Evolution and systems biology

Csaba Pál

Synthetic and Systems Biology Unit, Biological Research Center

Hungarian Academy of Sciences

Szeged

To the memory of Dr. István Molnár (Monya)

Table of contents

I. Introduction

Integration of Mendelian genetics into evolutionary biology has allowed better understanding a wide range of biological problems, and unified several previously isolated fields including biogeography, taxonomy, ecology and population genetics. In spite of the apparent achievements of the Modern Synthesis, several important issues have remained unanswered[1]:

1) How do genes evolve[2,3]? Recent advances in genomics catalyzed a move from investigating individual genes to characterizing the impact of cellular subsystems. These studies demonstrated that protein evolution is influenced not only by protein structure and function. Rather, gene expression level, context of biological networks also matter. An integrated theory that unites protein evolution with biochemistry and functional and structural genomics is still missing.

2) In the early 1920s Ronald Fisher advocated that evolution is a hill-climbing process: it proceeds through accumulation of beneficial mutations (Figure 1). By contrast, Sewall Wright suggested that accumulation of conditionally harmful mutations act as stepping stones by providing access to evolutionary pathways which are otherwise inaccessible[4]. After many decades, theoretical works on the subject are overwhelming, but the data (especially on a genomic scale) is scarce[5].

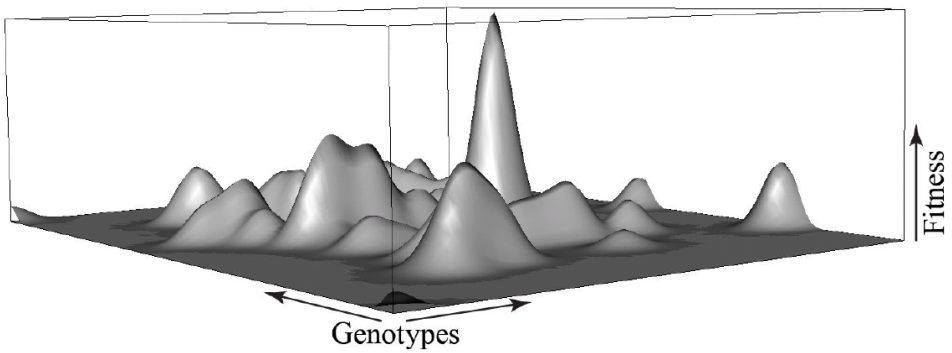


Figure 1. Adaptive landscape is a conceptual tool to visualize the relationship between genotypes and fitness. It assumes that genotypes have well defined replication rates in a fixed environment, irrespective of the frequency of other genotypes in the population. Fitness is defined as the "height" of the landscape. Genotypes next to each other are mutational neighbors. One of the most perplexing issues in evolutionary biology relates to the general forms of the landscape, i.e. the presence of isolated peaks and the possibility of evolution to proceed through suboptimal states (fitness valleys).

3) Evolutionary change is often considered to be contingent on initial conditions and chance events, and therefore unique on the one hand, and owing to predictable adaptive changes, replicable on the other hand. It remains unclear how far evolution is predictable at the genomic level[1]. Is it possible to infer which genes are most likely to be subject of adaptive mutations, and how adaptation at the phenotypic level will proceed? This goal requires investigating the relative roles of historical contingency, neutrality and adaptive changes during evolution. These problems have long been investigated at the level of individual proteins, but little is known about the evolution of large cellular subsystems[1].

4) Many steps in protein production are strikingly error-prone, although such errors can lead to reduction of fitness and genetic diseases. It has remained an open issue how cells achieve robustness to errors during information transmission[6,7] .

5) How do evolutionary novelties arise? An influential theory suggests that evolution initiates new enzymatic functions by utilizing the weak side activities of preexisting enzymes[8]. However, the extent to which underground reactions provide novelties is largely unknown[9].

Systems biology offers a new angle to study these problems in a consistent manner[1,10]. In a nutshell, it integrates detailed analyses of molecular networks, in silico modeling and laboratory evolution with the aim to study central issues in evolutionary biology[10] (Table 1-2).

Layer of prediction	Importance	Current state of knowledge	Difficulty	New tools and knowledge needed for integration
Distribution of mutational effects and epistatic interactions	General architecture of adaptation. Robustness against mutations	Wealth of systematic gene deletion studies and epistasis maps	Existing fitness landscape models are not biologically detailed. High-throughput experiments are restricted to a few environmental conditions, or they only consider null mutations	Realistic systems-biology models offer new predictions on mutational effects and mechanistic insights. New types of experimental data (for example, fitness profiling of point mutations or gene overexpression studies)
General patterns of genome evolution	Evolutionary forces driving protein and expression divergence, gene loss, horizontal gene transfer and gene duplicability	Impact of post-genomic features (for example, gene expression or network position)	No clear relationship between fitness and the post-genomic gene features studied	Predictions and measuring most relevant physiological data (for example, range of neutrality, optimal gene activity or physiological coupling between genes)
Specific evolutionary trajectories	Relative importance of chance and necessity in evolution. Predictive tools for applications	Map of adaptive landscape for single proteins. Insight from experimental evolutionary studies	Difficult to map adaptive landscapes for large cellular subsystems empirically. Interpretations dominate over predictions	Modelling the outcome of adaptive evolution at the molecular level. New experimental technologies to map adaptive landscapes

Table 1. Some major conceptual issues in evolutionary systems biology. Adapted from Papp et al. 2011.

Modelling approach	Examples	Scale	Data requirement	Phenotypes predicted	Advantages	Disadvantages
Digital organisms (that is, self-replicating computer programs)	Avida platform	Potentially large	No biological data required	Replication rate; performance of mathematical operations	Infers general principles of evolution	No direct connection to specific biological systems
Graph-theoretical models	Wide range of cellular interaction networks	Large	Large-scale molecular interaction data	Network properties, such as diameter and centrality	Low data requirement; insights into similarities of network architecture across species	Unclear how network architecture relates to cellular physiology and fitness
Kinetic biochemical models	Metabolic pathways, gene regulation and cell cycle	Small	Detailed: molecular interactions and kinetic information	Reaction fluxes; component concentrations	Conceptual understanding; realistic; quantitative; captures dynamics	Only available for small-scale systems
Logical models	Cell cycle, signalling and metabolism	From medium to large	Qualitative knowledge of molecular interactions	Activity states; viability; dynamic behaviour	Low data requirement; captures dynamics to some extent	Difficult to capture continuous molecular response in a discrete model
Constraint-based models	Flux balance analysis of genome-scale metabolic networks	From medium to large	Network reconstruction based on omics data; biochemical and physiological studies	Growth properties; reaction fluxes across conditions	No enzyme kinetic information required; testable predictions on a genomic scale	Basic models lack dynamics in time; metabolite concentrations are not captured

Table 2. Modeling frameworks for evolutionary systems biology. Our work primarily focused on constraint-based models, such as flux balance analysis. Adapted from Papp et al. 2011.

The logic is as follows. First, I will give a very brief overview of the main works I have been involved in. I will continue with focusing on four main topics, all of which illustrate the conceptual and methodological links between evolution and systems biology. The first chapter relates to the gene knock-out paradox. Why is it that most genes appear to be dispensable? These considerations will lead to the problem of compensatory evolution, a topic described in detail in chapter 2. The third chapter describes recent advances in bacterial genome engineering, and how this discipline can be employed to test central issues in evolution. The final chapter is devoted to a brief summary on antibiotic resistance and collateral sensitivity in microbes.

II. A brief summary of research

Genome evolution

In 2001, we demonstrated for the first time that highly expressed genes evolve slowly[11] (Figure 2.). Later, we argued that evolutionary rate of a protein is predominantly influenced by its expression level rather than functional importance[12]. Many consider these works as a start of a paradigmatic shift in the field of protein evolution[2]. Eugene Koonin wrote about the four major laws of genome evolution[13], and suggested that expression level-protein evolutionary rate is one of them.

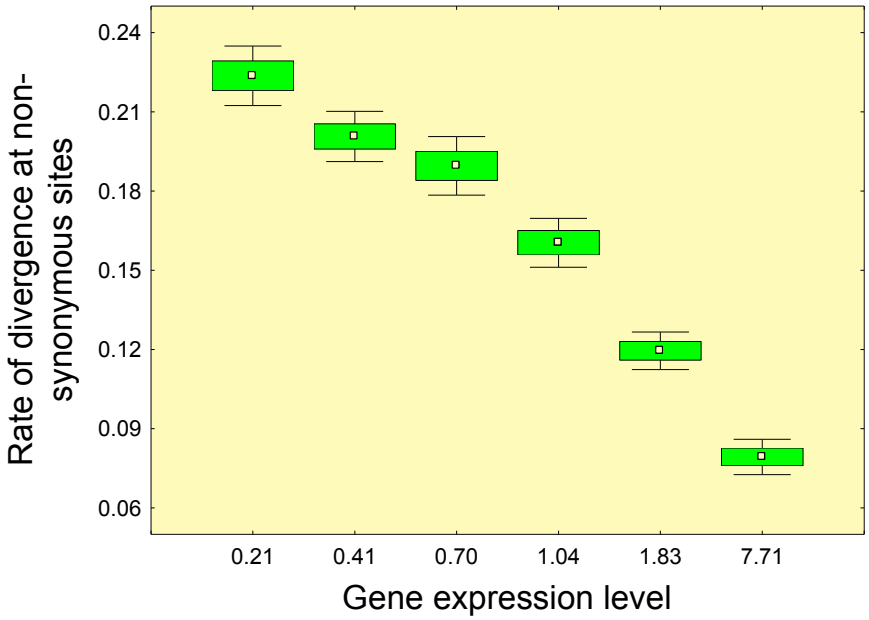


Figure 2. The figure shows the rate of protein evolution in yeast as a function of mRNA expression level. Rate of evolution was estimated by nucleotide sequence distances at non-synonymous sites. For details, see Pal et al. 2001 and Pal et al. 2006.

Next, we studied molecular mechanisms underlying dosage sensitivity[14]. In this paper, we developed and tested what is now known as the dosage balance hypothesis[15]. The hypothesis offers a synthesis on seemingly unrelated problems such as the evolution of dominance, gene duplicability and co-evolution of protein complex subunits. Predictions of the hypothesis have been confirmed in many eukaryotic organisms, and now it appears to be an important unifying model with implications on human genetic diseases[16].

In 2007, we demonstrated that antagonistic co-evolution with parasites has a large impact on the evolution of bacterial mutation rate[17]. This paper showed for the first time how biotic interactions shape mutation rate evolution.

Recently, the Pál lab explored the consequences of compensatory adaptation on gene content evolution[5]. It is well known that while core cellular processes are generally conserved during evolution, the underlying genes differ somewhat between related species. We demonstrated that gene loss initiates adaptive genomic changes that rapidly restores fitness, but this process has substantial pleiotropic effects on cellular physiology and evolvability upon environmental change[5].

Network evolution

We also had a pivotal role in establishing the emerging field of evolutionary systems biology[1]. Our research focused on understanding the extent to which evolution is predictable at the molecular level. We realized that genome-scale metabolic network modeling combined with experimental tools offers an unprecedented opportunity to study some of the most difficult problems in evolution, such as mutational robustness[18], horizontal gene transfer[19], genome reduction[20], epistasis[21,22], promiscuous enzyme reactions[9], and complex adaptations[23]. The approach developed by our group is now a major trend, and has been adopted by others[24].

Antibiotic resistance

The Pál lab currently studies the problem of antibiotic resistance using tools and concepts borrowed from systems biology[25-29]. By combining laboratory evolution, genome sequencing, and functional analyses, we charted the map of evolutionary trade-offs between antibiotics. We made the striking discovery that mutations that cause multidrug resistance in bacteria, simultaneously enhance sensitivity to many other unrelated drugs (collateral sensitivity), and explored the underlying molecular mechanisms[25]. As a result, the concept of collateral sensitivity is emerging as one of the leading concepts in antibiotic resistance research [30].

Genome engineering

Finally, the Pál lab is an advocate of the emerging field of evolutionary genome engineering[31]. These technologies enable the modification of specific genomic locations in a directed and combinatorial manner, and allow studying central evolutionary issues in which natural genetic variation is limited or biased. However, current tools have been optimized for a few laboratory model strains, lead to the accumulation of numerous undesired, off-target modifications, and demand extensive modification of the host genome prior to large-scale editing. We presented a simple, all-in-one solution[32,33] . The method is unique as it allows systematic comparison of mutational effects and epistasis across a wide range of bacterial species.

III. Evolution of gene dispensability

Key publications: Papp, Pal & Hurst 2004, Pal et al. 2005 (see Appendix)

In most organisms, deletion of a single gene generally has no impact on fitness and survival[34]. Only 20% of the single knock-outs in yeast *Saccharomyces cerevisiae* are

essential for growth, and similarly low figures have been observed in the worm *Caenorhabditis elegans*, *Bacillus subtilis*, and many other organisms (Table 3).

Organism	Number of protein coding genes	Estimated % of essential genes
Helicobacter pylori	1590	17%
Haemophilus influenzae	1743	38%
Staphylococcus aureus	2595	25%
Mycobacterium tuberculosis	3924	15%
Bacillus subtilis	4100	6.6%
Escherichia coli K12	4289	16.5%
Salmonella typhimurium LT2	4450	11%
Saccharomyces cerevisiae	5780	19%
Caenorhabditis elegans	19099	7%

Table 3. Distribution of essential genes in model organisms. Adapted from reference[34]. Details and references therein.

These patterns raise many problems: Are these genes truly dispensable to the organism? Why is it that a knockout can grow well in the laboratory? Here I briefly address advance in our knowledge by paying particular attention to metabolism.

If certain genes would be truly dispensable, it would require that a deletion of the gene would not be under selection. Unfortunately, current lab assays have limitations, for two reasons[34]. They don't have the ability to measure fitness at the necessary resolution and second, they fail to identify genes that contribute to fitness in nature, but not in standard laboratory conditions.

Recent works indicate that seemingly dispensable proteins are generally under strong selection, i.e. they evolve much slower than expected for non-

functional, neutrally diverging sequences[12,35]. Thus, although knowledge on the exact physiological or evolutionary roles of these proteins is often patchy, to say the least, they apparently do something useful for the organism.

A case study on yeast metabolism

Both computational and empirical studies indicate that dispensability is more apparent than real: many genes have important functions in special conditions only[18,21,36]. In 2004[18], we addressed this issue first using the genome scale metabolic network model of baker's yeast (*Saccharomyces cerevisiae*) (Figure 3−4.).

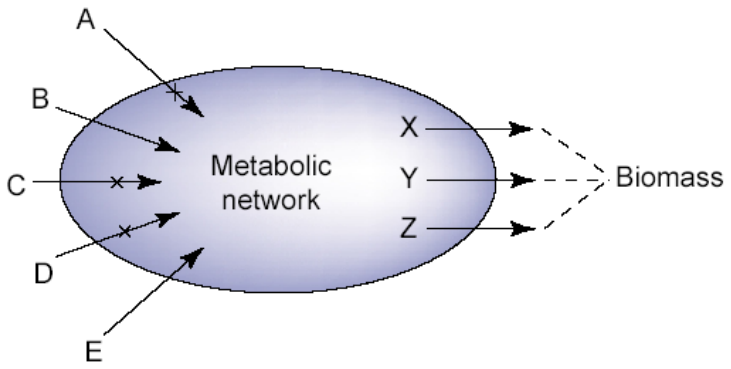


Figure 3. The essence of computational flux balance analysis. The analysis starts with the reconstruction of the metabolic network of a specific organism from genomic and detailed biochemical studies. The reaction set contains data on transport processes, direction and stoichiometry of reactions, and major metabolic components (X,Y,Z) important for the cell. Also the nutrients available in the environment (B,E) must be predefined in a way to mimic the experimental nutrient conditions. Finally, given the set of constraints – e.g. the reaction set and outer nutrients available for the cell – flux balance analysis calculates biomass production (a proxy of growth rate) in steady state.

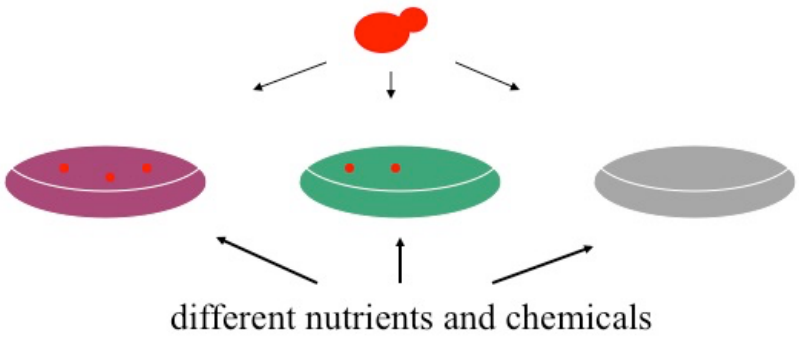


Figure 4. The predictions of flux balance analyses are tested on the wild-type and mutant yeast strains under a variety of conditions.

The metabolic network of yeast was reconstructed from a large set of prior biochemical studies, and includes 809 metabolites connected by 851 different biochemical reactions[18]. Using this network, we defined a solution where fluxes of all metabolic reactions in the network satisfy the relevant constraints, given the nutrients available in the environment. Next, we calculated the optimal use of the metabolic network to produce major biosynthetic components for growth under a set of 282 predefined and ecologically relevant nutrient conditions.

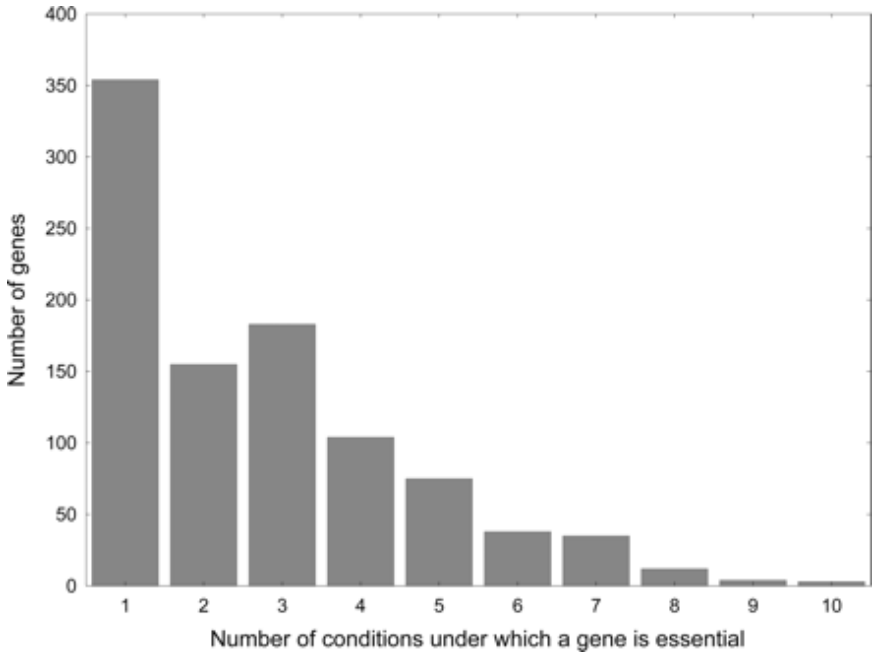
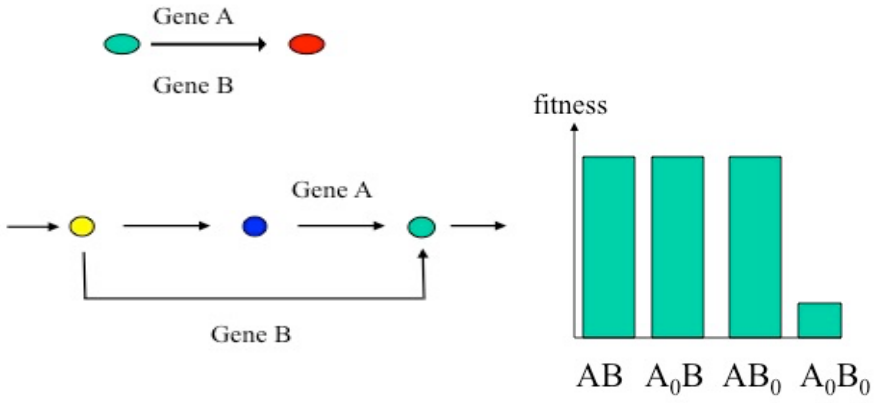


Figure 5. The figure shows the result of flux balance analysis. At least 20% of the 'dispensable' yeast metabolic genes are essential under some special environmental conditions. Adapted from Papp et al. 2004.

The model indicates that most metabolic genes have severe fitness defects only under a small fraction of the 282 different growth conditions investigated (Figure 5). Thus, most genes appear to be important in specific environments only[18]. Several empirical studies supported this claim. First, direct measurement of enzymatic fluxes in yeast demonstrated that about half of the apparently dispensable genes are inactive under laboratory conditions[37,38]. Even more importantly, a recent high-throughput chemogenomic study indicates that as high as 97% of the 5000 apparently nonessential genes in yeast make contribution to fitness under at least one environment[36]. Moreover, deleterious phenotypes are generally restricted to a small fraction of the tested environments[36].

The above figures do not exclude the possibility for other mechanisms of gene dispensability [39].



A B

Figure 6. Two major mechanisms contributing to robustness to gene deletion in metabolic networks. A gene may appear to be dispensable if another copy executes the same enzymatic reaction (a form of genetic redundancy). Alternatively, two genes may appear on alternative pathways producing the same end-product (distributed robustness) (A). As a result, only the genotype with deletion of both A and B (A0B0) shows fitness deficit (B).

