujeto a un análisis comparativo contra el ancestro. Por medio del uso de secuenciación de genoma completo (NGS), transcriptomica (microarreglos de genoma completo) y microarreglos fenotípicos, en el capitulo 2 y 3, se logró una observación completa de los cambios genéticos y fenotípicos ocurridos. Se encontró que varias soluciones genéticas llevaron a la adaptacióñas formas evolucionadas mostraron incremento de su eficacia biológica durante experimentos competitivos con la cepa ancestral. Adicionalmente, el nivel de adaptación fue influenciado por la principal fuerza ambiental (disponibilidad de oxígeno).

Posteriormente, se observó una relación directamente proporcional entre la regulación positiva de expresión genética y el numero de mutaciones, sugiriendo que un gran número de vías metabólicas fueron activadas. El número y naturaleza de los cambios genéticos y de expresión genética, correlacionaron con una mejor habilidad para lidiar con las características ambientales. Se observó una reproducibilidad consistente de las vías evolutivas actuando sobre blancos genuinos de selección, ya que respuestas genéticas y fenotípicas se encontraron consistentemente en todas la poblaciones. Lo anterior sugirió una respuesta provocada por el medio de cultivo (LB) como factor clave y determinante. Específicamente, una respuesta

relacionada con el operon gal (*galR*- un represor de el operon gal- y *galE* – relacionado con el metabolismo de galactosa). Las formas mutadas exhibieron mejor crecimiento en galactosa así como también en disacáridos conteniendo galactosa. Desde una perspectiva global, dichas innovaciones fenotípicas sugirieron una relación fuerte entre los cambios genómicos y las condiciones ecológicas (el ambiente visto desde una perspectiva amplia así como también una perspectiva local).

Aunque cambios genómicos y fenotípicos fueron observados en las formas evolucionadas, sugiriendo selección especifica de hábitat, una heterogeneidad considerable fue observada entre las poblaciones replica pertenecientes al mismo tratamiento e incluso hasta un nivel intra-poblacional. Esto sugirió la presencia de deriva genética. Así, polimorfismos genómicos y fenotípicos (de los cuales algunos relacionados con mutaciones a nivel regulatorio) fueron observados en cultivos paralelos. A la luz de resultados fenotípicos diferenciales, se concluyo que las respuestas paralelas estuvieron afectadas por los diferentes antecedentes genómicos. Se pudo presumir una relación entre el relativamente alto numero de cambios genómicos, la multiplicidad de recursos y la complejidad presente en el medio de cultivo. El análisis de polimorfismos en una de la poblaciones evolucionadas (régimen aeróbico) revelo un perfil metabólico en las formas seleccionadas claramente distinto pero teniendo además una relación interactiva. Así, en el capitulo 4 un numero de colonias obtenidas a partir de la población seleccionada fueron caracterizadas metabólicamente. Esto reveló la existencia de dos tipos metabólicos y una relación de tipo con morfología de colonia. Mientras un tipo fue capaz de consumir rápidamente carbohidratos, el otro tipo creció más lentamente pero con la capacidad de usar un producto presumiblemente excretado, como resultado de la acción metabólica del primer tipo. Esta propiedad ofreció la posibilidad al segundo tipo de mantener la población bajo las condiciones de recursos limitados.

El capitulo 5 investigó la presencia de características fenotípicas adicionales (mejor resistencia de estrés y capacidades metabólicas). Así, distintas capacidades fueron encontradas en la población coexistente, sugiriendo la presencia de una población dicotómica. En ella, un tipo adoptó la capacidad de usar estrategias alternativas que conllevan a la vía metabólica central adjudicándole la capacidad de sobrevivir a la escasez de recursos de carbono limitados (glucosa). Otro tipo reveló una mejor resistencia o tolerancia a las condiciones de estrés. La coexistencia estable e interactiva entre estas formas reveló la presencia de *trade-offs,* relacionados con metabolismo versus tolerancia a estrés. Las condiciones experimentales (complejidad del medio de cultivo y las condiciones fluctuantes entre crecimiento y estrés) aparentemente llevaron a la coexistencia de los dos tipos observados. Esto proporciona un claro ejemplo de partición de nicho como resultado de una adaptación evolutiva divergente. Así, el ambiente (definido por la disponibilidad de oxígeno y por las fuentes de carbón y energía) no solo fue determinante en el espacio del nicho, sino también en las oportunidades del nicho creadas atreves del crecimiento y actividades propias de los organismos. La complejidad de los entornos industriales (y naturales) tiene el potencial de desencadenar la creación de formas coexistentes y adaptadas.

Acknowledgments

" ...I have learned that everybody wants to live on top of the mountain,
without knowing that true happiness is obtained in the journey
taken and the form used to reach the top of the hill".