Gene deletions may be compensated for by a gene duplicate with a redundant function, and reorganization of metabolic fluxes across alternative pathways may buffer gene loss[18,39] (Figure 6). In agreement with expectation, duplicated genes in yeast and worm are less likely to be essential than single copy genes. We hasten to note however, that this pattern may also reflect that genes encoding non-essential functions preferentially undergo gene duplication[40]. The presence of alternative pathways (a form of distributed robustness) is a more controversial issue, but clear-cut examples from metabolism nevertheless exist.

To approach which of the two mechanisms – gene duplicates with redundant functions versus alternative pathways – are more important, we again turned to yeast metabolism[18]. We focused on essential enzymatic reactions, i.e. the ones predicted to stop growth when deleted. Overall, we estimate that duplicates account for between 15−28 percent of incidences of gene dispensability, while alternative metabolic pathways can only explain 4 to 17 percent of gene dispensability. These figures were later confirmed by experimental enzymatic flux measurements in the same species (Figure 7). These experiments suggest that, for 207 viable mutants of active reactions, network redundancy through duplicate genes is the major (75%), and alternative pathways is the minor (25%) molecular mechanism of genetic network robustness. These results do not exclude the possibility that distributed robustness via alternative pathways is more common in other cellular systems.

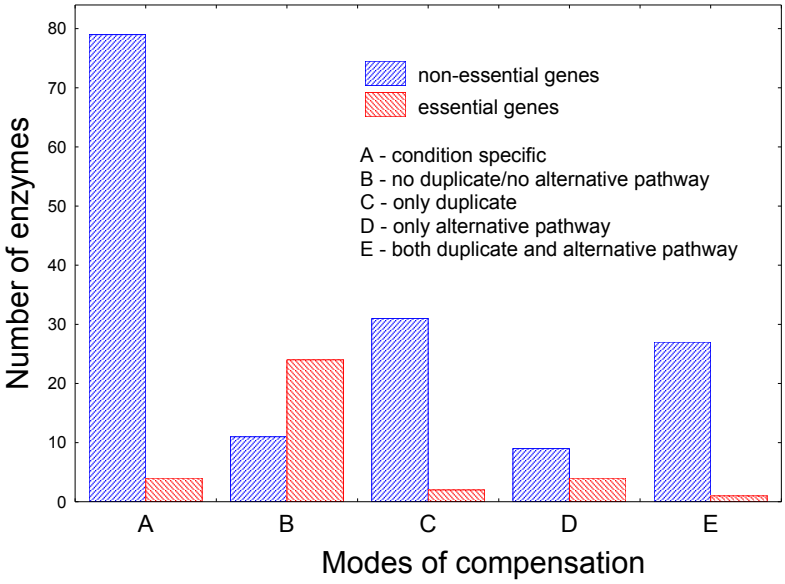


Figure 7. Gene dispensability in the metabolic networks of yeast. The classes are: (A) enzymatic reactions predicted to have zero flux under nutrient-rich conditions, but non-zero flux in at least one other environment (condition specific); (B) single-copy

genes predicted to catalyze essential reactions; (C) duplicate genes predicted to catalyze essential reactions; (D) single-copy genes predicted to catalyze dispensable reactions; and (E) duplicate genes predicted to catalyze dispensable reactions. When comparing groups B and C, of the 68 metabolic genes that are predicted to catalyze essential reactions, 33 are known to have a duplicated isoenzyme. Only about 6% of those that have an isoenzyme are observed to be essential in vivo, whereas the proportion of essential genes is roughly 69% among those without an isoenzyme. When comparing groups B and D, of the 47 single-copy genes 35 are predicted to catalyze essential reactions whereas 12 are predicted to be dispensable.

Next we asked whether the spread and retention of a duplicate was selected because it provided backup against mutations[18]. Prior theoretical works demonstrated that under realistic mutation rates and population size settings, most organisms are unlikely to evolve backup against mutations. So, why are duplicates in the genome? Flux balance analysis of the yeast metabolic network has shown that essential reactions are not more likely than nonessential reactions to be catalyzed by isoenzymes. Instead, isozymes appear at positions in the network where a high flux is needed. This suggests that duplicates were retained to permit a selectively advantageous increase in flux rates, a secondary consequence of which can be buffering[18].

The situation is similar for robustness provided by network architecture. A priori it is difficult to see how biological networks might evolve step-by-step to permit distributed robustness. A recent simulation study showed that robust network architecture emerged as a side consequence of selection for fast microbial growth rather than for enhanced robustness against mutations[41]. Another way to ask about the evolution of distributed robustness in networks is to ask about the evolution of

gene pairs that are not sequence related, but can compensate null mutations in each other. At least 51 percent of such synthetic lethal interactions are restricted to particular environmental conditions [21]. These results are compatible with a side effect model, where the enzymes are essential under nutrient specific conditions, not because they provide backup.

IV. Compensatory evolution

Relevant publication: Szamecz et al 2014 (Appendix)

Genetic disorders in human populations are surprisingly frequent[42]. However, individuals carrying the same deleterious mutations often have different or no symptoms at all. Moreover, mutations deleterious in human are frequently fixed in other closely related species[43,44]. Why is it so? In this short chapter, we argue that evolutionary adaptation is inherently linked to the incorporation of mutations with pleiotropic side consequences. Therefore, organisms undergo major changes during evolution not simply to adapt to novel environments, but also to compensate for the deleterious side-effects of adaptive mutations.

Premise 1. Harmful mutations are commonplace

All humans carry deleterious mutations in their genome sequence[45]. A recent analysis indicates that an average healthy person has 100 nonfunctionalized alleles, 20 of which are homozygous but with only mild phenotypic consequences[46]. In yeast, as high as 12% of the coding SNPs are predicted to be slightly deleterious [47].

Premise 2. Mutational effects depend strongly on the genetic context

In spite of the prevalence of harmful mutations, mutational effects vary due to epistatic interactions with other mutations. The evidences come from many different sources:

Human populations. Classic ‘‘monogenic’’ disorders show clear genetic background effects. For example, patients carrying the same deleterious allele present a broad range of clinical symptoms, most likely due to the action of modifier loci[48]. Strikingly, a recent large-scale study identified 13 adults harboring mutations for severe Mendelian conditions, with no clinical manifestation of the indicated disease[49]. The study indicates that penetrance of disease is influenced and potentially buffered by other mutations in the genome.

Systematic mutational screens. Studies in yeast, *C. elegans,* and human cell lineages revealed that the severity of phenotypes due to loss-of-function mutations differ significantly across genetic backgrounds[50]. Most notably, Vu and colleagues compared loss-of-function phenotypes of 1,400 genes in two *C. elegans* isolates that differ genetically by 1 SNP per 800 bp[51]. Strikingly, 20% of the genes have different loss-of-function phenotypes in two individuals and the differences in mutant phenotypes were predictable from expression[51].

Similarly, recent studies surveyed the set of essential genes in human cancer cell lineages[52,53]. Although they identified a coherent and overlapping set of essential genes in two related haploid cell lines, the essentiality of some genes is context-dependent and affects viability in a cell type-specific manner[54].

Laboratory evolution. The best evidence comes from studies on individual proteins. They unequivocally demonstrate that mutational effects are context dependent: mutations neutral or deleterious in one genetic background can be beneficial in

another[55]. Moreover, such studies indicate that adaptive evolution frequently demand prior fixation of other, so called permissive mutations[56-58]. These mutations do not alter the molecular function of the protein, but are necessary to tolerate large-effect mutations that cause shift in specificity and are generally destabilizing protein structure.

Premise 3. Mutational effects are condition specific

It has also long been noted that mutational effects very much depend on the environment. In most organisms, inactivation of a single gene generally has no major effect on survival in a particular condition. Only 20% of the single knock-outs in yeast *Saccharomyces cerevisiae* are essential for growth, and similarly low figures have been observed in many other species[59]. However, gene dispensability is more apparent than real. Most genes appear to be important in specific environments only. A recent high-throughput chemogenomic study indicates that as high as 97% of the 5000 apparently nonessential genes in yeast make contribution to fitness under at least one condition[36]. Moreover, deleterious phenotypes are generally restricted to a small fraction of the tested environments[36]. Similarly, in diploid yeast, haploproficiency phenotypes (increased growth rate when one copy is deleted) are surprisingly frequent, but are restricted to specific environmental contexts only[60].

Premise 4. Mutations with antagonistic effects are prevalent

Traditionally, mutations are divided into three categories: deleterious mutations, effectively neutral, and beneficial mutations. The above considerations (premises 2 and 3) demonstrate that categorization of mutations depends very much on the genomic background and the environments considered. Highly deleterious mutations can be neutral or even beneficial in other genetic or environmental conditions. Here we argue that mutations with such antagonistic effects are very common, and they influence evolutionary processes. First, a wealth of comparative and experimental

data have confirmed that, when organisms evolve to a given environment, the beneficial changes accumulated in one trait are generally linked to detrimental changes in other traits[61,62].

Such negative trade-offs shape the evolution of gene content as well. Laboratory evolution studies showed that adaptive loss-of-function mutations have an important role in the adaptation to a new environment[63]. As loss-of-function mutations are much more frequent than gain of function mutations, the contribution of gene loss to adaptive evolution might be higher than previously anticipated. Probably the most convincing study comes from the Zhang lab[62]. By measuring the fitness difference between the wild-type and null alleles of approximately 5,000 nonessential genes in yeast, the authors found that in any given environment, yeast expresses hundreds of genes that harm rather than benefit the organism.

Premise 5. Mutations, highly deleterious in one species, are fixed in another.

Recent comparative genomic studies revealed that disease-associated mutations in human are present in mouse strains with no apparent phenotypic consequences[43,44]. The best hypothesis to explain these patterns are that the majority of fixations of disease mutations in mice are due to compensatory genetic changes, which minimize the phenotypic consequences of these mutations.

Premise 6. Defects can readily be mitigated through compensatory mutations

Recent laboratory studies in bacteria and yeast showed that defects in a broad range of molecular processes can readily be compensated during evolution[5,64]. Notably, deletion of 9% of the essential genes can be overcome by evolution of alternative pathways, suggesting that gene dispensability can readily evolve in the laboratory[65]. Compensatory evolution appears to be common at different levels of biological organization (for references, see [5]).

A case study: compensatory evolution following gene deletion

In our work[5], we addressed one of the most long-standing debates in evolution. Here we focused on a special, largely neglected aspect of this problem and asked whether deleterious gene loss events promote adaptive genetic changes, and what might be the side consequences of such processes[5]. To achieve such an ambitious goal, we integrated approaches of several disciplines, including laboratory experimental evolution and genomic analyses, coupled with bioinformatics and detailed molecular studies[5] (Figure 8).

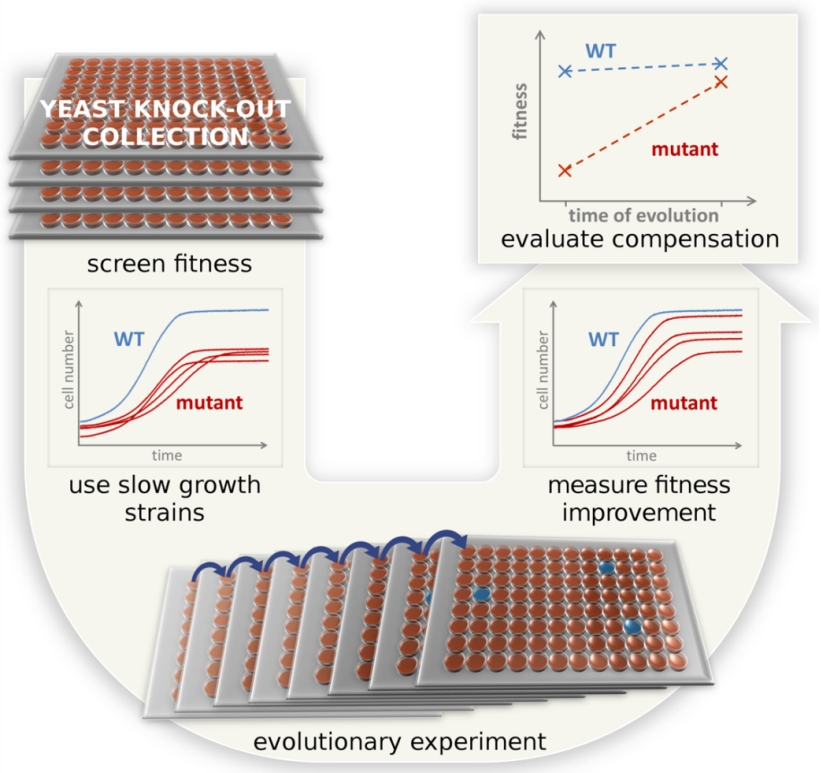


Figure 8. An experimental scheme to study compensatory evolution in strains with single gene defects. Briefly, we started laboratory evolution with over 180 single gene knock-out mutant yeast (*Saccharomyces cerevisiae*) strains, all of which initially showed low fitness compared to the wild-type control in a standard laboratory

medium. Populations were cultivated in parallel, resulting in over 700 independently evolving lineages. To control for potential adaptation unrelated to compensatory evolution, we also established 22 populations starting from the isogenic wild-type (WT) genotype, referred to as evolving wild types. All lineages were subjected to high-throughput fitness measurements by measuring growth capacity in liquid medium.

The analysis reached several important results:

Compensatory evolution following gene loss is pervasive. At least 68% of the deleterious but non-lethal null mutations can be buffered through accumulation of adaptive mutations elsewhere in the genome (Figure 9.).

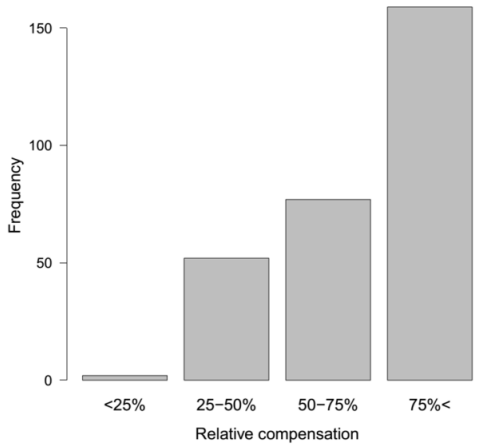


Figure 9. Fraction of initial fitness defects compensated in knock-out mutant yeast strains following evolution in the laboratory.

Full restoration of the lost molecular function is rare. The work revealed that the evolved lines diverge from each other and reach new fitness peaks. The wild-type

physiological state is generally not restored and pleiotropic side effects are prevalent (Figure 10).

Compensatory evolution generates cryptic variation across populations. Accordingly, compensatory evolution generates cryptic differences between diverging lines which can be revealed upon environmental change.

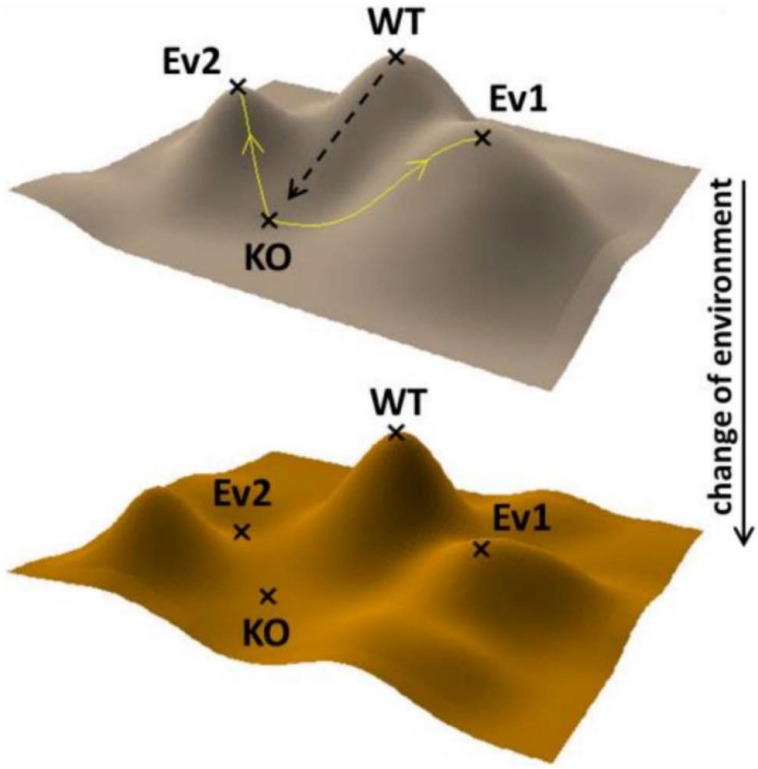


Figure 10. Schematic representation of the impact of compensatory evolution on the fitness landscape. Gene loss leads to a fitness valley (from WT to KO), while compensatory evolution can drive the population to different adaptive peaks (Ev1 versus Ev2). The upper fitness landscape shows the environment where compensatory evolution took place. The dashed arrow represents the original gene deletion event. Yellow lines represent different evolutionary routes. WT, wild type; KO, ancestor strain with a gene deletion

Based on these results, we proposed that a substantial fraction of the gene content variation across species is due to the action of compensatory evolution and may not need to reflect changes in environmental conditions and consequent passive loss of genes.

V. Evolutionary genome engineering

Key publication: Nyerges et al. 2016 (Appendix)

Genome-scale engineering enables editing specific genomic locations in a directed and combinatorial manner[66]. Recent advances in this field offer an unprecedented opportunity to design complex molecular circuits with predefined functions[67]. Most studies have either focused on engineering novel pathways that produce specific molecules for medicine and industry or attempted to construct genomic chassis that are more amenable for further rational design. We recently argued that genome engineering offers extremely powerful discovery tools for understanding the evolution of natural cellular systems[31]. While genome engineering had limited impact on evolutionary research so far, I predict that it will change in the near future: technical advancements in genome engineering have the potential to transform evolutionary biology into a more predictive discipline[31].

Laboratory evolutionary experiment on microbes coupled with whole-genome sequence analysis offer powerful tools to investigate evolution in real time[31]. Current works largely focus on complex phenotypes of whole organisms, where genetic basis is not understood properly. However, this approach has several limitations:

1) **Natural genetic variation is limited in the laboratory**. Several crucial evolutionary innovations lack within population variation, on which selection could act.

2) **Evolution in the laboratory is slow**. Given the limited timescale of microbial laboratory evolution experiments, only relatively few mutations are fixed in most laboratory evolved populations. Therefore, comparison of these results to macroevolutionary trends is often difficult.

3) **No appropriate control of mutational processes**. Studying the evolution of a particular cellular subsystem is hindered by the fact that beneficial mutations can occur outside the subsystem under investigation.

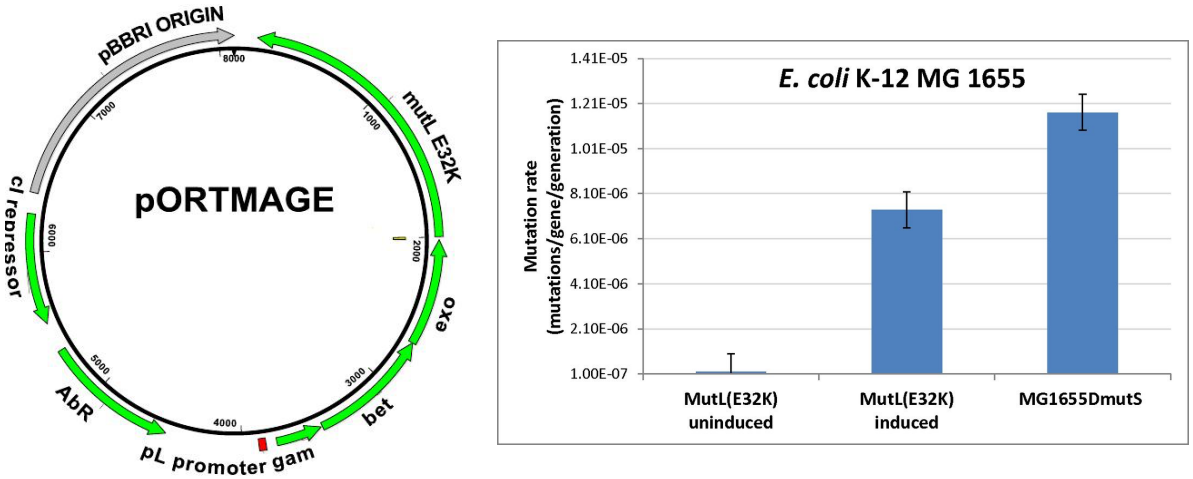
Genome-scale engineering (i.e. the simultaneous modification of multiple genomic loci) provides a novel approach to study evolution in real time, as it can potentially handle the above-mentioned problems[68]. Among others, genome engineering offers a) rapid editing and directed evolution of large genomic segments or entire chromosomes, b) synthesis and combinatorial shuffling of small DNA segments (promoters, coding regions) or complete genomes, c) chemical synthesis and integration of large segments or even whole genomes into new host organism. For details, see ref [31].

Development of a reliable genome engineering protocol for bacteria

Recently, we addressed some of the most long-standing problems in genome engineering[32,33]. Currently available tools for bacterial genome manipulation suffer from three major limitations. They i) have been optimized for a few laboratory model strains (such as *Escherichia coli* MG1655), ii) demand extensive modification of the host genome prior to large-scale genome engineering, and iii) lead to the accumulation of numerous unwanted, off-target modifications, sometimes

outnumbering the desired ones. Clearly, these issues have serious implications on wide-spread biotechnological applicability. Moreover, although CRISPR/Cas9 is applicable to a range of organisms, there seems to be a technical limit when it comes to using CRISPR/Cas9 for simultaneous modification of multiple loci[69,70].

Building on prior development of multiplex automated genome engineering [71], our work addressed these problems and presented a simple, all-in-one solution. Briefly, we first characterized a dominant mutation in a key protein of the methyl-directed mismatch repair (MMR) system and used it to precisely disrupt mismatch-repair in target cells [33] (Figure 11).



A) B)

Figure 11. pORTMAGE. A) General map of the pORTMAGE plasmid. Expression of the mutL E32K gene [along with the three λ Red recombinase enzymes (exo, bet, and gam)] is controlled by the cI857 temperature-sensitive λ repressor. B) Mutation rate measurement of *E. coli* K-12 MG1655 (MG) harboring the AhTC inducible pZA31tetR-mutLE32K plasmid for MutL(E32K) expression, as well as the

MG1655ΔmutS strain for comparison. A rifampicin resistance assay was used to calculate mutation rates. Adapted from Nyerges et al. PNAS 2016.

With the integration of this advance, we developed a new workflow for genome-scale engineering and demonstrated its applicability for high-throughput genome editing by efficient modification of multiple loci (Figure 12).

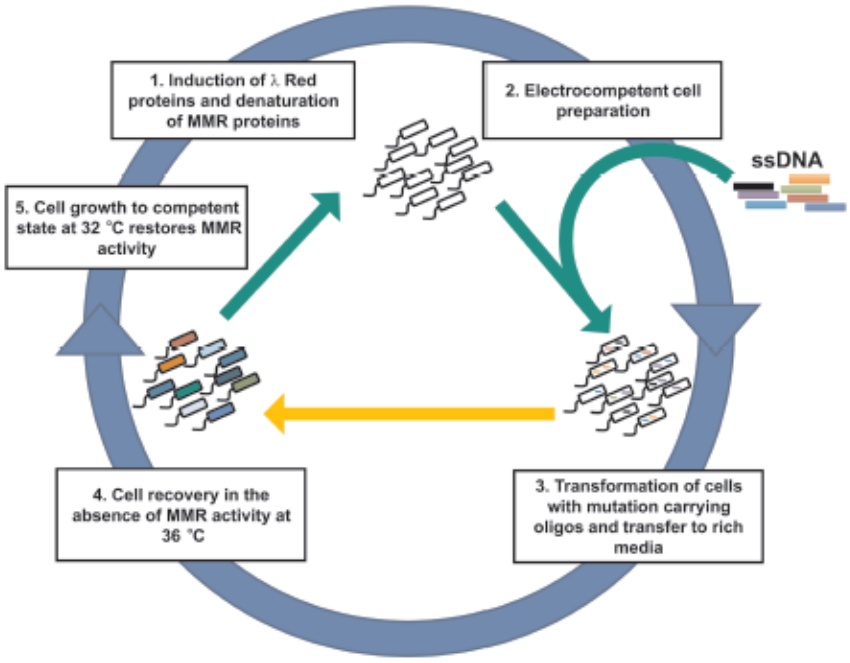


Figure 12. Representation of a modified Multiplex Automated Genome Engineering (MAGE) protocol. Cells are grown and transformed with single stranded oligonucleotides carrying the desired mutations. These oligonucleotides are incorporated into the target genomes in various combinations. Cyclical repetition of MAGE yields highly diverse population of cells. Adapted from Nyerges et al. 2014.

Whole genome sequencing revealed that none of the modified strains carried any observable off-target mutation, a major advance over prior approaches[33]. Due to the highly conserved nature of the bacterial MMR system, the application of dominant mutations in this system provides a unique solution to portability. By placing the entire synthetic operon that enables efficient genome engineering into a broad-host vector, we successfully adapted MAGE to a wide range of hosts and applied the strategy for genome editing in biotechnologically and clinically relevant enterobacteria[33].

To demonstrate the usefulness of our system, we applied pORTMAGE to study a set of antibiotic resistance conferring mutations in *Salmonella enterica* and *E. coli.* Despite over 100 million years of divergence between the two species, mutational effects remained generally conserved, a result with implications for future systematic studies[33] (Table 4).

Gene	Mutation	Antibiotic	*E. coli* relative MIC	*S. enterica* relative MIC
mprA	Arg110Leu	NIT	1.20	1.73
marR	Val84Glu	AMP	2.20	1.15
soxR	Leu139*	ERY	2.31	1.73
phoQ	Gly384Cys	NIT	1.20	0.83
trkH	Thr350Lys	STR	3.18	1.59
gyrA	Ser83Leu	CPR	16.00	16.00
gyrA	Ser83Leu	NAL	97.66	610.35
fis	Thr70Pro	ERY	1.44	1.44
acrR	Gln78*	ERY	1.20	1.00
ompC	Met1	NIT	1.20	1.20
ycbZ	Ser438Arg	ERY	1.23	1.51

Table 4. Minimum inhibitory concentrations (MICs) of *Escherichia coli* and *Salmonella enterica* strains with a single specific mutation. The measured MIC for each strain was compared with the MIC of the wild-type strain, resulting in the

relative MIC value. The antibiotic abbreviations are as follows: AMP, ampicillin; CPR, ciprofloxacin; ERY, erythromycin; NAL, nalidixic acid; NIT, nitrofurantoin; STR, streptomycin. Adapted from Nyerges et al. PNAS 2016.

In sum, with just one transformation, pORTMAGE allows any strain of interest across a range of enterobacteria to become an efficient host for genome-scale editing. pORTMAGE simultaneously eliminates off-target mutagenesis. Within a year after the publication, at least 45 research groups started using pORTMAGE.

Our findings have broad implications with regards to chassis engineering for the production of valuable biomaterials through the rapid optimization of biosynthetic pathways across a wide range of bacteria, a process previously requiring tedious laboratory optimization. Moreover, based on our proof-of-principle experiments, we predict that pORTMAGE will open a new avenue of research in diverse fields, such as functional genomics and evolutionary biology. For the first time, pORTMAGE allows systematic comparison of mutational effects and epistasis across a wide range of bacterial species.

Evolution of genome minimization

Next, we addressed one of the central issues in evolution: why are some bacterial genomes highly reduced[34,72]? According to the prevailing view that has emerged in the past 15 years, massive genome shrinkage in bacteria is driven by non-adaptive processes, such as genetic drift and mutational bias[73,74]. However, the recent discovery of genome reduction in free-living bacteria with immense population sizes challenged this view and led to the alternative hypothesis that simplified genomes are the result of selection for efficient use of nutrients[75]. The issue remains unsettled not least because little is known about how genome reduction alters cellular traits.

For example, it remains poorly understood whether genome reduction results in faster and more efficient cell growth owing to a reduced burden of DNA replication. Recently, we employed genome engineering to construct *Escherichia coli* strains with successively reduced genomes[76] (Figure 13).

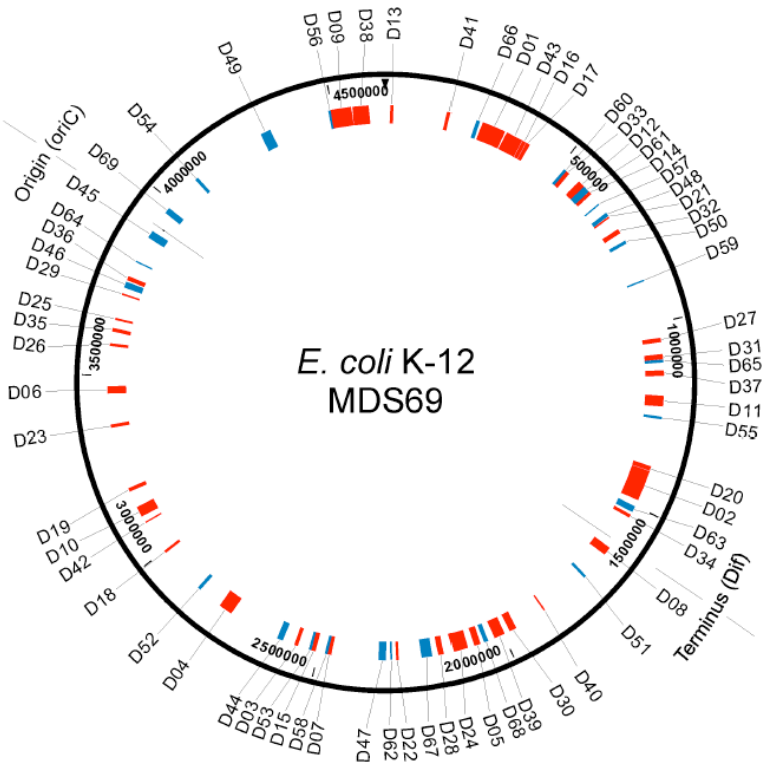


Figure 13. Comparison of the starting *E.coli* K-12 genome and the derived multi-deletion strain 69 (MDS69). Deleted genomic regions are indicated. Adapted from Karcagi et al. 2016.

Our strain collection gives a unique opportunity to investigate the evolutionary consequences of genome reduction, for three reasons: i) the extent of genome reduction was as high as 20%, ii) the resulting 69 strains of the multiple-deletion series represent different stages of genome reduction and iii) the deleted segments harbor genes that have been repeatedly lost and gained in relatives of *E. coli*[76]. Next,

we systematically tested the impacts of genomic reduction on several cellular traits, including growth rate, metabolic yield, nutrient utilization profile, cell size, and transcriptome profile[76]. Prior knowledge of the impact of genome reduction on these traits was very limited.

Our analysis yielded two major insights[76]: First, we found no evidence for increased cellular efficiency as a result of genome reduction. On the contrary, removal of seemingly non-essential genomic segments had widespread and strong pleiotropic effects on cellular physiology. This indicates that the energetic benefit gained by short genomic deletions is vanishingly small compared to the deleterious side effects of these deletions. Thus, bacterial genome reduction is unlikely to be solely driven by natural selection for decreased DNA synthesis costs.

Second, our systematic assays revealed that accessory genomic regions, that preferentially harbor horizontally transferred genes, have important contributions to fitness both in standard laboratory environments (Figure 14) and under stress (Table 5).

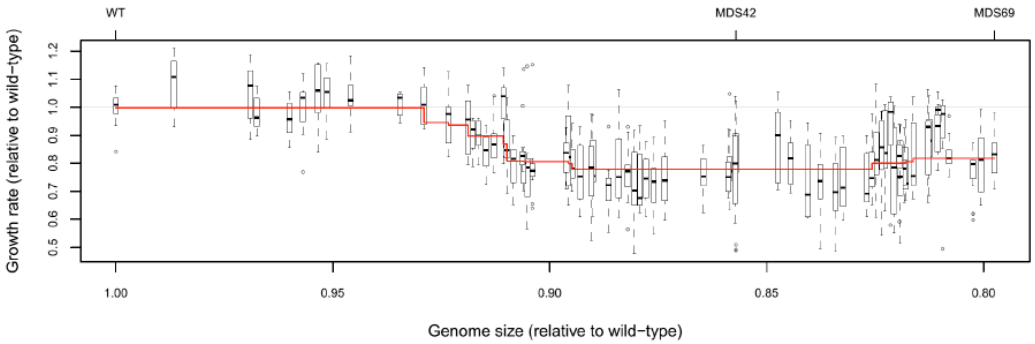


Figure 14. Growth rates of the wild type (*E.coli* K-12) and multi-deletion strains in standard laboratory medium. For details, see Karcagi et al. 2016.

Class of environment	Number of environments		
	wild-type grows	MDS42 defective	MDS69 defective
Carbon source	84	4	8
Nitrogen source	303	18	25
Phosphorous source	59	2	3
Sulfur source	35	3	3
Nutrient supplements	94	0	0
Osmotic / ionic stress	60	5	5
pH stress	47	22	21
Chemical inhibitors	226	26	65
Total	*908*	*80*	*130*

Table 5. Summary of growth profiles of the wild-type and land-mark multi-deletion strains (MDS42 and MDS69) in 908 environments.

These results provide strong support to the notion that accessory genes of the bacterial pangenome are under strong selection, and are not just a collection of transient neutral DNA segments. Accordingly, our work indicates that bacterial genes derived by horizontal transfer are indispensable, and many appear to have important functional roles even in stress-free environments[76]. Finally, we argued that selection for eliminating specific gene functions detrimental in particular environments, and not a reduced genome per se, could be the driving force behind rapid evolution of genome reduction in microbes with large population sizes.

The future of evolutionary genome engineering

Two factors will influence future applications. First, the nascent field of genome engineering is expected to integrate concepts and protocols of other evolutionary disciplines and computational systems biology[1] (Figure 15). Second, novel technologies are expected to transform this discipline[31,77].

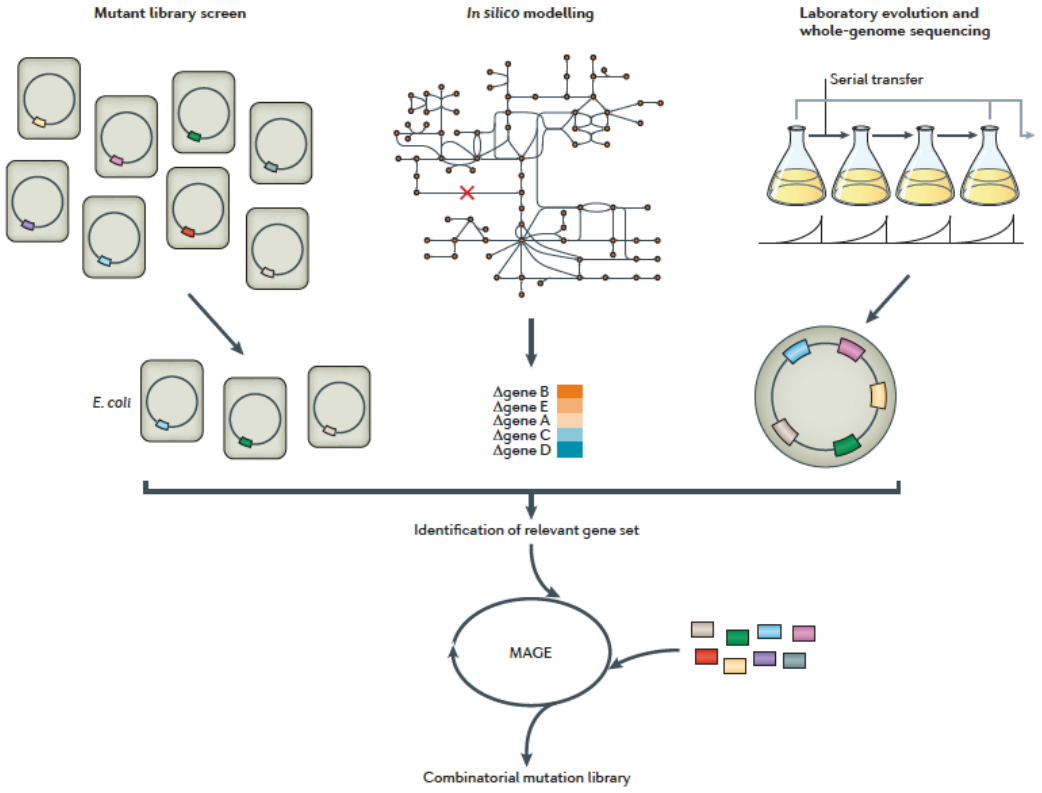


Figure 15. Tools for evolutionary genome engineering. The analysis should integrate screens of genome-scale mutant libraries, computational modeling of cellular networks (such as flux balance analysis), and laboratory evolution. These methods enable researchers to identify gene sets relevant to the phenotypic trait investigated (such as production of a biomaterial). As a next step, directed evolution should focus on mutagenesis – selection on the identified loci.

We expect major breakthroughs in the following areas:

Reconstruction of ancestral networks, subsystems or genomes

Ancestral protein sequences can be inferred using phylogenetic methods. Reconstruction of these ancestral sequences through gene synthesis and integration into native genomes allows functional characterization[78,79]. Successful examples so far include enzymes, highly conserved regulatory proteins or protein complexes[80]. Among others, these studies delivered insight into ecological niches of ancestral species and mechanisms underlying evolutionary innovations through gene duplication[81]. The next step will be to use multiplex automated genome engineering and related protocols to reconstruct larger subsystems or even the complete genomes of ancestral species[31].

Laboratory evolution of complex adaptations

The forces by which complex cellular features – such as linear metabolic pathways or multimeric protein complexes – emerge is one of the major problems of evolutionary cell biology[82,83]. Many of such complex adaptations require simultaneous acquisition of multiple, very specific and rare mutations in a single lineage. Thus, the time for establishment of such adaptations is expected to be very slow in nature. The process is also highly dependent on the frequency of appropriate mutations or horizontal transfer events. As multiplex automated genetic engineering can generate over 4.3 billion combinatorial genomic variants per day at selected loci, it can potentially accelerate the laboratory evolution of complex adaptations[31,67,71].

VI. Antibiotic resistance and collateral sensitivity in bacteria

Representative publications: Lazar et al MSB 2013, Lazar et al. Nature Communications 2014 (Appendix)

Understanding how evolution of microbial resistance towards a given antibiotic enhance (cross-resistance) or decrease (collateral sensitivity) fitness in the presence of other drugs is a challenge of profound importance for several fields of basic and applied research[84]. Despite its obvious clinical importance, our knowledge is still limited, not least because this problem has been addressed largely by small-scale clinical studies. By combining laboratory evolution (Figure 16A), genome sequencing, and functional analyses (Figure 16B), recent works charted the maps of cross-resistance/collateral sensitivity interactions between antibiotics in *E. coli*[85-87], and explored the mechanisms driving these evolutionary patterns[86].

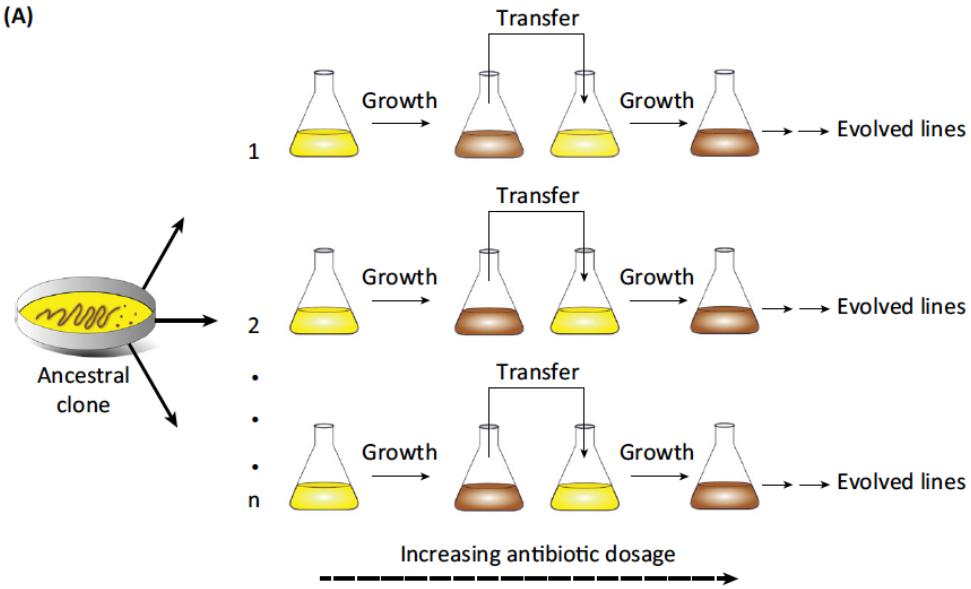


Figure 16A. In prior works[85,86], we initiated laboratory evolutionary experiments starting with a single clone of *E. coli* K12. Parallel evolving bacterial populations were exposed to gradually increasing concentrations of one of 12 clinically relevant antibiotics, leading to up to 328-fold increase in the minimum inhibitory concentrations (MICs) relative to the wild-type. In all cases, the resistance levels were equal to or above the EUCAST clinical break-points. 52% of the evolved strains showed resistance to multiple antibiotics. Adapted from Pal et al. 2015 [29].

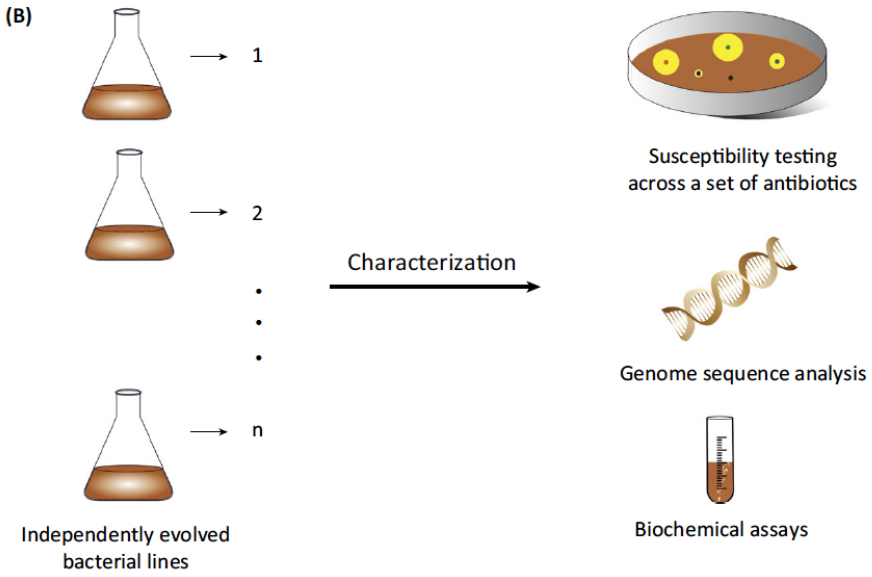


Figure 16B. The laboratory evolved lineages were subjected to in-depth phenotypic and genomic analysis with the aim to explore the accompanying changes in drug sensitivity and the underlying molecular mechanisms thereof. Adapted from Pal et al. 2015[29].