Gabriel García Márquez

I would like to express my special appreciation and thanks to my supervisor Professor Dirk van Elsas, for giving me the opportunity to follow this PhD, for encouraging my research, and for all the motivation and advices to complete this work. Thanks for all the critical and helpful inputs. Thanks because although this was a new and challenging subject to the group, you believed in me to perform it. You taught me the importance to grow as a guided but independent researcher. After this, I learned that a scientific mind should not have boundaries.

I would also like to acknowledge our colleagues from University of Copenhagen and DTU for their collaboration. Thanks to Professor Søren Sørensen for his valuable advice. My short stay in Copenhagen was really crucial to this work. Thanks also to Martin A. Hansen for helping me dive into the waters of Linux. I really enjoyed my time in DK and thank all the colleagues of the Molecular Microbial Ecology group, who were very kind and also introduced me to what is now my favorite sport, running. I also want to thank the members of the Molecular Genetics group (GBB) of the University of Groningen, Professor Oscar Kuipers for his corrections and for allowing me to work in their facilities. I am very grateful to Dr. Ákos T. Kovács, for his patience, time, advice and guidance in the transcriptomics work.

I own special thanks to all the members of the Microbial Ecology group. The ones who already left and were very helpful from the beginning of my work. Many thanks to Pablo, Patricia, Eelco, Riitta, Anna. Ozgül, "my vecinita", you were and still are a good friend. The sun that you left hanged on our shared lab-bench kept shining on my days. Sasha, thanks for all the help with the numbers. My dear Pakistani friend Rashid, I missed your company and all the life and scientific talks in the office. My dear "señorita" Silvia, with whom I shared memorable times from the very beginning, thanks for being an example of order and discipline to me and for the scientific and life advices. Michele, thanks because your door was always open to help me, I miss our conversations. Also I want to thank people who passed by the lab for a short time and left nice memories during this time: Elisa, Simone, Armando, Sandy, Jonas, Esmeralda, Domen, Muhammad, Manoj and my student Wicher. I am also very grateful to the current members of the group whom in one way or another have listened to my stories: Irene, Daniel, Victor, Francisco, Larisa, Maria Julia, Memoona, Pu, Miao W. Special thanks to Cyrus because your positive and 'awesome' answer to my morning greeting was always a good start of the day. To Irshad, your crazy laughs also made my days, thanks for the constant compliments to my work. Miaozhi, thanks for your help in the lab when I was in need for an extra hand and for being so nice to me. To Stephanie J., the 'Black and Bloom' times were always so enjoyable, thanks also for the English corrections. Diego and Maryam, it is always nice to know that 'compatriotas' are so close to help and to share familiar moments. To my later office mates and friends: Mukil, thanks for your company and useful advices and Steph G., I found in you a good, helpful and special friend. Finally, thanks to Jolanda, you were always open to help, thanks for sharing your artistic talents with me. To Joana because

your scientific opinions are always constructive, also for believing in me and having trust on my scientific capacities. I also want to thank Joukje for her bureaucratic help.

When I arrived to the Netherlands I was looking for a project to confirm that beyond my capacities as a microbiologist working for the industrial world I had a scientific mind. I think I could not have been luckier having to work on this subject to confirm it. Before I started my PhD I knew that the more I was learning about microorganisms, the more I was being fascinated with them. After learning a bit about evolution and observing how they were responding to what I was imposing them, my fascination and admiration to these 'simple-looking' but practically perfect forms of life became endless.

The time of my PhD was accompanied with very important events of my life, good ones and bad ones. I had a difficult start. As my friend Ozgül well described, many things happened to me 'like in a black comedy'. From moving around Groningen like a pilgrim to everybody's house (by the way, Dick, Joana, Michele, Silvia, thanks for the hospitality in that time) to having unfortunate news like loosing my Grandpa and bad news compromising my health. Luckily, with time the things improved. I finally settled down, adapted and even took important steps in life like getting married and becoming a mother. All these events together with my professional development made of this time an important and very special landmark of my life.

I also want to thank all the people that close or far, in one way or another accompanied me during this process. I owe my gratitude to Hans, Yvonne, John, Rolf, Suzanne and Thierry, thanks for all your support and for being always a warm family and having open arms for me. My endless gratitude goes to my family in Colombia, the distance was not an issue to feel your support, no words can even describe what I feel for you and the gratitude I have for all your love. My sister Diana, thanks for introducing me to NL and Ale 'mi mirris', I couldn't be happier and more proud that you chose to study Biology; we will have great scientific conversations in the future. I love you and miss you both, thanks for being both my confidents and 'amiguis'. Madre y Padre, les debo mi vida pero también las alas que me regalaron para volar y cumplir mis sueños, infinitas gracias, los amo.

Last but not least, thanks to my lovely husband, Hans. You filled my days with the certainty that I could count on you, as my partner, my friend, my confident, my advisor, my personal trainer. Your unconditional love and continuous support to follow this and all my dreams are blessings in my life; I could not have made it without you. Finally, to Anna Lucía who came to my life and made me feel the most amazing feelings as a mother. Her daily smiles are fuel and each one of them means the world to me.