The exceptionally large scale of these works allowed to derive several conceptually novel conclusions[29]. First, antibiotic cross-resistance is frequent and computationally predictable by integrating the accumulated knowledge on functional and chemical antibiotic properties[85]. Second, mutations that cause multi-drug resistance simultaneously enhance sensitivity to many other drugs[86,87]. Third, these works offered an insight into the mechanisms underlying collateral sensitivity[86]. In this short chapter, we summarize recent advances in this emerging research area. We highlight the potential and limitations of current approaches, review the underlying molecular mechanisms of these phenomena, and suggest new research directions

for future studies. Specifically, we discuss how these advances could be exploited for the development of novel antimicrobial strategies.

Multi-drug resistance emerges in response to evolution against a single drug

To chart the map of cross-resistance, recent works initiated parallel laboratory evolutionary experiments to adapt to increasing dosages of one of 12 clinically relevant antibiotics (Table 6).

Antibiotic name	Abbreviation	Mode of Action	Bactericidal or Bacteriostatic
Ampicillin	AMP	Cell wall	Bactericidal
Cefoxitin	FOX	Cell wall	Bactericidal
Ciprofloxacin	CPR	Gyrase	Bactericidal
Nalidixic Acid	NAL	Gyrase	Bactericidal
Nitrofurantoin	NIT	Multiple mechanisms	Bactericidal
Kanamycin	KAN	Protein synthesis, 30S, Aminoglycosides	Bactericidal
Tobramycin	TOB	Protein synthesis, 30S, Aminoglycosides	Bactericidal
Tetracycline	TET	Protein synthesis, 30S	Bacteriostatic
Doxycycline	DOX	Protein synthesis, 30S	Bacteriostatic
Chloramphenicol	CHL	Protein synthesis, 50S	Bacteriostatic
Erythromycin	ERY	Protein synthesis, 50S	Bacteriostatic
Trimethoprim	TRM	Folic acid biosynthesis	Bacteriostatic

Table 6. Antibiotics employed in the study by Lazar and colleagues and the corresponding modes of action.

Evolved populations reached up to 300-fold increas in the minimum inhibitory concentrations relative to the ancestor[85-87]. As a next step, the corresponding changes in susceptibilities of the lab-evolved populations were measured against a panel of other antibiotics, allowing us to infer a network of cross-resistance interactions (Figure 17). Laboratory-evolved lines were subjected to whole-genome sequence analysis and biochemical assays to decipher the underlying molecular mechanisms of these interactions. These studies revealed that:

a) The cross-resistance network is dense, indicating that exposure to a single

antibiotic frequently yields multidrug resistance.

b) The populations frequently evolve asymmetric cross protection, where stress A protects against stress B but not vice versa.

c) The network of cross-resistance is predictable based on antibiotic properties.

d) Finally, laboratory studies recapitulated major patterns of antibiotic cross-resistance observed in the clinics.

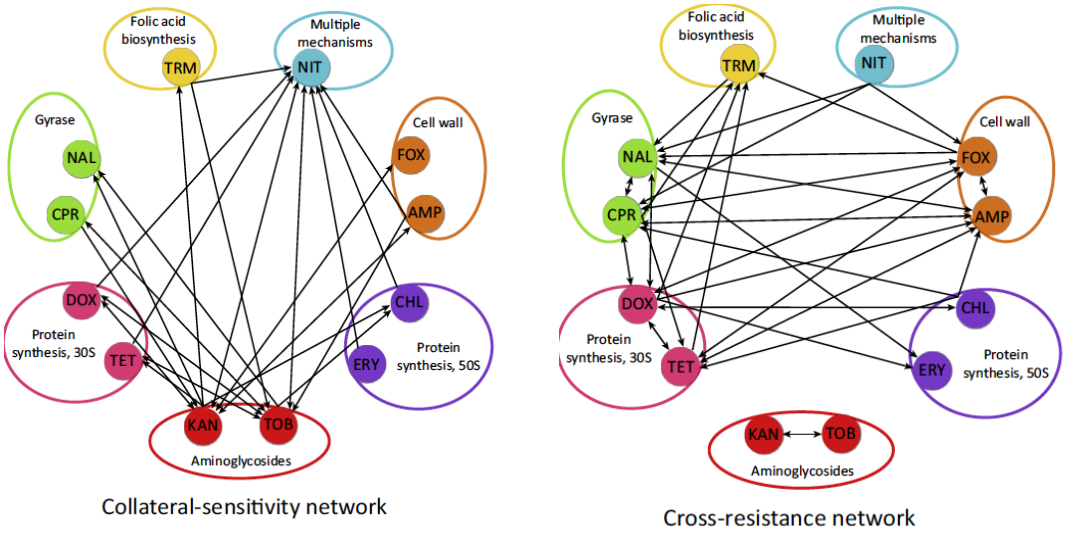


Figure 17. Based on the high-throughput measurement of antibiotic susceptibilities in laboratoryevolved bacteria, two networks can be deciphered. An arrow from antibiotic A to B indicates that evolution of resistance to A generally increases (collateral sensitivity) or decreases (cross-resistance) susceptibility to B. Adapted from Pal et al. 2015.

These works also identified a strong signature of parallel evolution at the molecular level that emerged across populations adapted to different antibiotics, and such parallel mutations delivered resistance to multiple antimicrobial agents[29,85]. The

molecular mechanisms underlying antibiotic cross-resistance appeared to be very diverse, including mutations in multi drug efflux pumps, metabolic genes, and genes involved in bacterial defense against c) oxidative, d) nutritional and e) membrane stresses. These works also suggested that genome-wide transcriptional rewiring mediated by global transcriptional regulatory genes has an important contribution to cross-resistance patterns.

Perhaps the most remarkable aspect of these findings is that cross-resistance is delivered by mutations with wide pleiotropic effects[85,86]. Therefore, cross-protection may be more general[88], and opens the possibility that stressful conditions unrelated to antibiotic pressure may, as a byproduct, select for enhanced antibiotic tolerance in nature.

Evolution of multi-drug resistance promotes hypersensitivity to certain drugs

The phenomenon

Prior studies demonstrated that evolution of resistance to a single antibiotic is frequently accompanied by increased resistance to multiple other antimicrobial agent[29]. However, very little is known about the occurrence of collateral sensitivity (i.e. when evolution of resistance yields enhanced sensitivity to other antibiotics). Given the prevalence of resistance conferring mutations with pleiotropic effects, researchers speculated that such collateral sensitivity interactions could frequently emerge. Large-scale laboratory evolution studies demonstrated that it was indeed so. Strikingly, not only cross-resistance, but also collateral sensitivity interactions frequently occur during evolution of antibiotic resistance (Figures 17 and 18).

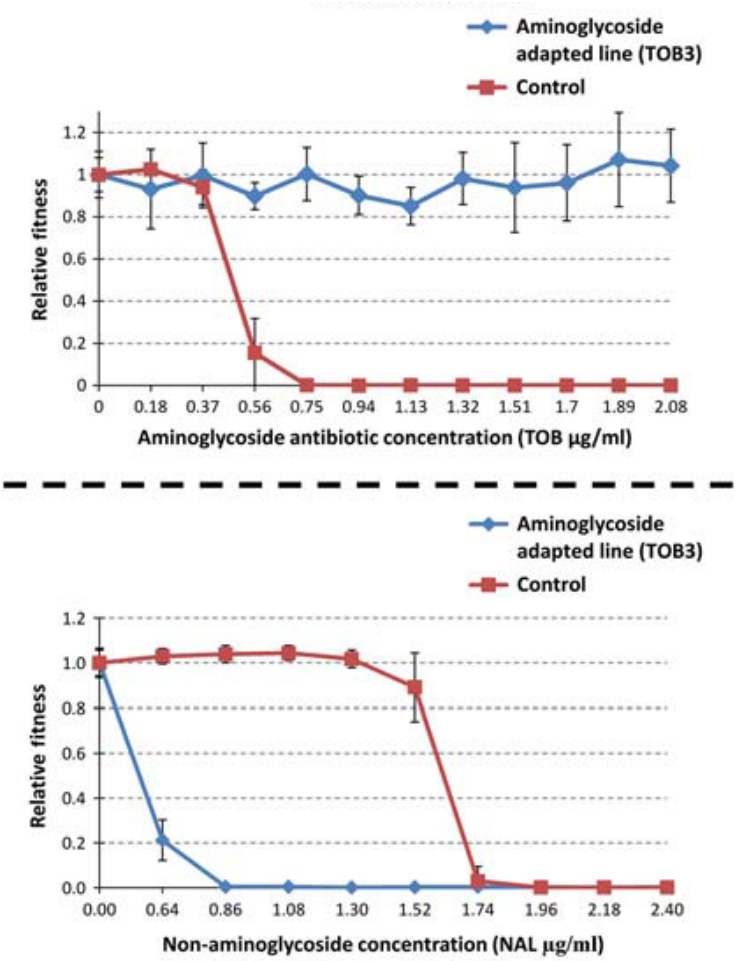


Figure 18.. An example of collateral sensitivity. Dose response curve of the wild-type control and a tobramycin (aminoglycoside) resistant bacterial strain (TOB3). TOB3 shows resistance to tobramycin, but surprisingly, it has elevated susceptibility to a drug with unrelated mode of action (gyrase inhibitor, nalidixic acid). Adapted from Lazar et al. 2013.

The mechanisms

Understanding the mechanisms underlying collateral sensitivity interactions is still at an embryonic stage. We mention one example here: resistance mechanisms to various antibiotics via alteration of membrane potential have been reported in both laboratory studies and clinical settings, and such changes underlie the

hypersensitivity of bacteria to other antibiotics[86]. These results indicate the existence of antagonistic mechanisms by which bacteria modulate intracellular antibiotic concentration through altering membrane polarity[86] (Figure 19).

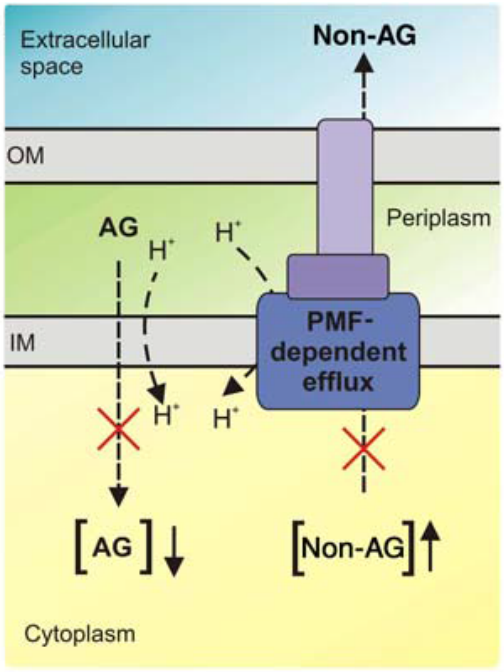


Figure 19. A mechanism underlying collateral sensitivity. Altering the membrane potential across the inner bacterial membrane has two opposing effects: it reduces the uptake of many aminoglycoside-related antibiotics but simultaneously leads to the reduced activity of PMF-dependent efflux pumps. Adapted from Lazar et al. 2013.

Development of novel multi-drug therapies

The experimental map of cross-resistance/collateral sensitivity could serve as a unique resource and potentially permit informed decisions in medicin[29]. For example, the choice of optimal antibiotic combinations depends on both the presence of physiological drug interactions and the availability of mutations that deliver resistance

to both drugs simultaneously. It has been shown that cross-resistance between two antibiotics is largely independent of whether they show synergistic effects in combination[89]. Combination of large-scale information on antibiotic synergism and cross-resistance could be especially informative for future development of multidrug therapies. For example, it remains controversial whether temporal rotation of antibiotics could select against the development of resistance[30]. These works strongly indicate that the success of such a strategy depends on the choice of antibiotics: treatment with a single antibiotic and then switching to a cross-sensitive partner may be a viable strategy. An alternative approach relies on the simultaneous administration of two agents in collateral sensitivity interaction to inhibit both the wild-type and the resistant subpopulations, and thereby prevent the emergence of resistance[30,89] (Figure 20).

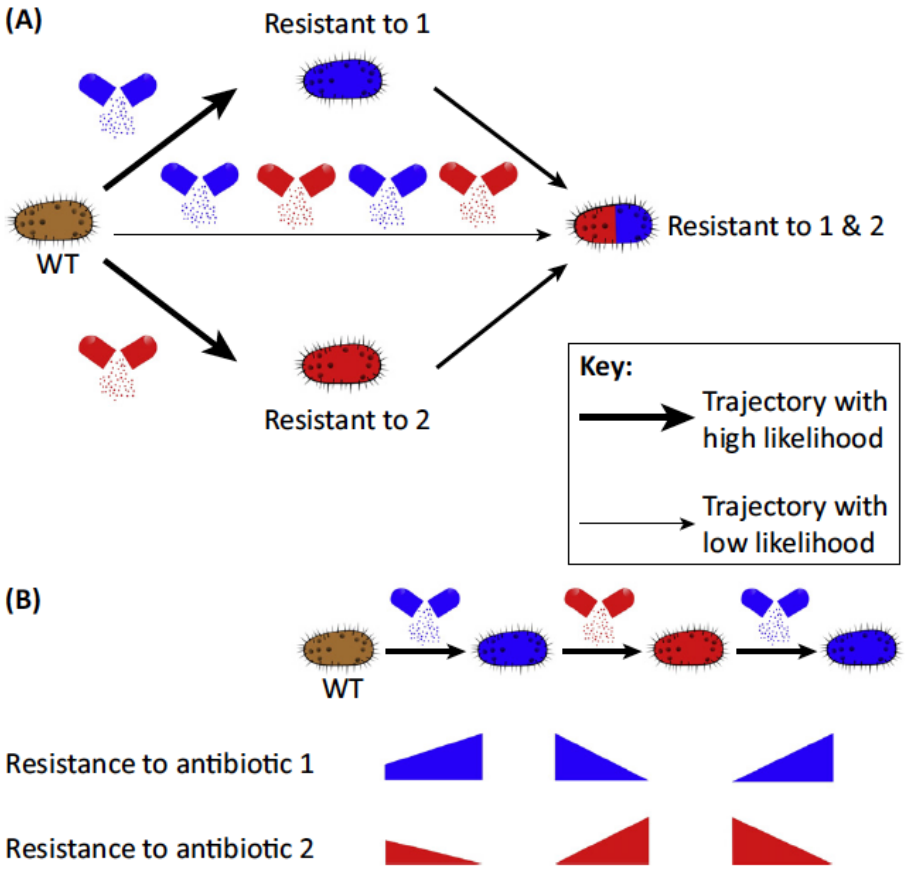


Figure 20. Potential applications of collateral sensitivity to eradicate antibiotic resistant bacteria. Antibiotic pairs showing collateral sensitivity could be administered simultaneously as drug combination (a) or in an alternating fashion (b). Abbreviation: WT, wild type. See Pal et al. 2015 for more details.

Testing the long-term efficacy of novel therapeutic agents

By analyzing the maps of cross-resistance, researchers unveiled some general principles governing the evolution of cross-resistance patterns. By integrating available data on antibiotic properties, it has been shown that cross-resistance is partly predictable[85]. These results pave the way towards in silico methods to estimate the cross-resistance propensity of novel antimicrobial compounds before entry into clinical usage. At least five key issues need to be investigated in more depth by future studies:

1) Evolutionary conservation of cross-resistance maps and the underlying molecular mechanisms across bacterial species.

2) Exploiting the fitness costs of plasmid mediated antibiotic resistance mechanisms.

3) Confirmation of laboratory results with in vivo and clinical studies. Indeed, comparison of existing large-scale clinical data on multidrug resistance with results of laboratory evolution studies has a central importance.

4) Integrating information from metagenomic approaches that aim to identify resistance genes from environmental reservoirs.

5) Establishing how the evolvability of further resistance is influenced by cross-resistance and collateral sensitivity interactions.

VII. The future of evolutionary systems biology

The emerging field of evolutionary systems biology investigates central issues in evolutionary biology by focusing on specific cellular subsystems and integrating a variety of methodologies[1]. The goal of computational analyses is at least threefold. First, they calculate the impact of genetic mutations on cellular phenotypes that are difficult to estimate experimentally on a large-scale or across environments. Second, they provide insights into complex evolutionary problems such as the causes of gene dispensability or the evolution of minimized genomes. Third, these approaches will transform evolutionary biology into a more predictive discipline.

Such advances are important for the following reasons. For the first time, it is becoming possible to investigate the evolution of metabolic networks and other cellular subsystems in exceptional detail across related microbial species. Researchers now can ask how robustness to mutations and other emergent properties rely on changes in genome architecture and ecology. It also paves the way for network archaeology: that is, the reconstruction and analysis of the functional properties of ancestral cellular networks[1].

More practically, systems biology could promote the identification of new drug targets shared by pathogens. Indeed, there is an urgent need for new experimental technologies to investigate mutational effects and evolution in a high-throughput manner. Given the limited timescale of microbial laboratory evolution experiments, most existing protocols are inadequate to study long-term evolution of a given cellular subsystem[1]. Novel genome-engineering techniques provide a solution, as multiplex automated genome engineering generates huge genetic diversity in very specific manner[1].

Accordingly, systems biology can greatly benefit from concepts and methods of genome engineering[31]. By constructing rare genomic alterations or specific combinations of mutations, genome engineering could facilitate complex changes of cellular subsystems. Combination of rational and evolutionary design strategies is important both for understanding natural systems and for the construction of genetic regulatory circuits for biotechnological purposes.

These considerations have important medical implications, including the problem of antibiotic resistance. Although there has been much progress in our understanding of collateral sensitivity, there are several key questions that remain unanswered[29]. It will be crucial to decipher the long-term impact of collateral sensitivity on resistance evolution. The associated costs that render microbes vulnerable to certain antibiotic may only be temporary, and that compensatory evolution can rapidly restore fitness[29]. Future works should elucidate to what extent, and how, mutations ameliorating the fitness cost of resistance under drug-free conditions re-wire the collateral-sensitivity interactions between antibiotics. Alternatively, collateral sensitivity may have a long-lasting effect with a substantial impact on reaching clinically significant resistance levels[29].

I anticipate that these novel experimental techniques, along with computational models of specific cellular subsystems, will allow researchers to reinvestigate key issues in the fields of network evolution and antibiotic resistance.

Acknowledgments

I am very lucky to work with wonderful people around me. With them, science is fun. I'm very grateful for all what I have learned from Balázs Papp, Laurence Hurst,

György Posfai and members of the Pál and Papp labs. I'm particularly grateful for the Hungarian Academy of Sciences for all the supports I have received. Finally, I would not have reached my goals without BBB.

References

1. Papp, B., Notebaart, R. A. & Pál, C. Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet* **12,** 591-602 (2011).
2. Zhang, J. & Yang, J. -R. Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics* (2015).
3. Pál, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nat Rev Genet* **7,** 337-348 (2006).
4. Wright, S. Surfaces of selective value revisited. *The American Naturalist* **131,** 115-123 (1988).
5. Szamecz, B., Boross, G., Kalapis, D., Kovács, K., *et al.* The genomic landscape of compensatory evolution. *PLoS biology* **12,** e1001935 (2014).
6. Drummond, D. A. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* **10,** 715-724 (2009).
7. Kalapis, D., Bezerra, A. R., Farkas, Z., Horvath, P., *et al.* Evolution of Robustness to Protein Mistranslation by Accelerated Protein Turnover. *PLoS Biol* **13,** e1002291 (2015).
8. Copley, S. D. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci* **40,** 72-78 (2015).
9. Notebaart, R. A., Szappanos, B., Kintses, B., Pál, F., *et al.* Network-level architecture and the evolutionary potential of underground metabolism. *Proc Natl Acad Sci U S A* **111,** 11762-11767 (2014).
10. Soyer, O. S. & O'Malley, M. A. Evolutionary systems biology: What it is and why it matters. *BioEssays* n/a (2013).
11C Pál, B Papp, LD Hurst, Highly expressed genes in yeast evolve slowly., in *Genetics* [UNKNOWN REFERENCE TYPE]
12. Pál, C., Papp, B. & Hurst, L. D. Genomic function: Rate of evolution and gene dispensability. *Nature* **421,** 496-7; discussion 497-8 (2003).
13. Koonin, E. V. Are there laws of genome evolution? *PLoS Comput Biol* **7,** e1002173 (2011).
14. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424,** 194-197 (2003).
15. Birchler, J. A. & Veitia, R. A. Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A* (2012).
16. Makino, T. & McLysaght, A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proceedings of the National Academy of Sciences* **107,** 9270-9274 (2010).
17. Pal, C., Maciá, M. D., Oliver, A., Schachar, I. & Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450,** 1079-1081 (2007).
18. Papp, B., Pál, C. & Hurst, L. D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429,** 661-664 (2004).
19. Pál, C., Papp, B. & Lercher, M. J. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37,** 1372-1375 (2005).
20. Pál, C., Papp, B., Lercher, M. J., Csermely, P., *et al.* Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440,** 667-670 (2006).
21. Harrison, R., Papp, B., Pál, C., Oliver, S. G. & Delneri, D. Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci U S A* **104,** 2307-2312 (2007).
22. Szappanos, B., Kovács, K., Szamecz, B., Honti, F., *et al.* An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet* **43,** 656-662 (2011).
23. Szappanos, B., Fritzemeier, J., Csörgő, B., Lázár, V., *et al.* Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat Commun* **7,** 11607 (2016).

24. Barve, A. & Wagner, A. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* **500,** 203-206 (2013).
25. Lázár, V., Pal Singh, G., Spohn, R., Nagy, I., *et al.* Bacterial evolution of antibiotic hypersensitivity. *Mol Syst Biol* **9,** 700 (2013).
26. Ocampo, P. S., Lázár, V., Papp, B., Arnoldini, M., *et al.* Antagonism between bacteriostatic and bactericidal antibiotics is prevalent. *Antimicrob Agents Chemother* **58,** 4573-4582 (2014).
27. Méhi, O., Bogos, B., Csörgő, B., Pál, F., *et al.* Perturbation of iron homeostasis promotes the evolution of antibiotic resistance. *Mol Biol Evol* **31,** 2793-2804 (2014).
28. Lázár, V., Nagy, I., Spohn, R., Csörgő, B., *et al.* Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nat Commun* **5,** 4352 (2014).
29. Pál, C., Papp, B. & Lázár, V. Collateral sensitivity of antibiotic-resistant microbes. *Trends Microbiol* **23,** 401-407 (2015).
30. Baym, M., Stone, . K. & Kishony, R. Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* **351,** aad3292 (2016).
31. Pál, C., Papp, B. & Pósfai, G. The dawn of evolutionary genome engineering. *Nat Rev Genet* **15,** 504-512 (2014).
32. Nyerges, Csörg\Ho, B., Nagy, I., Latinovics, D., *et al.* Conditional DNA repair mutants enable highly precise genome engineering. *Nucleic acids research* gku105 (2014).
33. Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., *et al.* A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc Natl Acad Sci U S A* **113,** 2502-2507 (2016).
34. Fehér, T., Papp, B., Pal, C. & Pósfai, G. Systematic genome reductions: theoretical and experimental approaches. *Chem Rev* **107,** 3498-3513 (2007).
35. Hurst, L. D. & Smith, N. G. Do essential genes evolve slowly? *Curr Biol* **9,** 747-750 (1999).
36. Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320,** 362-365 (2008).
37. Blank, L. M., Kuepfer, L. & Sauer, U. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* **6,** R49 (2005).
38. Kuepfer, L., Sauer, U. & Blank, L. M. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res* **15,** 1421-1430 (2005).
39. de Visser, J. A., Hermisson, J., Wagner, G. P., Ancel Meyers, L., *et al.* Perspective: Evolution and detection of genetic robustness. *Evolution* **57,** 1959-1972 (2003).
40. He, X. & Zhang, J. Higher duplicability of less important genes in yeast genomes. *Mol Biol Evol* **23,** 144-151 (2006).
41. Pfeiffer, T., Soyer, O. S. & Bonhoeffer, S. The Evolution of Connectivity in Metabolic Networks. *PLOS Biology* **3,** e228 (2005).
42. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research* **33,** D514-D517 (2005).
43. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky--Muller incompatibilities in protein evolution. *Proceedings of the National Academy of Sciences* **99,** 14878-14883 (2002).
44. Gao, L. & Zhang, J. Why are some human disease-associated mutations fixed in mice? *Trends in Genetics* **19,** 678-681 (2003).
45. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nature Reviews Genetics* **16,** 333-343 (2015).
46. MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335,** 823-828 (2012).
47. Doniger, S. W., Kim, H. S., Swain, D., Corcuera, D., *et al.* A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet* **4,** e1000183 (2008).
48. Nadeau, J. H. Modifier genes in mice and humans. *Nature Reviews Genetics* **2,** 165-174 (2001).
49. Chen, R., Shi, L., Hakenberg, J., Naughton, B., *et al.* Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nature biotechnology* (2016).

50. Dowell, R. D., Ryan, O., Jansen, A., Cheung, D., *et al.* Genotype to phenotype: a complex problem. *Science* **328,** 469 (2010).
51. Vu, V., Verster, A. J., Schertzberg, M., Chuluunbaatar, T., *et al.* Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. *Cell* **162,** 391-402 (2015).
52. Wang, J et al. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost-Benefit Tradeoff. PLoS Biol **13**:1, e100204 (2015)
53. Blomen VA et al. Gene essentiality and synthetic lethality in haploid human cells Science **350**:6264, 1092-1096 (2015)
54. Boone, C. and Andrews, BJ. The indispensable genome Science **350**:6264, 1028-1029 (2015)
55. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends in Genetics* (2011).
56. Bridgham, J. T., Ortlund, E. A. & Thornton, J. W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461,** 515-519 (2009).
57. Harms, M. J. & Thornton, J. W. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* **512,** 203-207 (2014).
58. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328,** 1272-1275 (2010).
59. Zhang, R., Ou, H. Y. & Zhang, C. T. DEG: a database of essential genes. *Nucleic Acids Res* **32,** D271-D272 (2004).
60. Delneri, D., Hoyle, D. C., Gkargkas, K., Cross, E. J., *et al.* Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat Genet* **40,** 113-117 (2008).
61. MacLean, R. C., Bell, G. & Rainey, P. B. The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. *Proc Natl Acad Sci U S A* **101,** 8072-8077 (2004).
62. Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep* **2,** 1399-1410 (2012).
63. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat Rev Genet* **17,** 379-391 (2016).
64. Moore, F. B. G., Rozen, D. E. & Lenski, R. E. Pervasive compensatory adaptation in Escherichia coli. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267,** 515-522 (2000).
65. Liu, G., Yong, M. Y., Yurieva, M., Srinivasan, K. G., *et al.* Gene Essentiality Is a Quantitative Property Linked to Cellular Evolvability. *Cell* (2015).
66. Esvelt, K. M. & Wang, H. H. Genome-scale engineering for systems and synthetic biology. *Molecular Systems Biology* **9,** (2013).
67. Woodruff, L. B. & Gill, R. T. Engineering genomes in multiplex. *Curr Opin Biotechnol* **22,** 576-583 (2011).
68. Carr, P. A. & Church, G. M. Genome engineering. *Nature biotechnology* **27,** 1151-1162 (2009).
69. DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* **41,** 4336-4343 (2013).
70. Mans, R., van Rossum, H. M., Wijsman, M., Backx, A., *et al.* CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *FEMS Yeast Res* **15,** (2015).
71. Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894-898 (2009).
72. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35,** 829-837 (2013).
73. Nikolaev, E. V., Burgard, A. P. & Maranas, C. D. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. *Biophys J* **88,** 37-49 (2005).
74. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10,** 13-26 (2012).
75. Giovannoni, S. J., Thrash, J. C. & Temperton, B. Implications of streamlining theory for microbial ecology. *The ISME journal* **8,** 1553-1565 (2014).
76. Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., *et al.* Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol Biol Evol* **33,** 1257-1269 (2016).
77. Biot-Pelletier, D. & Martin, V. J. Evolutionary engineering by genome shuffling. *Appl Microbiol Biotechnol* (2014).

78. Gaucher, E. A., Thomson, J. M., Burgan, M. F. & Benner, S. A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425,** 285-288 (2003).
79. Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451,** 704-707 (2008).
80. Benner, S. A., Sassi, S. O. & Gaucher, E. A. Molecular paleoscience: systems biology from the past. *Adv Enzymol Relat Areas Mol Biol* **75,** 1-132, xi (2007).
81. Voordeckers, K., Brown, C. A., Vanneste, K., van der Zande, E., *et al.* Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *PLOS Biology* **10,** e1001446 (2012).
82. Lynch, M. & Abegg, A. The rate of establishment of complex adaptations. *Mol Biol Evol* **27,** 1404-1414 (2010).
83. Lynch, M. Scaling expectations for the time to establishment of complex adaptations. *Proc Natl Acad Sci U S A* **107,** 16577-16582 (2010).
84. SZYBALSKI, W. & BRYSON, V. Genetic studies on microbial cross resistance to toxic agents. I. Cross resistance of Escherichia coli to fifteen antibiotics. *J Bacteriol* **64,** 489-499 (1952).
85. Lázár, V., Nagy, I., Spohn, R., Csörg\Ho, B., *et al.* Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nature communications* **5,** (2014).
86. Lázár, V., Pal Singh, G., Spohn, R., Nagy, I., *et al.* Bacterial evolution of antibiotic hypersensitivity. *Molecular systems biology* **9,** (2013).
87. Imamovic, L. & Sommer, M. O. A. Use of Collateral Sensitivity Networks to Design Drug Cycling Protocols That Avoid Resistance Development. *Science Translational Medicine* **5,** 204ra132 (2013).
88. Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J. & Tagkopoulos, I. Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in Escherichia coli. *Molecular Systems Biology* **9,** 643 (2013).
89. Munck, C., Gumpert, H. K., Wallin, A. I. N., Wang, H. H. & Sommer, M. O. Prediction of resistance development against drug combinations by collateral responses to component drugs. *Science translational medicine* **6,** 262ra156-262ra156 (2014).

Appendix

1) Papp, B., Pal, C., Hurst, L.D. (2004)
Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast.
Nature 429: 661-4.

2) Pal, C., Papp, B., Lercher M.J. (2005)
Adaptive evolution of bacterial metabolic networks by horizontal gene transfer.
Nature Genetics 37: 1372-5.

3) Szamecz, B, Boross, G, Kalapis, D, Kovács, K, Fekete, G, Farkas, Z, Lázár, V, Hrtyan, M, Kemmeren, P, Groot Koerkamp MJA, Rutkai, E, Holstege, FCP, Papp, B, Pál, C (2014)
The genomic landscape of compensatory evolution.
PLoS Biol 12(8): e1001935

4) Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffer, K., Bogos, B., Pósfai, G., Pál, C. (2016)
A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species.
Proc Natl Acad Sci U S A. 2016 Feb 16. pii 201520040.

5) Lázár, V, Singh, G P, Spohn, R, Nagy, I, Horváth, B, Hrtyan, M, Busa-Fekete, R, Bogos, B, Méhi, O, Csörgő, B, Pósfai, G, Fekete, G, Szappanos, B, Kégl, B, Papp, B, Pál, C (2013)
Bacterial evolution of antibiotic hypersensitivity
Molecular Systems Biology 9:700

6) Lázár, V, Nagy, I, Spohn, R, Csörgő, B, Györkei, A, Nyerges, Á, Horváth, B, Vörös, A, Busa-Fekete, R, Hrtyan, M, Bogos, B, Méhi, O, Fekete, G, Szappanos, B, Kégl, B, Papp, B, Pál, C (2014)
Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network.
Nature Communications 5: 4352

14. Jones, G. & Rayner, J. M. V. Flight performance, foraging tactics and echolocation in free-living Daubenton's bats *Myotis daubentoni* (Chiroptera: Vespertilionidae). *J. Zool.* **215,** 113–132 (1988).
15. Krull, D., Schumm, A., Metzner, W. & Neuweiler, G. Foraging areas and foraging behavior in the notch-eared bat, *Myotis emarginatus* (Vespertilionidae). *Behav. Ecol. Sociobiol.* **28,** 247–253 (1991).
16. Schumm, A., Krull, D. & Neuweiler, G. Echolocation in the notch-eared bat, *Myotis emarginatus*. *Behav. Ecol. Sociobiol.* **28,** 255–261 (1991).
17. Britton, A. R. C., Jones, G., Rayner, J. M. V., Boonman, A. M. & Verboom, B. Flight performance, echolocation and foraging behaviour in pond bats, *Myotis dasycneme* (Chiroptera: Vespertilionidae). *J. Zool.* **241,** 503–522 (1997).
18. Siemers, B. M. & Schnitzler, H.-U. Natterer's bat (*Myotis nattereri* Kuhl, 1818) hawks for prey close to vegetation using echolocation signals of very broad bandwidth. *Behav. Ecol. Sociobiol.* **47,** 400–412 (2000).
19. Swift, S. M. & Racey, P. A. Gleaning as a foraging strategy in Natterer's bat *Myotis nattereri*. *Behav. Ecol. Sociobiol.* **52,** 408–416 (2002).
20. Fenton, M. B. & Bogdanowicz, W. Relationships between external morphology and foraging behaviour: bats in the genus *Myotis*. *Can. J. Zool.* **80,** 1004–1013 (2002).
21. Arlettaz, R., Jones, G. & Racey, P. A. Effect of acoustic clutter on prey detection by bats. *Nature* **414,** 742–745 (2001).
22. Siemers, B. M., Stilz, P. & Schnitzler, H.-U. The acoustic advantage of hunting at low heights above water: behavioural experiments on the European 'trawling' bats *Myotis capaccinii*, *M. dasycneme* and *M. daubentonii*. *J. Exp. Biol.* **204,** 3843–3854 (2001).
23. Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* 138–162 (Oxford Univ. Press, Oxford, 1991).
24. Ruedi, M. & Mayer, F. Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic ecomorphological convergences. *Mol. Phyl. Evol.* **21,** 436–448 (2001).
25. Simmons, J. A. & Stein, R. A. Acoustic imaging in bat sonar: Echolocation signals and the evolution of echolocation. *J. Comp. Physiol. A* **135,** 61–84 (1980).
26. von Helversen, O. & von Helversen, D. Acoustic guide in bat pollinated flower. *Nature* **398,** 759–760 (1999).
27. Schmidt, S., Hanke, S. & Pillat, J. The role of echolocation in the hunting of terrestrial prey—new evidence for an underestimated strategy in the gleaning bat, *Megaderma lyra*. *J. Comp. Physiol. A* **186,** 975–988 (2000).
28. Morse, P. M. & Ingard, K. U. *Theoretical Acoustics* (Princeton Univ. Press, Princeton, 1986).
29. Kingston, T., Jones, G., Akbar, Z. & Kunz, T. H. Echolocation signal design in Kerivoulinae and Murinae (Chiroptera: Vespertilionidae) from Malaysia. *J. Zool.* **249,** 359–374 (1999).
30. Müller, R. & Kuc, R. Foliage echoes: A probe into the ecological acoustics of bat echolocation. *J. Acoust. Soc. Am.* **108,** 836–845 (2000).

Supplementary Information accompanies this paper on **www.nature.com/nature**.

Acknowledgements We thank all those who assisted with fieldwork. We also thank A. Boonman, A. Denzinger, J. Ostwald, D. Menne, E. Müller, P. Pilz, M. Sánchez-Villagra and P. Stilz for discussions, H. Harty for language assistance and B. Fenton for comments. Our research was funded by the Deutsche Forschungsgemeinschaft (DFG) and a PhD scholarship by Studienstiftung des deutschen Volkes to B.M.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.M.S. (bjoern.siemers@uni-tuebingen.de).

Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast

Balázs Papp[1,2], Csaba Pál[1,3] & Laurence D. Hurst[1]

[1]*Department of Biology and Biochemistry, University of Bath, BA2 7AY Bath, Somerset, UK*
[2]*Collegium Budapest (Institute for Advanced Study) Szentháromság utca 2, Budapest H-1014, Hungary*
[3]*MTA, Ecology and Theoretical Biology Research Group, Eötvös Loránd University, Pázmány Péter Sétány 1/C, Budapest H-1117, Hungary*

Under laboratory conditions 80% of yeast genes seem not to be essential for viability[1]. This raises the question of what the mechanistic basis for dispensability is, and whether it is the result of selection for buffering or an incidental side product. Here we analyse these issues using an *in silico* flux model[2–5] of the yeast metabolic network. The model correctly predicts the knockout fitness effects in 88% of the genes studied[4] and *in vivo* fluxes. Dispensable genes might be important, but under conditions not yet examined in the laboratory. Our model indicates that this is the dominant explanation for apparent dispensability, accounting for 37–68% of dispensable genes, whereas 15–28% of them are compensated by a duplicate, and only 4–17% are buffered by metabolic network flux reorganization. For over one-half of those not important under nutrient-rich conditions, we can predict conditions when they will be important. As expected, such condition-specific genes have a more restricted phylogenetic distribution. Gene duplicates catalysing the same reaction are not more common for indispensable reactions, suggesting that the reason for their retention is not to provide compensation. Instead their presence is better explained by selection for high enzymatic flux.

Although many single-gene deletions have negligible effects on growth rates under laboratory conditions[1,6], the causes and evolution of gene dispensability has remained a controversial issue[7–9]. The capacity of organisms to compensate mutations partly stems from gene duplicates[8], whereas alternative metabolic pathways might also have a role[7,10–12]. The one previous systematic analysis on a eukaryotic organism[13] used a gene's rate of evolution as a proxy for dispensability, a supposition now considered questionable[14]. A third possibility, and one that has received relatively little attention, is that genes only seem to be non-essential, and that they have important roles under environmental conditions yet to be replicated in the laboratory[8,15].

To investigate the causes of gene dispensability, the metabolic capabilities of the *Saccharomyces cerevisiae* network were calculated using flux balance analysis (FBA)[16]. The previously reconstructed network[2,4] consists of 809 metabolites as nodes (including external metabolites), connected by 851 different biochemical reactions (including transport processes). The method first defines a solution space of fluxes of all metabolic reactions in the network that satisfy the governing constraints (that is, steady state of metabolites, flux capacity, direction of reactions, nutrients available in the environments; see Methods). Next, the optimal use of the metabolic network to produce major biosynthetic components for growth can be found among all possible solutions using various optimization protocols[3,4]. The FBA and MOMA[5] (minimization of metabolic adjustment) protocols enable us to predict the phenotypic behaviour of nutritional changes and gene deletions, along with the concomitant changes in flux distributions.

We start by asking how well the mathematical model predicts experimentally measured fluxes, and the effects of gene deletions. We then use it to address the relative importance of the suggested mechanisms for gene dispensability. Finally, we ask whether dispensability is a directly selected feature or a side consequence.

Owing to the availability of systematic knockout studies[1] and some experimentally measured fluxes under four different growth conditions[17], we can directly test the predictive power of the mathematical protocol. We initiated the model to mimic the growth conditions used in these experimental studies. The model correctly predicts: (1) relative differences in flux values; (2) presence or absence of fluxes in 91–95% of the cases; (3) the fitness effects of 88% of single-gene deletions under nutrient-rich growth conditions[4] (see Supplementary Tables S1–S3). Although the model ignores details of gene regulation, the predicted variations in the activity of metabolic pathways across environments are consistent with observations (Supplementary Tables S1 and S2; see also ref. 3). The method, although robust, is not perfect. Although the frequency of experimentally verified essential genes in the group of genes with zero predicted flux is low on rich medium, it is not zero (8.8% for genes with zero flux compared with 28.8% for the rest; $\chi^2 = 18.54$, $P < 10^{-4}$, 1 degree of freedom (d.f.)). The few essential genes in this group probably represent incomplete biochemical knowledge, missing components from the biomass equation, or pleiotropic gene functions[4].

To investigate the possible causes of empirically observed gene dispensability, we compared the predicted and experimentally measured effects of enzyme deletions under nutrient-rich conditions (see Methods). Enzymes were classified into five mutually exclusive groupings on the basis of the presence of isoenzymes, predicted dispensability and flux distribution: (A) enzymes that are inactive under nutrient-rich conditions but active under some other environments; (B) single-copy enzymes that encode essential reactions; (C) duplicated isoenzymes that encode essential reactions; (D) single-copy enzymes that encode dispensable reactions with non-zero enzymatic flux; (E) duplicated isoenzymes that encode dispensable reactions with non-zero enzymatic flux (Fig. 1; see also Supplementary Table S4).

One possible reason that a gene might be non-essential is that its function is not required under a given circumstance. Indeed, the model predicts that a large fraction of experimentally 'verified' non-essential genes should have zero enzymatic flux under nutrient-rich conditions (68.3%). This result indicates that many enzymes make no contribution to the production of biomass components under this condition.

Can we find conditions under which the apparently non-essential genes with zero predicted flux have an important fitness contribution? As with a previous study[3], we set up nine different growth conditions that might have been representative during the evolution of this species (Fig. 2), and performed enzyme-based deletion studies. Under nutrient-rich medium, the fraction of essential reactions and reactions with non-zero flux are especially low (Fig. 2). Importantly, more than half of the experimentally verified non-essential genes that are predicted to have zero flux under nutrient-rich condition appear to have non-zero flux under some other conditions (54.5%, 79 out of 145 cases). These we define as 'conditionally active' genes. This contrasts with the unconditionally active (non-zero flux under all conditions examined) genes and those for which we cannot find conditions under which they are active. These results suggest that 37–68% of the seemingly dispensable genes are environmentally specific (Supplementary Table S4).

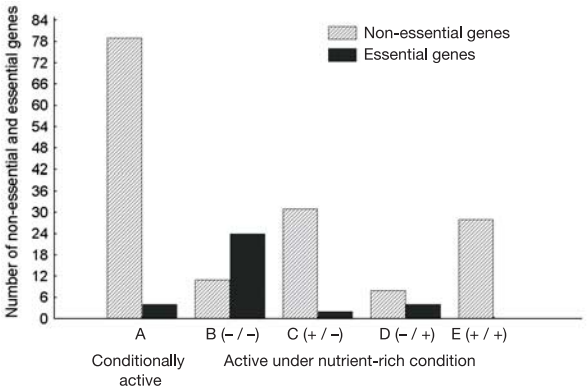


Figure 1 Number of experimentally verified essential and non-essential genes in different categories. The classes are: (A) predicted to have zero flux under nutrient-rich conditions, but non-zero flux in at least one other environment; (B) single-copy genes predicted to catalyse essential reactions; (C) duplicate genes predicted to catalyse essential reactions; (D) single-copy genes predicted to catalyse dispensable reactions; and (E) duplicate genes predicted to catalyse dispensable reactions. When comparing groups B and C, of the 68 metabolic genes that are predicted to catalyse essential reactions, 33 are known to have a duplicated isoenzyme. Only about 6% of those that have an isoenzyme are observed to be essential *in vivo*, whereas the proportion of essential genes is roughly 69% among those without an isoenzyme ($\chi^2 = 28.1$, d.f. $= 1$, $P < 10^{-6}$). When comparing groups B and D, of the 47 single-copy genes 35 are predicted to catalyse essential reactions whereas 12 are predicted to be dispensable. The fraction of essential genes is indeed higher in the former class (about 69% versus about 33%, $\chi^2 = 4.6$, $P < 0.05$). A plus sign indicates the presence and a minus sign the absence of isoenzymes/flux compensation.

Many of the conditionally active genes (76%) are predicted to catalyse reactions that are essential under specific conditions. It remains to be seen whether experiments will actually confirm the detailed predictions.

If the above classifications are correct, one should expect differences in the phylogenetic distribution of enzymes with unconditional and conditional activity, as the latter group can more easily be lost during evolution if the appropriate environment becomes rare. We investigated this issue using a database[18] of enzymatic reactions across 133 sequenced genomes. We found that enzymes having non-zero fluxes under only a few environmental conditions tend to have a more limited phylogenetic distribution than enzymes with unconditional activity (Fig. 3).

Why is it that there are dispensable genes associated with non-zero predicted fluxes under nutrient-rich conditions? To shed light on the relative importance of the compensation mechanisms (duplication versus flux reorganization in the network), we first compared the fraction of experimentally verified essential genes between single-copy enzymes and duplicated isoenzymes. For the comparison, only genes that are predicted to encode essential reactions were considered (Fig. 1, class B and C). The low fraction of essential enzymes with isoenzymes strongly supports previous claims that dispensability partially results from redundant gene duplicates[8]. The two exceptions (failure of compensation) might be due to lack of duplicate enzyme activity in the same subcellular compartment (Supplementary Table S5). Assuming that all or none of the non-essential genes of class E are due to gene duplication rather than flux reorganization, we obtained a lower (14.6%) and upper estimate (27.8%) for the contribution of gene duplication to dispensability (Supplementary Table S4).

The ability of duplicates to buffer each other's loss may be considered a special case of a more general mode of compensation, in which the metabolic network adjusts the metabolic flux, and, in so doing, mitigates the loss of individual genes. Compensation occurs only if the original enzyme has a contribution to biomass production (non-zero flux), but the underlying reaction is dispensable for growth[19] (class D and E, Fig. 1). To see the effect of flux reorganization on *in vivo* gene dispensability independent of duplicate gene copies, we compared the fraction of experimentally verified essential genes between class B and D under the assumption that essential and dispensable reactions should differ in the network's ability to compensate for their loss. Indeed, this is what we observed (Fig. 1). However, this mode of compensation can only explain 3.8–17% of gene dispensability (Fig. 1; see also Supplementary Table S4).

What factors might limit the compensatory capability of the metabolic network? Our model demonstrates that the extent of flux

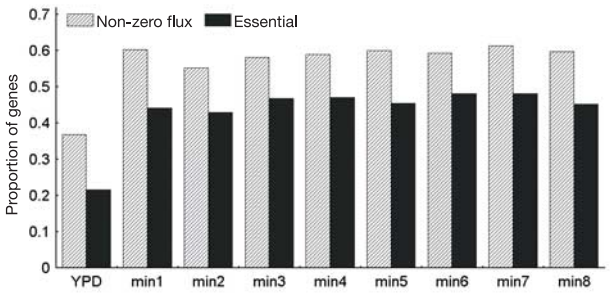


Figure 2 The proportion of genes predicted to have non-zero flux and to be essential under different growth conditions. Single-enzyme knockouts were simulated under nine different growth conditions listed below. The total number of investigated enzymes was 310 for all conditions (isoenzymes were counted only once). Environments were: YPD, rich glucose, low O_2; min1, minimal glucose, low O_2; min2, minimal glucose, anaerobic; min3, minimal ethanol, low O_2; min4, minimal acetate, low O_2; min5, minimal glucose, carbon limited; min6, minimal glucose, nitrogen limited; min7, minimal glucose, phosphate limited; min8, minimal glucose, sulphur limited.

reorganization positively correlates with the predicted fitness effect of the compensated knockout (Supplementary Fig. S1), suggesting that the yeast metabolic network has difficulties in tolerating large flux reorganization.

Although it is clear that the presence of isoenzymes has a large effect on gene dispensability, is it likely that dispensability has evolved to enable such compensation or might it instead be a side product? In the former case, one would expect gene duplicates to be preferentially maintained if they specify a crucial function to provide a shield against intracellular noise[20]. If it were so, then one would expect the most important reactions of the network to be under the control of isoenzymes. In contrast to expectations, essential reactions are not more likely to be catalysed by isoenzymes compared to non-essential reactions (Supplementary Table S6). One possible alternative explanation for the maintenance of isoenzymes is that selection favours enhanced dosage of the same product to provide high enzymatic flux[21]. We found strong support for this theory: the average predicted flux of reactions catalysed by isoenzymes is higher under all conditions than that of reactions catalysed by single-copy enzymes (Supplementary Table S7).

Although many suggest that the high number of dispensable genes is evidence for the selection of robustness to perturbation, our results support a different conclusion. For the most part, knockout studies performed under nutrient-rich conditions provide a substantial underestimate of the number of genes that are essential under some environmental conditions (Fig. 2). Moreover, non-essential genes may make small but significant contributions to fitness even under routine growth conditions, but the effects are not large enough to be detected[8,15]. Of those that seem to be truly dispensable (non-zero flux and viable knockout), at least in the case of gene duplicates, the dispensability is better explained as a side consequence, rather than the result of selection to favour resilience. These results, along with previous studies[21,22], indicate that the dosage requirements have an important influence on the evolutionary maintenance of gene duplicates in yeast.

Is it likely that environmental specificity explains much of the apparent dispensability seen in other organisms? Recent systematic deletion studies[1,23–25] suggest that despite the apparent differences in metabolic complexity and the extent of gene duplication across free-living bacterial and eukaryotic species, the fraction of essential genes under a given laboratory condition is generally low, in the range of 7–19%. In contrast to these low figures in free-living species, the fraction of essential genes is 55–73% in the *Mycoplasma genitalium* genome[26]. This is not simply due to a rarity of gene duplicates. We suggest that, being a parasite with strict host and tissue specificity, *M. genitalium* should have relatively few condition-specific genes. We can test this hypothesis by comparing the proportion of single-copy genes that are non-essential in yeast and in *Mycoplasma*. In agreement with the theory, in yeast this is at least 62%, whereas this drops to 24% in *Mycoplasma*. More direct evidence comes from a data set on the growth phenotypes of mutant strains in *Escherichia coli*[27]: most genes show severe fitness defects only under a small fraction (10%) of the 282 conditions investigated (Supplementary Fig. S2). Moreover, in agreement with the results on yeast metabolic genes, condition-specific genes of *E. coli* show limited phylogenetic distribution (Supplementary Fig. S3).

These issues are important, not least because they affect our ability to test reliably hypotheses concerning the evolution of genes and genomes. For example, the abundance of environmentally specific genes in yeast might explain why dispensability under nutrient-rich conditions only very weakly correlates with the rate of protein evolution[14]. □

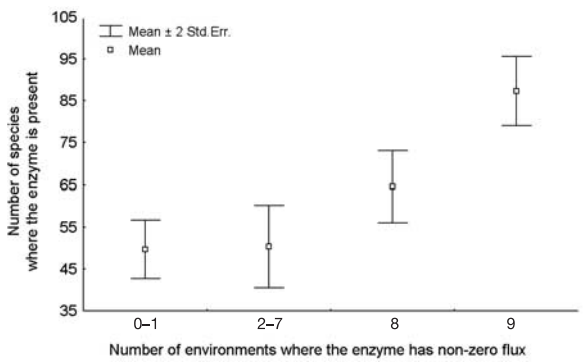


Figure 3 Relationship between phylogenetic distribution and condition specificity. Enzymes having non-zero fluxes in most of the simulated environments have wide phylogenetic distributions (analysis of variance: $F = 17.72$; d.f. $= 3, 281$; $P < 10^{-9}$). Data are means (square symbols) ± 2 standard errors.

Methods

Filtered data sets used in this study

To investigate the metabolic network we used a previously compiled list of enzymatic reactions in yeast[2]. The metabolic reconstruction gives accurate information on the stoichiometry and direction of enzymatic reactions and on the presence of isoenzymes. Cytosolic, mitochondrial and extracellular metabolites are treated separately, and the data set also includes a list of transport reactions between compartments. Reactions catalysed by isoenzymes were considered as a single flux, eliminating duplicate reactions. For data analyses we restricted our attention to unambiguously classified enzymes; that is, those with complete EC numbers. Sequence similarity between isoenzyme pairs was computed by a pair-wise BLASTP[28] search (we used an *E*-value of <0.01 as a cutoff to recognize even distant duplicated isoenzymes). We checked whether the duplicated isoenzymes act in protein complexes, using both the compiled list on yeast metabolism and the MIPS CYGD[29] catalogue on known protein complexes, and these pairs ($N = 21$) were excluded from further analysis. Classification of the dispensability of genes on glucose-rich medium (essential versus non-essential) was as provided by the *Saccharomyces* Genome Deletion Project (http://www-sequence.stanford.edu/group/yeast_deletion_project/), which contains information on large-scale knockout studies[1]. To minimize confounding factors in designation of dispensability, multienzyme polypeptides, genes participating in protein complexes (according to the MIPS CYGD catalogue of annotated complexes) and genes with overlapping reading frames were excluded from all of the analyses. The KEGG database[18] was used to identify the enzymatic reactions of 133 bacterial and eukaryotic species with complete genome sequences (a filtered set of genomes consisting of only one genome per genus gives similar results).

Basic metabolic network model

Flux distribution and metabolic network capabilities were investigated by modification of a previously elaborated genome-scale metabolic flux balance model of *S. cerevisiae*[2–4]. The model starts by specifying the mass balance constraints around intracellular metabolites. These constraints specify a series of linear equations of individual reaction fluxes that must be fulfilled to enable steady state of metabolites. Mathematically, this is represented by $S\mathbf{v} = 0$, where S is the $m \times n$ stoichiometric matrix, with m as the number of metabolites, and n as the number of reactions. An S_{ij} element of the stoichiometric matrix represents the contribution of a jth reaction to metabolite i. The vector $\mathbf{v}$ represents the individual fluxes of the network. Besides mass balance equations, reversibility/irreversibility constraints are also imposed on individual internal fluxes ($\mathbf{v}_i > 0$ for irreversible reactions). Import flux of external metabolites was constrained to be zero when not available under the studied environment. The system also includes a biomass reaction (with rate $\mathbf{v}_{\text{growth}}$) that represents the relative contribution of metabolites to the cellular biomass of yeast (see Supplementary equation S1). Linear programming was used to find a particular flux distribution that maximizes $\mathbf{v}_{\text{growth}}$ under the described constraints and defined nutrient uptake rates. We used this optimal flux configuration as the wild type under the given growth conditions. We have investigated nine different environments (see Fig. 2).

Calculating the fitness effect of gene knockouts

Enzyme deletions were simulated by constraining the flux of the corresponding reactions to zero and calculating the knockout flux configuration under the assumption that knockout metabolic fluxes undergo a minimal flux redistribution with respect to the flux configuration of the wild type (minimization of metabolic adjustment, MOMA protocol[5]). Using a different optimization protocol[3,4] gives almost exactly the same results (data not shown). Thus, calculation of knockout $\mathbf{v}_{\text{growth}}$ requires quadratic programming to find a point in flux space, which is closest to wild type. The software tool Cplex 7.5 was used to solve these linear and quadratic optimization problems. We scaled fitness relative to the wild type. Essential enzymes are defined as knockout strains having a growth rate of at most one-half of the wild type. We observed a clear bimodal distribution of knockout fitnesses: enzymes predicted to be non-essential have minimal or no effect on growth (Supplementary Fig. S4). If the optimization problem for a given knockout was infeasible we treated the enzyme as essential. Flux and knockout phenotype predictions were not attempted for enzymes located on dead-end pathways or for enzymes with functions not represented in the biomass equation[4] (for example, glycoprotein, haem and chitin metabolism, transfer RNA synthetases). In the case of reactions catalysed by isoenzymes, the

duplicates were deleted to obtain predictions on the dispensability of the underlying enzymatic reaction.

Comparison of *Mycoplasma* and *Saccharomyces* genomes

We calculated the frequency of non-essential genes in the *M. genitalium* and the *S. cerevisiae* genomes (only single-copy genes were considered). Gene duplicates were identified using a BLAST protein search, with at least 25% amino acid similarity (using different thresholds do not affect our results). The list of putative essential *Mycoplasma* genes is from ref. 26. We found 1,881 out of 3,003 single-copy yeast genes that are non-essential. This figure is 83 out of 356 genes for *Mycoplasma*.

Received 14 January; accepted 30 April 2004; doi:10.1038/nature02636.

1. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
2. Forster, J., Famili, I., Fu, P., Palsson, B. O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13,** 244–253 (2003).
3. Famili, I., Forster, J., Nielsen, J. & Palsson, B. O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl Acad. Sci. USA* **100,** 13134–13139 (2003).
4. Forster, J., Famili, I., Palsson, B. O. & Nielsen, J. Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae*. *OMICS* **7,** 193–202 (2003).
5. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99,** 15112–15117 (2002).
6. Steinmetz, L. M. *et al.* Systematic screen for human disease genes in yeast. *Nature Genet.* **31,** 400–404 (2002).
7. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24,** 355–361 (2000).
8. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421,** 63–66 (2003).
9. Gu, X. Evolution of duplicate genes versus genetic robustness against null mutations. *Trends Genet.* **19,** 354–356 (2003).
10. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420,** 190–193 (2002).
11. Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. & Palsson, B. O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28,** 250–258 (2003).
12. Emmerling, M. *et al.* Metabolic flux responses to pyruvate kinase knockout in Escherichia coli. *J. Bacteriol.* **184,** 152–164 (2002).
13. Kitami, T. & Nadeau, J. H. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genet.* **32,** 191–194 (2002).
14. Pál, C., Papp, B. & Hurst, L. D. Rate of evolution and gene dispensability. *Nature* **421,** 496–497 (2003).
15. Thatcher, J. W., Shaw, J. M. & Dickinson, W. J. Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl Acad. Sci. USA* **95,** 253–257 (1998).
16. Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97,** 5528–5533 (2000).
17. Daran-Lapujade, P. *et al.* Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*: a chemostat culture study. *J. Biol. Chem.* **279,** 9125–9138 (2004).
18. Nakao, M. *et al.* Genome-scale gene expression analysis and pathway reconstruction in KEGG. *Genome Inform. Ser. Workshop Genome Inform.* **10,** 94–103 (1999).
19. Schilling, C. H., Edwards, J. S., Letscher, D. & Palsson, B. O. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **71,** 286–306 (2000).
20. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51,** 329–347 (2000).
21. Seoighe, C. & Wolfe, K. H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2,** 548–554 (1999).
22. Papp, B., Pal, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424,** 194–197 (2003).
23. Gerdes, S. Y. *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185,** 5673–5684 (2003).
24. Kobayashi, K. *et al.* Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA* **100,** 4678–4683 (2003).
25. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421,** 231–237 (2003).
26. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286,** 2165–2169 (1999).
27. Glasner, J. D. *et al.* ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* **31,** 147–151 (2003).
28. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
29. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30,** 31–34 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank J. Förster and B. Palsson for discussions and providing information on the yeast flux balance model. We also thank J. Glasner for providing growth data on the *E. coli* wild-type strain. B.P. and C.P. were supported by the Hungarian National Research Grant Foundation (OTKA).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.D.H. (l.d.hurst@bath.ac.uk).

Temporal difference models describe higher-order learning in humans

Ben Seymour[1], John P. O'Doherty[1], Peter Dayan[2], Martin Koltzenburg[3], Anthony K. Jones[4], Raymond J. Dolan[1], Karl J. Friston[1] & Richard S. Frackowiak[1,5]

[1]*Wellcome Department of Imaging Neuroscience, 12 Queen Square, London WC1N 3BG, UK*
[2]*Gatsby Computational Neuroscience Unit, Alexandra House, 17 Queen Square, London WC1N 3AR, UK*
[3]*Institute of Child Health, University College London, 30 Guilford St, London WC1N 1EH, UK*
[4]*University of Manchester Rheumatic Diseases Centre, Hope Hospital, Manchester M6 8HD, UK*
[5]*Fondazione Santa Lucia, 00179 Rome, Italy*

The ability to use environmental stimuli to predict impending harm is critical for survival. Such predictions should be available as early as they are reliable. In pavlovian conditioning, chains of successively earlier predictors are studied in terms of higher-order relationships, and have inspired computational theories such as temporal difference learning[1]. However, there is at present no adequate neurobiological account of how this learning occurs. Here, in a functional magnetic resonance imaging (fMRI) study of higher-order aversive conditioning, we describe a key computational strategy that humans use to learn predictions about pain. We show that neural activity in the ventral striatum and the anterior insula displays a marked correspondence to the signals for sequential learning predicted by temporal difference models. This result reveals a flexible aversive learning process ideally suited to the changing and uncertain nature of real-world environments. Taken with existing data on reward learning[2], our results suggest a critical role for the ventral striatum in integrating complex appetitive and aversive predictions to coordinate behaviour.

Substantial evidence in humans and other animals has outlined a network of brain regions involved in the prediction of painful and aversive events[3–6]. Most of this work has concentrated on its simplest realization, namely first-order pavlovian fear conditioning; however, the predictions in this paradigm are rudimentary, showing little of the complexities associated with sequences of predictors that are critical in psychological investigations of prognostication[7]. These latter studies led to a computational account called temporal difference learning[1,8], which has close links with methods for prediction, and optimal action selection, in engineering[9]. When applied to first-order appetitive conditioning, temporal difference learning provides a compelling account of neurophysiological data, both with respect to the phasic activity of dopamine neurons in animal studies, and with blood-oxygenation-level-dependent (BOLD) activity in human functional neuroimaging studies[10–15]. However, beyond this simple paradigm, the utility of temporal difference models to describe learning remains largely unexplored. Here we provide a neurobiological investigation based on aversive and, importantly, sequential conditioning.

We used fMRI to investigate the pattern of brain responses in humans during a second-order pain learning task. Fourteen healthy subjects were shown two visual cues in succession, followed by a high- or low-intensity pain stimulus delivered to the left hand (Fig. 1a) (see Methods). Subjects were told that they were performing a study of reaction times and were asked to judge whether the cues appeared on the left or on the right side of a display monitor. The second cue in each sequence was fully predictive of the strength of the subsequently experienced pain; however, the first cue only allowed a probabilistic prediction. Thus, in a small percentage of

nature genetics

Adaptive evolution of bacterial metabolic networks by horizontal gene transfer

Csaba Pál[1,2], Balázs Papp[2,3] & Martin J Lercher[1,4]

Numerous studies have considered the emergence of metabolic pathways[1], but the modes of recent evolution of metabolic networks are poorly understood. Here, we integrate comparative genomics with flux balance analysis to examine (i) the contribution of different genetic mechanisms to network growth in bacteria, (ii) the selective forces driving network evolution and (iii) the integration of new nodes into the network. Most changes to the metabolic network of *Escherichia coli* in the past 100 million years are due to horizontal gene transfer, with little contribution from gene duplicates. Networks grow by acquiring genes involved in the transport and catalysis of external nutrients, driven by adaptations to changing environments. Accordingly, horizontally transferred genes are integrated at the periphery of the network, whereas central parts remain evolutionarily stable. Genes encoding physiologically coupled reactions are often transferred together, frequently in operons. Thus, bacterial metabolic networks evolve by direct uptake of peripheral reactions in response to changed environments.

Although horizontal gene transfer shapes bacterial genomes[2,3], most large-scale analyses have ignored its influence on the evolution of biological networks. Theoretical models[1] and systematic analyses[4–6] of the evolution of metabolic networks concentrate on the effects of gene duplicates. Similarly, the selective forces that influence the growth of biochemical networks are largely unknown. Here, we analyze these issues using the previously reconstructed metabolic network[7] of *Escherichia coli* K-12, composed of 904 proteins and 931 unique biochemical reactions, including coenzymes and transport processes of specified external nutrients.

In eukaryotes, gene duplicates are the main source of evolutionary novelties. Is gene duplication also the dominant genetic mechanism contributing to growth of bacterial biochemical networks? In sharp contrast to the eukaryotic yeast *Saccharomyces cerevisiae*, *E. coli* K-12 contains few duplicated enzymes in its metabolic network, almost all of which seem to be ancient (**Fig. 1**). Detailed phylogenetic analysis (**Supplementary Methods** online) indicated that only 1 of 451 investigated duplicated enzymes in *E. coli* arose since the divergence from *Salmonella* ~100 million years ago[8], despite vast differences in lifestyle and genome content between those two species[9]. Moreover, this one duplicate pair (ornithine carbamoyltransferase 1 and 2) functions in the same enzymatic reaction. Therefore, gene duplication had little effect on the topology of the *E. coli* metabolic network over the last 100 million years.

An alternative source of network growth is horizontal gene transfer. To identify transfer events, we first established the phylogeny of 51 proteobacteria species including *E. coli* K-12 and several of its close relatives, using 5 additional species to root the phylogenetic tree. The

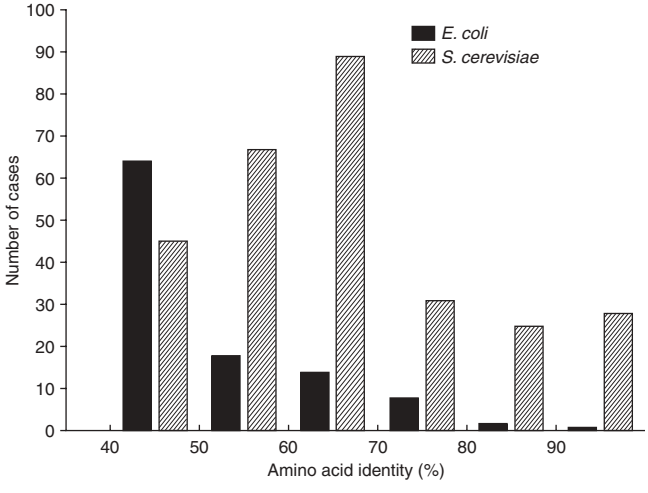


Figure 1 Comparison of duplicate genes in the metabolic networks of yeast (*S. cerevisiae*)[29] and *E. coli*. The distributions of amino acid sequence similarities differ strongly between the two species ($N = 107$ genes (*E. coli*), $N = 285$ genes (*S. cerevisiae*), $P < 0.001$ from Kolmogorov-Smirnov two-sample test). Amino acid sequence similarities between all gene pairs in each network were computed by BLAST, retaining all pairs with more than 40% amino acid similarity. The result remains after excluding remnants of genome duplicates in yeast[30] ($N = 107$ (*E. coli*), $N = 243$ (*S. cerevisiae*), $P < 0.001$ from Kolmogorov-Smirnov two-sample test).

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany. [2]MTA, Theoretical Biology and Ecology Research Group, Eötvös Loránd University, Budapest H-1117, Hungary. [3]Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK. [4]Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. Correspondence should be addressed to M.J.L. (lercher@embl.de).

Received 6 May; accepted 8 September; published online 20 November 2005; doi:10.1038/ng1686

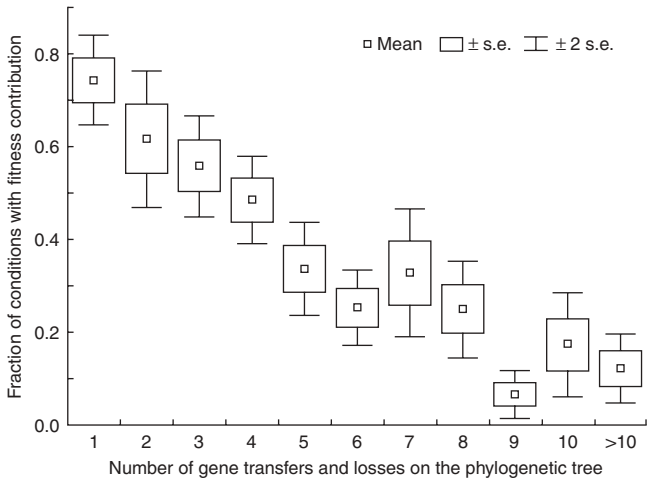


Figure 2 Environment-specificity of proteins increases with the frequency of horizontal transfer and loss events of the encoding genes ($N = 689$; ANOVA: $P < 10^{-7}$, $F = 16.32$, d.f. = 10). s.e., standard error.

maximum-likelihood tree reconstructed from 47 concatenated protein sequences was well-supported by bootstrap analyses (**Supplementary Fig. 1** online), and comparison with four independent phylogenetic studies confirmed the branching order of all previously investigated species sets (**Supplementary Methods**). Following earlier studies[10–12], we then used the presence or absence of proteins among the 51 species to identify the most parsimonious scenarios for horizontal gene transfers and gene losses across the reconstructed tree. We did this for each of 2,325 orthologous families with members in *E. coli* K-12 (numbers of inferred transfers at each node are listed in **Supplementary Table 1** online). Our results rely on the biologically reasonable assumption that gene losses are approximately twice as likely to occur as are transfer events[10,11] (gain/loss penalty ratio = 2); we obtained very similar results with other parameter settings (**Supplementary Methods**).

Consistent with expectations and earlier observations[13], a large fraction (30%) of the most recently transferred genes are annotated[14] with virus- or transposon-related functions (**Supplementary Fig. 2** online). For recently acquired genes, our results are in good agreement with those from complementary approaches[13] based on irregular GC content and use of suboptimal codons (**Supplementary Table 2** online). We found a gradual decay of both GC and codon usage irregularities with the age of the transfer event (**Supplementary Fig. 3** online), providing support for the previously hypothesized 'amelioration' of compositional biases over evolutionary time[13]. Under realistic parameter settings, we estimated that 15–32 genes were transferred horizontally into the *E. coli* metabolic network since its divergence from the *Salmonella* lineage, vastly outnumbering the one identified gene duplication over the same period.

Although gene duplication may have been an important source for network changes during the early evolution of pathways[1], the above analyses suggest that horizontal gene transfer was the dominant genetic mechanisms in the recent expansion of metabolic networks in bacteria. Which forces may be responsible for the low contribution of gene duplication to bacterial network growth? The foremost difficulty for the expansion of gene families is preserving both copies until they develop functionally distinct roles[2]. Moreover, the initial preservation of duplicated genes probably depends on the effect of enhanced gene dosage, which will be deleterious except under specific selection pressures[15]. Most gene duplicates are quickly removed from bacterial populations[16].

What are the selective pressures driving the acquisition of foreign genes? In comparisons with a systematic experimental gene knockout study[17], we found that only 7% of the genes horizontally transferred into the metabolic network of *E. coli* are essential under nutrient-rich laboratory conditions, compared with 23% of other genes ($N = 761$ genes, $\chi^2 = 26.53$, degrees of freedom (d.f.) = 1, $P < 5 \times 10^{-7}$). This observation is consistent with at least two hypotheses. First, transferred genes may provide small but evolutionarily important contributions to fitness, even under the examined routine growth conditions[18]. Alternatively, horizontal gene transfers might confer condition-specific advantages, facilitating adaptation to new environments. To assess the fitness contribution of all metabolic *E. coli* K-12 genes under different environments *in silico*, we carried out flux balance analyses of the metabolic network[19] (very similar results were obtained with minimization of metabolic adjustment analyses[20]; **Supplementary Methods** and **Supplementary Table 2**). Assuming a steady state of metabolite concentrations, we determined the flux distribution that maximized the production of a physiological combination of major biosynthetic components, the biomass, for a given set of available nutrients (**Supplementary Methods**).

Using a previously described protocol[19], we investigated systematically the effect of gene deletions on fitness in different environments, approximating fitness by the rate of biomass production. We examined 136 simulated environments, characterized by their main carbon source and the availability of oxygen, which had been shown *in silico* to support aerobic and/or anaerobic growth[21] (**Supplementary Table 3**). Those genes that contributed most to the evolution of metabolic networks (*i.e.*, that were frequently gained or lost during the

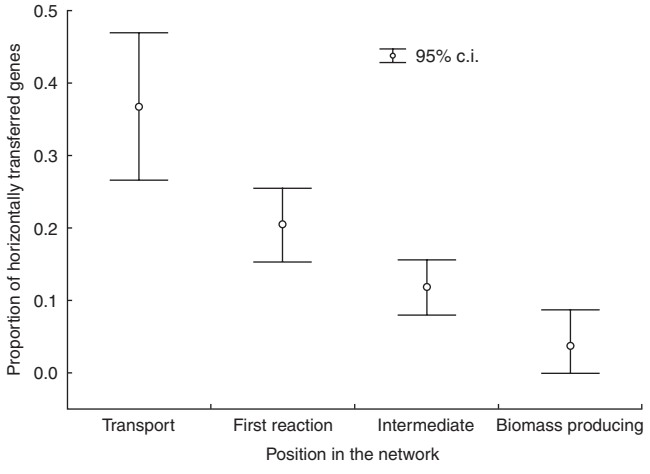


Figure 3 Proteins at the periphery of the metabolic network are much more likely to have undergone horizontal gene transfer into the *E. coli* lineage since its split from the *Vibrio* lineage. Genes are divided into the following groups: (i) transport proteins involved in nutrient uptake (87 genes); (ii) enzymes catalyzing the first reaction after uptake (240 genes); (iii) enzymes catalyzing internal reactions (271 genes); and (iv) enzymes producing major biosynthetic components (55 genes). $P < 10^{-7}$, $\chi^2 = 37.03$, d.f. = 3. Cofactors and metabolites involved in large numbers of reactions were excluded from the metabolic map, including NAD^+, NADH, NADPH, $NADP^+$, H^+, ATP, ADP, orthophosphate, CO_2, pyrophosphate, FAD, $FADH_2$ and H_2O. Genes with ambiguous network positions were excluded from the analysis. Results remain for gain/loss penalty = 1 ($\chi^2 = 65.39$, d.f. = 3, $P < 10^{-13}$), as well as when transfer events across the whole phylogenetic tree are considered (data not shown). c.i., confidence interval.

Table 1 Physiologically coupled enzyme pairs are frequently transferred or lost together

Interaction	Event	Individual events	Fraction of co-events	OR (95% c.i.)
Fully coupled	Transfer	59	37%	64.6 (24.2–168.8)
Fully coupled	Loss	1,624	53%	50.0 (41.8–59.6)
Directionally coupled	Transfer	78	30%	60.3 (24.3–147.2)
Directionally coupled	Loss	2,833	21%	9.6 (8.3–11.1)

'Individual events' is the total number of individual gene gains (or losses) investigated. 'Fraction of co-events' is the fraction of the total gains (or losses) of genes involved in physiologically coupled pairs that occur together with their coupled partner. Only branches originating from an ancestral node in which both genes are absent (or present) were considered in the analysis of 'co-gains' (or 'co-losses'). Odds ratios (ORs) quantify how much more likely gain (or loss) of a gene is when its coupled partner gene is gained (or lost) along the same phylogenetic branch; all odds ratios are highly significant (Fisher's exact test, $P < 10^{-12}$). Similar results were obtained with different model settings (**Supplementary Table 2**). c.i., confidence interval.

evolution of proteobacteria) were generally environment-specific, whereas those genes that were invariant among proteobacteria contributed to fitness in most environments (**Fig. 2**). Previous analyses showed quantitative agreement between the predictions of the flux balance model and experimentally measured nutrient uptake, enzymatic fluxes and the effects of gene deletions under several conditions in *E. coli*[19], but detailed predictions under several other conditions remain to be validated. The above result remained valid when the analysis was restricted to conditions where growth of *E. coli* K-12 was experimentally shown; it was also robust to changes in the method to identify gene transfers and losses and in the optimization protocols to calculate the impact of gene deletions[19,20] (**Supplementary Table 2**). The environment-specificity of horizontally transferred genes might explain why most of them are not translated into proteins under laboratory conditions[22]. We conclude that the evolution of the network is largely driven by adaptation to new environments and not by optimization in fixed environments.

Having established the genetic mechanisms and the selective forces that govern network evolution, we next turned to the topological effect of horizontal gene transfer on the network. The above results suggest that addition and deletion of reactions might be concentrated on those network parts that interact with the environment. The number of independent horizontal transfer events was highly variable across different enzymatic pathways (**Supplementary Table 2**), and genes in the central pathways of the network (*e.g.*, glycolysis) had undergone few transfer events across the tree. To analyze further the relationship between network position and gene transfers, we classified proteins according to their involvement in nutrient uptake, first reactions after uptake, intermediate steps of metabolism and production of major biosynthetic components. As predicted, proteins contributing to peripheral reactions (nutrient uptake and first metabolic step) were more likely to be transferred, whereas enzymes catalyzing central reactions (intermediate steps and biomass production) were largely invariant across species (**Fig. 3**).

Are genes added or lost from metabolic networks one at a time, or does network evolution proceed by steps involving whole sets of genes simultaneously? Modules of physiologically coupled genes might be the best candidates for simultaneous acquisition or loss during evolution. We identified physiologically coupled enzyme pairs by flux-coupling analysis[23]. Two special cases were considered: fully and directionally coupled enzyme pairs. In fully coupled enzyme pairs, the flux catalyzed by one protein is always the same as that catalyzed by the other except for a constant factor, as in linear pathways. Fully coupled enzymatic pairs provide a very rigorous and stringent definition of biochemical modules, as only together can such pairs fulfill their metabolic function. Directional coupling indicates that removal of one enzyme shuts down flux through the other but not vice versa. As predicted, both fully and directionally coupled enzymes were much more often gained or lost together on the same branch of the proteobacterial phylogenetic tree than would be expected by chance (**Table 1**). This suggests that physiological modules tend to be conserved during evolution, contrary to previous results based on more loosely defined modules[24].

Moreover, 30% of the fully coupled pairs are encoded in the same operon in *E. coli*[25], a fraction much higher than would be expected for randomly chosen pairs (0.5%). The fraction of pairs sharing the same operon rises to at least 75% when considering only fully coupled pairs that were gained together during evolution leading to *E. coli*. These latter results confirm that the gains of physiologically fully coupled pairs together most likely occurred in one step, the uptake of at least part of an operon.

Future studies will aim to characterize the molecular details of the evolutionary network dynamics, for example, by analyzing how the enzymatic composition of the network affects its ability to adapt to

Table 2 Some operons containing horizontally transferred genes with physiologically coupled, environment-specific functions

Operon name	Predicted nutrient for which operon is required	Literature information	Genes in operon and physiological coupling
atoDAE	Acetoacetate, butyrate	Short fatty acids	**atoE** ↔ **atoD/atoA***
codBA	NA	Cytosine	**codB** → codA
cynTSX	NA	Cyanate	**cynX** ↔ **cynS** → cynT
fucPIKUR	Fucose	Fucose	**fucP** ↔ **fucI** ↔ **fucK***
melAB	Melibiose	Melibiose	**melB** ↔ **melA**
mtlADR	Mannitol	Mannitol	**mtlA** ↔ **mtlD***
AraBAD	Arabinose	Arabinose	**araA** ↔ **araB** ↔ araD

The operons listed are required for the uptake or catalysis (mostly the first or second step after uptake) of specific nutrients. Members of these operons are physiologically coupled. Nutrient conditions where the operons are required are derived from the model and from literature compiled from RegulonDB and Ecocyc[14]. Genes in the operons are listed as ordered on the metabolic map. Genes involved in transport processes are underlined. Genes horizontally transferred to *E. coli* are depicted in bold. Physiological coupling between genes is denoted by arrows (↔, fully coupled; →, directionally coupled). Genes that are not part of the metabolic network are excluded from the analysis. Unless otherwise indicated (*), evidence for horizontal transfer is consistent under all investigated parameter settings. NA, not analyzed.

new environments. As a first step, we examined whether the gradual evolution of metabolic pathways can be understood by analyzing the details of physiological coupling between enzymes. For example, one might expect that an enzyme whose function depends on the presence of another enzyme would have been acquired by the *E. coli* genome more recently than its partner. This is indeed observed in 70% of those directionally coupled gene pairs in which both genes were acquired on different branches leading to *E. coli* ($N = 386$, sign test $P < 10^{-16}$). Future studies will also have to examine how the number of physiological interactions influences the probability of successful gene transfer[26]. Furthermore, given that the physiological adaptation to new environments is accompanied by major flux reorganizations along the high-flux backbone of the metabolic network[27], the role of horizontally transferred genes in these reorganizations needs to be examined.

In summary, metabolic networks in bacteria evolve in response to changing environments, not only by changes in enzyme kinetics through point mutations, but also by the uptake of peripheral genes and operons through horizontal gene transfers (a list of examples is given in **Table 2**). Our results indicate that systems biology cannot stop at the boundaries of the metabolic network: to understand network evolution, we need to extend our analysis to the environment, both inanimate (providing nutrients) and animate (providing genetic material).

METHODS

Gene gains and losses. Based on gene presence and absence obtained from STRING[28], we reconstructed the most parsimonious scenarios[10–12] for gene loss and horizontal transfer events (gene gains) on the rooted phylogeny using generalized parsimony as implemented in PAUP* (**Supplementary Methods**). All results were obtained using relative penalties for horizontal gene transfer and deletions of 2:1 (gain/loss penalty = 2)[10,11]; different settings gave similar results (**Supplementary Table 2**).

To analyze CDgains, we started with nodes of the phylogenetic tree in which the two investigated enzymes (*e.g.*, A and B) were absent. We then constructed a contingency table by counting the occurrence of the four possible evolutionary scenarios (gain of A, gain of B, gain of A and B, and no gain) along all branches starting from these nodes. The odds ratio quantifies how much more likely the gain of a gene A is if its physiologically coupled partner gene B is gained along the same phylogenetic branch. We used an analogous procedure for loss events, analyzing all branches starting from nodes in which both A and B were present. Gene families with more than one member in *E. coli* K-12 were excluded from the analysis.

Metabolic network analysis. We examined the reconstructed metabolic network (*i*JR904 GSM/GPR) of *E. coli* K-12. We followed previously established protocols[19,20] to investigate the effect of gene deletions under 136 environmental conditions. Flux balance analysis involves two fundamental steps: (i) specification of mass balance constraints around intracellular metabolites and (ii) maximization of the production of biomass components (the list of environments and biomass components is given in **Supplementary Table 3** online). Physiologically coupled reactions and blocked reactions were identified as described previously[23]. We found 772 fully coupled reaction pairs and 1,542 directionally coupled reaction pairs.

More methodological details (including ortholog identification, inference of phylogenetic genome tree and age estimation of gene duplicates) are given in **Supplementary Methods**.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank P. Bork, L. Hurst and S. McWeeney for suggestions on previous versions of the manuscript; C. von Mering for discussions and providing early access to the updated STRING database; and E. Nikolaev for discussions on flux coupling analysis. C.P. and B.P. are supported by the Hungarian Scientific Research Fund. B.P. is a Fellow of the Human Frontier Science Program. C.P. acknowledges support by an EMBO Long-Term Fellowship. M.J.L. acknowledges support by the Deutsche Forschungsgemeinschaft.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Schmidt, S., Sunyaev, S., Bork, P. & Dandekar, T. Metabolites: a helping hand for pathway evolution? *Trends Biochem. Sci.* **28**, 336–341 (2003).
2. Lawrence, J.G. & Hendrickson, H. Lateral gene transfer: when will adolescence end? *Mol. Microbiol.* **50**, 739–749 (2003).
3. Lerat, E., Daubin, V., Ochman, H. & Moran, N.A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130 (2005).
4. Teichmann, S.A. *et al.* The evolution and structural anatomy of the small molecule metabolic pathways in *Escherichia coli*. *J. Mol. Biol.* **311**, 693–708 (2001).
5. Rison, S.C., Teichmann, S.A. & Thornton, J.M. Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J. Mol. Biol.* **318**, 911–932 (2002).
6. Alves, R., Chaleil, R.A. & Sternberg, M.J. Evolution of enzymes in metabolism: a network perspective. *J. Mol. Biol.* **320**, 751–770 (2002).
7. Reed, J.L., Vo, T.D., Schilling, C.H. & Palsson, B.O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**, R54 (2003).
8. Lawrence, J.G., Hartl, D.L. & Ochman, H. Molecular considerations in the evolution of bacterial genes. *J. Mol. Evol.* **33**, 241–250 (1991).
9. Ochman, H. & Groisman, E.A. The origin and evolution of species differences in *Escherichia coli* and *Salmonella typhimurium*. *EXS* **69**, *479–493 (1994).*
10. Snel, B., Bork, P. & Huynen, M.A. Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* **12**, 17–25 (2002).
11. Mirkin, B.G., Fenner, T.I., Galperin, M.Y. & Koonin, E.V. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* **3**, 2 (2003).
12. Boussau, B., Karlberg, E.O., Frank, A.C., Legault, B.A. & Andersson, S.G. Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc. Natl. Acad. Sci. USA* **101**, 9722–9727 (2004).
13. Lawrence, J.G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95**, 9413–9417 (1998).
14. Keseler, I.M. *et al.* EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33**, D334–D337 (2005).
15. Papp, B., Pál, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661–664 (2004).
16. Hooper, S.D. & Berg, O.G. On the nature of gene innovation: duplication patterns in microbial genomes. *Mol. Biol. Evol.* **20**, 945–954 (2003).
17. Gerdes, S.Y. *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185**, 5673–5684 (2003).
18. Thatcher, J.W., Shaw, J.M. & Dickinson, W.J. Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl. Acad. Sci. USA* **95**, 253–257 (1998).
19. Price, N.D., Reed, J.L. & Palsson, B.O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**, 886–897 (2004).
20. Segre, D., Vitkup, D. & Church, G.M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* **99**, 15112–15117 (2002).
21. Reed, J.L. & Palsson, B.O. Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res.* **14**, 1797–1805 (2004).
22. Taoka, M. *et al.* Only a small subset of the horizontally transferred chromosomal genes in *Escherichia coli* are translated into proteins. *Mol. Cell. Proteomics* **3**, 780–787 (2004).
23. Burgard, A.P., Nikolaev, E.V., Schilling, C.H. & Maranas, C.D. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* **14**, 301–312 (2004).
24. Snel, B. & Huynen, M.A. Quantifying modularity in the evolution of biomolecular systems. *Genome Res.* **14**, 391–397 (2004).
25. Salgado, H. *et al.* RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res.* **32**, D303–D306 (2004).
26. Jain, R., Rivera, M.C. & Lake, J.A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96**, 3801–3806 (1999).
27. Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z.N. & Barabasi, A.L. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839–843 (2004).
28. von Mering, C. *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **33**, D433–D437 (2005).
29. Forster, J., Famili, I., Fu, P., Palsson, B.O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
30. Kellis, M., Birren, B.W. & Lander, E.S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).

OPEN ACCESS Freely available online

PLOS BIOLOGY

The Genomic Landscape of Compensatory Evolution

Béla Szamecz[1☯], Gábor Boross[1☯], Dorottya Kalapis[1], Károly Kovács[1], Gergely Fekete[1], Zoltán Farkas[1], Viktória Lázár[1], Mónika Hrtyan[1¤], Patrick Kemmeren[2], Marian J. A. Groot Koerkamp[2], Edit Rutkai[3], Frank C. P. Holstege[2], Balázs Papp[1]*, Csaba Pál[1]*

1 Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Center, Szeged, Hungary, **2** Molecular Cancer Research, University Medical Center Utrecht, Utrecht, the Netherlands, **3** Institute for Biotechnology, Bay Zoltán Non-Profit Ltd., Szeged, Hungary

Abstract

Adaptive evolution is generally assumed to progress through the accumulation of beneficial mutations. However, as deleterious mutations are common in natural populations, they generate a strong selection pressure to mitigate their detrimental effects through compensatory genetic changes. This process can potentially influence directions of adaptive evolution by enabling evolutionary routes that are otherwise inaccessible. Therefore, the extent to which compensatory mutations shape genomic evolution is of central importance. Here, we studied the capacity of the baker's yeast genome to compensate the complete loss of genes during evolution, and explored the long-term consequences of this process. We initiated laboratory evolutionary experiments with over 180 haploid baker's yeast genotypes, all of which initially displayed slow growth owing to the deletion of a single gene. Compensatory evolution following gene loss was rapid and pervasive: 68% of the genotypes reached near wild-type fitness through accumulation of adaptive mutations elsewhere in the genome. As compensatory mutations have associated fitness costs, genotypes with especially low fitnesses were more likely to be subjects of compensatory evolution. Genomic analysis revealed that as compensatory mutations were generally specific to the functional defect incurred, convergent evolution at the molecular level was extremely rare. Moreover, the majority of the gene expression changes due to gene deletion remained unrestored. Accordingly, compensatory evolution promoted genomic divergence of parallel evolving populations. However, these different evolutionary outcomes are not phenotypically equivalent, as they generated diverse growth phenotypes across environments. Taken together, these results indicate that gene loss initiates adaptive genomic changes that rapidly restores fitness, but this process has substantial pleiotropic effects on cellular physiology and evolvability upon environmental change. Our work also implies that gene content variation across species could be partly due to the action of compensatory evolution rather than the passive loss of genes.

Citation: Szamecz B, Boross G, Kalapis D, Kovács K, Fekete G, et al. (2014) The Genomic Landscape of Compensatory Evolution. PLoS Biol 12(8): e1001935. doi:10.1371/journal.pbio.1001935

Academic Editor: Nick H. Barton, Institute of Science and Technology Austria (IST Austria), Austria

Received April 15, 2014; **Accepted** July 18, 2014; **Published** August 26, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files. The raw gene expression dataset is available online at ArrayExpress (accession number E-MTAB-2352).

Funding: This work was supported by grants from the European Research Council (202591) and the EMBO Young Investigator Programme (to CP), the Wellcome Trust, and the 'Lendület Program' of the Hungarian Academy of Sciences (to CP and BP), the International Human Frontier Science Program Organization (to BP), the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (to BS), the TÁMOP-4.2.2 B-10/1-2010-0012 grant (to GB), and the TÁMOP 4.2.4. A/2-11-1-2012-0001 'National Excellence Program' (to KK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: pal.csaba@brc.mta.hu (CP); papp.balazs@brc.mta.hu (BP)

☯ These authors contributed equally to this work.

¤ Current address: Functional Genomics and Proteomics of Plants, CEITEC - Masaryk University, Brno, the Czech Republic

Introduction

Deleterious, but non-lethal mutations are constantly generated and can hitchhike with adaptive mutations [1]. Consequently, such deleterious alleles are widespread in eukaryotic populations [2,3]. For example, as high as 12% of the coding SNPs in yeast populations are deleterious [2]. Many of the observed functional variation in this species yield proteins with compromised or no activities [2], or lead to complete loss of genes with significant contribution to fitness (Text S1). Deleterious loss-of-function variants may occasionally revert to wild type, eventually perish from the population, or become compensated by mutations elsewhere in the genome. The third possibility, termed compensatory evolution, is the focus of our study. Theoretical works suggest that mutant subpopulations can cross fitness valleys by the simultaneous fixation of a compensatory mutation in the population [4,5]. This process can also work in large populations and is facilitated by linkage of the two alleles [5].

Compensatory evolution appears to be common at many levels of molecular interactions. It is involved in the maintenance of RNA and protein secondary structures, it mitigates the costs of antibiotic resistance [6,7], and allows rapid fitness recovery in populations with accumulated deleterious mutation loads [7–9]. Compensatory regulatory mutations also act to stabilize gene expression levels across species [10,11], and conserve DNA-encoded nucleosome organization [12]. The most detailed experimental analyses on compensatory mutations for fixed deleterious mutations were performed in DNA bacteriophages

Author Summary

While core cellular processes are generally conserved during evolution, the constituent genes differ somewhat between related species with similar lifestyles. Why should this be so? In this work, we propose that gene loss may initially be deleterious, but organisms can recover fitness by the accumulation of compensatory mutations elsewhere in the genome. To investigate this process in the laboratory, we investigated 180 haploid yeast strains, each of which initially displayed slow growth owing to the deletion of a single gene. Laboratory evolutionary experiments revealed that defects in a broad range of molecular processes can readily be compensated during evolution. Genomic analyses and functional assays demonstrated that compensatory evolution generates hidden genetic and physiological variation across parallel evolving lines, which can be revealed when the environment changes. Strikingly, despite nearly full recovery of fitness, the wild-type genomic expression pattern is generally not restored. Based on these results, we argue that genomes undergo major changes not simply to adapt to external conditions but also to compensate for previously accumulated deleterious mutations.

[8,13–15], bacteria [16,17], and *Caenorhabditis elegans* [7,9]. Three major patterns emerged from these studies. As the target size for compensatory mutations is typically much larger than that for reversion, compensation is more likely than reversion of deleterious mutations [13]. The rate of compensatory evolution increased with the severity of the deleterious fitness effects, and was not limited to functionally interacting partners of the mutated gene [15].

As regards the potential pleiotropic effects of compensatory evolution, our knowledge is rather limited, not least because it demands detailed exploration of the underlying molecular mechanisms of compensation. Compensatory mutations may enhance fitness either by reducing the need for the gene with the compromised function, or by restoring the efficiency of the affected molecular function [18]. For compensation of fitness costs of antibiotic resistance conferring mutations, restoration of function was the most common mechanism [18], but in other systems the relative importance of functional substitution and restoration is unknown. In the case of functional restoration (e.g., by enhanced dosage of a redundant duplicate of the disrupted gene), one might expect limited pleiotropic fitness effects of compensatory mutations across environmental conditions.

Compensatory evolution following gene loss is of special interest [17]. Gene loss may be initiated by genetic drift and/or selection through antagonistic pleiotropy [17,19]. As reversion to the wild-type state is less likely, gene loss may promote genetic changes that drive the populations to new adaptive peaks (Figure 1). It's reasonable to assume that compensatory mutations are generally specific to the gene defect, and multiple molecular mechanisms can restore fitness. Therefore, independently evolving populations carrying an inactivated gene are expected to diverge from each other. Moreover, if compensation mainly proceeds by reducing the need for the disrupted molecular function then compensatory evolution could have a large impact on cellular physiology and survival upon environmental change. Accordingly, the beneficial effects of compensatory mutations may frequently be conditional, and subsequent changes to the environment can reveal the hidden genetic variation across populations (Figure 1). The goal of the current study was to test this hypothesis by an integrated systems biology approach. Specifically, we aimed to determine the potential of the *Saccharomyces cerevisiae* genome to compensate for gene loss through compensatory evolution and to explore the long-term consequences of this process.

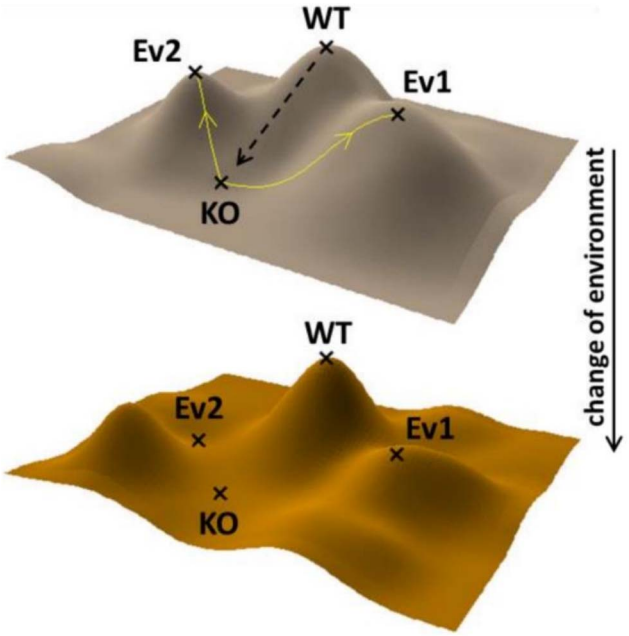


Figure 1. Compensatory evolution on the adaptive landscape. Schematic representation of the impact of compensatory evolution on the fitness landscape. The x and y axes on the landscape locate the network of neighboring genotypic states, while the z axis defines fitness in a single environment. Gene loss leads to a fitness valley (from WT to KO), while compensatory evolution can drive the population to different adaptive peaks (Ev1 versus Ev2). The upper fitness landscape shows the environment where compensatory evolution took place. The dashed arrow represents the original gene deletion event. Yellow lines represent different evolutionary routes. WT, wild type; KO, ancestor strain with a gene deletion.
doi:10.1371/journal.pbio.1001935.g001

Results

Rapid Compensatory Evolution Following Gene Loss Is Common

We initiated laboratory evolutionary experiments with 187 haploid single gene knock-out mutant strains, all of which initially showed slow (but non-zero) growth compared to the wild-type control in a standard laboratory medium (Figure 2A, for selection criteria, see Materials and Methods). These genes cover a wide range of molecular processes and functions (Table S1). Populations were cultivated in parallel (four replicate populations for each null mutation), resulting in 748 independently evolving lines. 0.5% of each culture was diluted into fresh medium every 48 hours, and populations were propagated for approximately 400 generations. To control for potential adaptation unrelated to compensatory evolution, we also established 22 populations starting from the isogenic wild-type genotype, referred to as evolving wild types. Next, all starting and evolved populations were subjected to high-throughput fitness measurements by monitoring growth rates in liquid cultures.

Fitness may increase during the course of laboratory evolution as a result of general adaptation to the environment and/or accumulation of compensatory mutations that suppress the

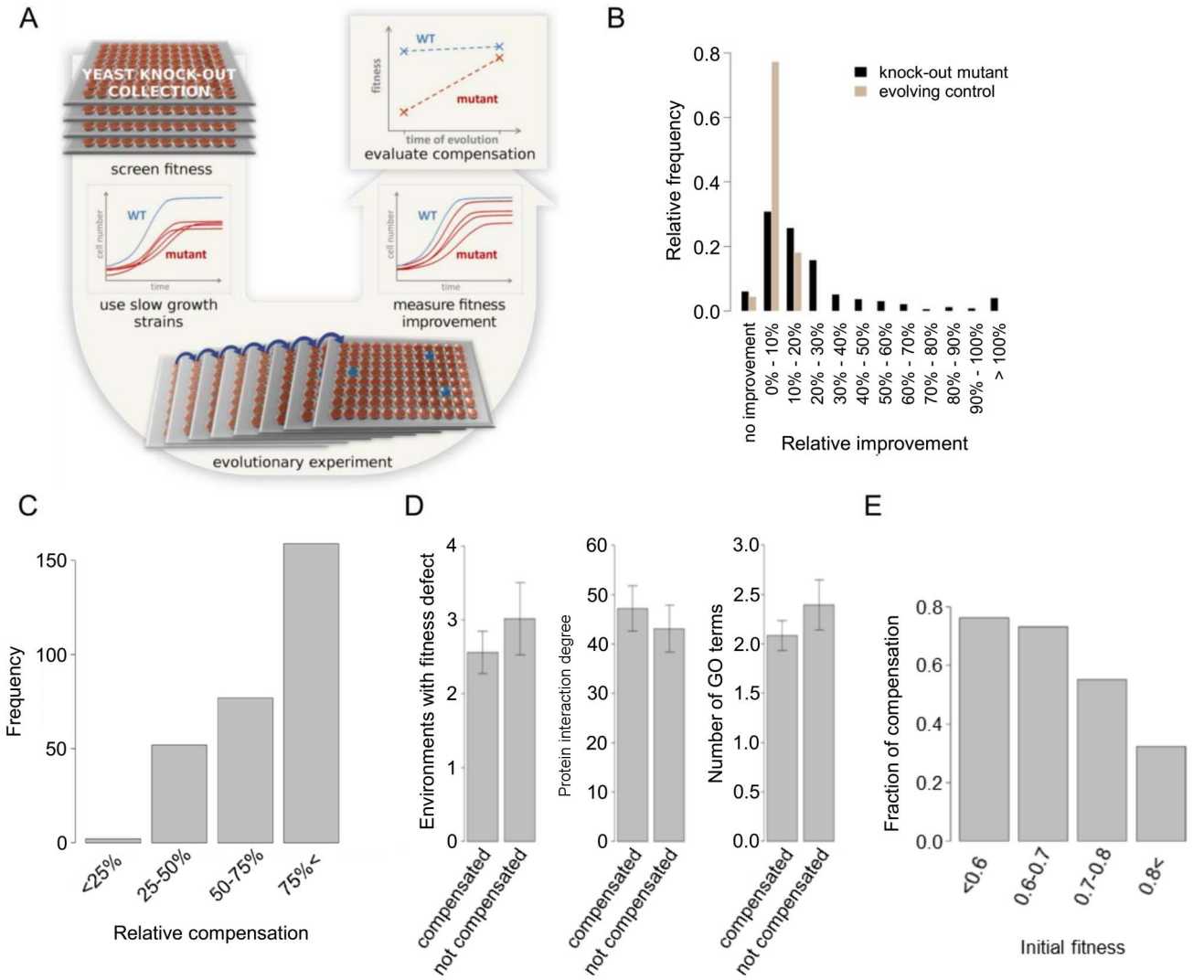


Figure 2. Compensation of fitness loss during laboratory evolution. (A) Experimental scheme to estimate evolutionary compensation of gene defects. See text for details. (B) Distribution of relative fitness improvement (RFI) of the knock-out mutant strains and the evolving control lineages (Table S1), where RFI = (evolved fitness/initial fitness)−1. (C) Relative compensation (RC) of the compensated knock-out mutant strains (Table S1), where RC is the fraction of the initial fitness defect that was compensated for during laboratory evolution (see Materials and Methods). (D) Compensation does not depend on pleiotropy (Table S1). The bars indicate mean ± standard error, Wilcoxon rank sum test *p*-values for the three comparisons are: 0.71, 0.44, and 0.36, respectively. (E) Genotypes with lower initial fitness were more likely to be compensated for during laboratory evolution (Table S1). Lines were divided into groups by initial fitness, the fraction of compensated lines among all the lines in the group is shown as bars (chi-squared test for trend in proportions, $p<10^{-13}$, number of lines in the groups from left to right: 38, 56, 201, 337).
doi:10.1371/journal.pbio.1001935.g002

deleterious effects of gene inactivation. Under the assumption that compensatory evolution was the dominant force in our experiments, fitness should not increase by the same extent in all lineages: genotypes that carry deleterious null mutations are further away from the optimal state and are hence expected to show large fitness gains (Figure 2A); this was indeed so. On average, the evolving wild-type control populations showed a small, but significant 5% fitness improvement. By contrast, the fitness of populations carrying a deleterious null mutation improved by 23% on average (Figure 2B), and many of them approximated wild-type fitness (Figure 2C; Table S1). On the basis of fitness measurements at multiple time points during laboratory evolution (see Methods), we also report that individual fitness trajectories often showed a saturating trend during the course of laboratory evolution (Figure S1).

The difference in fitness improvement is not due to the elevated mutation rate of mutant genotypes for two reasons. First, a previous study conducted a genome-wide screen with the aim to identify genes in *S. cerevisiae* that influence the rate of mutations [20]. While a large number of such genes have been found, only four of them were present in our gene set (*Δrad54*, *Δrad52*, *Δmre11*, and *Δrad50*). Second, fitness improvements of the corresponding single gene knock-out strains did not differ from the rest of the dataset (one-tailed Wilcoxon rank sum test, $p = 0.89$).

As previously [16], we defined compensatory evolution as a fitness increase that is disproportionally large relative to that in the evolving wild-type lines. Using this definition, 68% of the genotypes showed evidence of compensatory evolution (i.e., at least one of the four independently evolving populations fulfilled

the above criteria). The corresponding genes cover a wide range of molecular and cellular processes (Table S1).

Impact of Gene Pleiotropy and Dispensability on the Propensity for Compensation

Next, we compared the fitness improvements between evolved lines founded from the same gene deletion genotype versus those founded from different genotypes. This analysis revealed that not all genes were equally likely to be compensated as fitness gain differed significantly across genotypes (ANOVA, $F(186) = 3.9$, $p < 10^{-14}$) (see also Figure S2). It has been previously suggested that as mutations with especially large fitness effects tend to disrupt a broader range of molecular processes [21], such mutations may influence the number of mutational targets where compensatory evolution can occur [13]. We compiled three datasets that estimate different aspects of gene pleiotropy [22], including fitness under diverse environmental conditions (environmental pleiotropy), the number of protein-protein interactions (network pleiotropy), and the number of biological processes associated with a gene (multifunctionality). The extent of evolutionary compensation did not depend on any of the above mentioned features (Figure 2D). However, consistent with results of prior small-scale bacterial and viral evolutionary studies [13,16], null mutations with more severe defects were more likely to be compensated (Figure 2E). This pattern probably reflects that the availability of compensatory mutations across the genome strongly depends on the fitness effect of the deleted gene. We provide a simple explanation of this phenomenon in the Discussion.

Compensatory Evolution Promotes Genomic Diversification

To investigate the genomic changes underlying compensatory evolution, we re-sequenced the complete genomes of 41 independently evolved lines and the 14 corresponding ancestors, all of which showed large fitness improvements (Table S1). We focused on *de novo* mutations that accumulated during the course of laboratory evolution. Large-scale duplications (including segmental or whole chromosome duplication) were observed in 22% of the laboratory evolved lines. On average, six point mutations and 0.5 small insertions or deletions per clone were detected (Figure 3A; Table S2). The ratio of non-synonymous to synonymous mutations was significantly higher than expected by chance ($p = 0.003$, see Materials and Methods), indicating that the accumulation of these mutations was driven by adaptive evolution. On average, pairs of evolutionary lines founded from the same genotype shared 5.3% of their mutated genes, while the same figure was 0.1% for lines founded from different genotypes (Table S2). This result is in contrast to results of a prior bacterial study [23], where a strong signature of parallel evolution emerged at the gene level across parallel evolving laboratory populations. Despite the rarity of parallel evolution at the molecular level, a major unifying trend emerged: evolution preferentially affected genes that are functionally related to that of the disrupted gene (Figure 3B). Moreover, when the null mutation affected a protein complex subunit, another subunit of the same complex was mutated 10 times more often than expected by chance (Figure 3B). Taken together, these results indicate that deletion of any single gene drives adaptive genetic changes specific to the functional defect incurred.

Pre-Existing Genetic Redundancy Has No Major Impact on Compensatory Evolution

Although duplicated genes with partially overlapping function are frequent in the yeast genome, we found no evidence that genetic changes affecting a duplicate of the disrupted gene provide a general mechanism of compensation in our evolved lines. First, our dataset contains 128 genes showing evidence for compensation, and only 25% of these genes have a duplicate in the yeast genome (i.e., at least 30% amino acid similarity between the two copies). This figure is a gross overestimate, as it includes very distant duplicates that most likely diverged functionally (Materials and Methods). Second, the subset of genes with a gene duplicate were not more likely to be compensated during laboratory evolution than the rest of the dataset (Chi-squared test, $p = 0.54$). Third, genome sequence analysis of the evolved lines revealed only one clear example where evolution proceeded through increasing the dosage of a gene duplicate with redundant function of the deleted gene (Figure 3C). All three studied evolved lines of *Δrpl6b* showed an increased copy number of the left arm of Chromosome XIII (Figure 3C). *RPL6B* is a non-essential gene and encodes a ribosomal 60S subunit protein L6B. The duplicated genomic regions of *Δrpl6b* evolved lines carry *RPL6A*, a duplicate copy of *RPL6B*. The two genes share 94% amino acid identity, have highly overlapping functions, and deletion of both genes confer a synthetic lethal phenotype [24]. On the basis of these observations, we propose that doubling the copy number of *RPL6A* through segmental duplication could be partly responsible for the improved fitness in the evolved lines carrying the *RPL6B* deletion. The hypothesis was tested by increasing the copy number of *RPL6A* in wild-type and Δrpl6b genetic backgrounds, respectively. As expected, an enhanced copy number of *RPL6A* substantially improved the fitness of *Δrpl6b*, but not that of the wild type (Figure 3D).

Compensatory Evolution Does Not Restore Wild-Type Genomic Expression State

Compensatory evolution may restore wild-type physiology or generate novel alterations with respect to prior physiological states [25]. To investigate the relative contribution of these processes, eight genotypes carrying a deleterious gene deletion and one corresponding evolved line were selected for transcriptome analysis (see Materials and Methods for selection criteria). Using DNA microarrays, the global gene expression states were compared between the wild-type, the ancestral line, and the evolved lines carrying the same gene deletion (Figure 4A and 4B). As expected from prior studies [26], inactivation of genes with high fitness contribution altered the expression of a large number of genes across the genome (ranging between 81 to 588) (see Table S3). Next, the transcriptomic profiles were compared by calculating all pairwise combinations of Euclidean distances. The wild-type, the ancestral line, and the corresponding evolved lines generally showed substantial differences in their transcriptome profiles (Figure 4B), indicating that compensatory evolution drives the cell towards novel genomic expression states. Importantly, transcriptome profile distances between different genotypes was always higher than distances between replicate measurements of the same genotype (Figure 4B), implying that the substantial differences observed between evolved lines and wild-type cannot be attributed to measurement noise. As a further support, typically only 10%–30% of the genes with altered expression in the ancestral lines showed significant shift towards the wild-type expression level in the corresponding evolved lines (Figure 4C). Hence, despite substantial fitness improvements (>75% for all cases investigated), the majority of the gene expression changes due to gene deletion remained unrestored during evolution. These patterns were not attributable to growth rate regulated gene expression or copy number variation in the evolved lines (Figure S3).

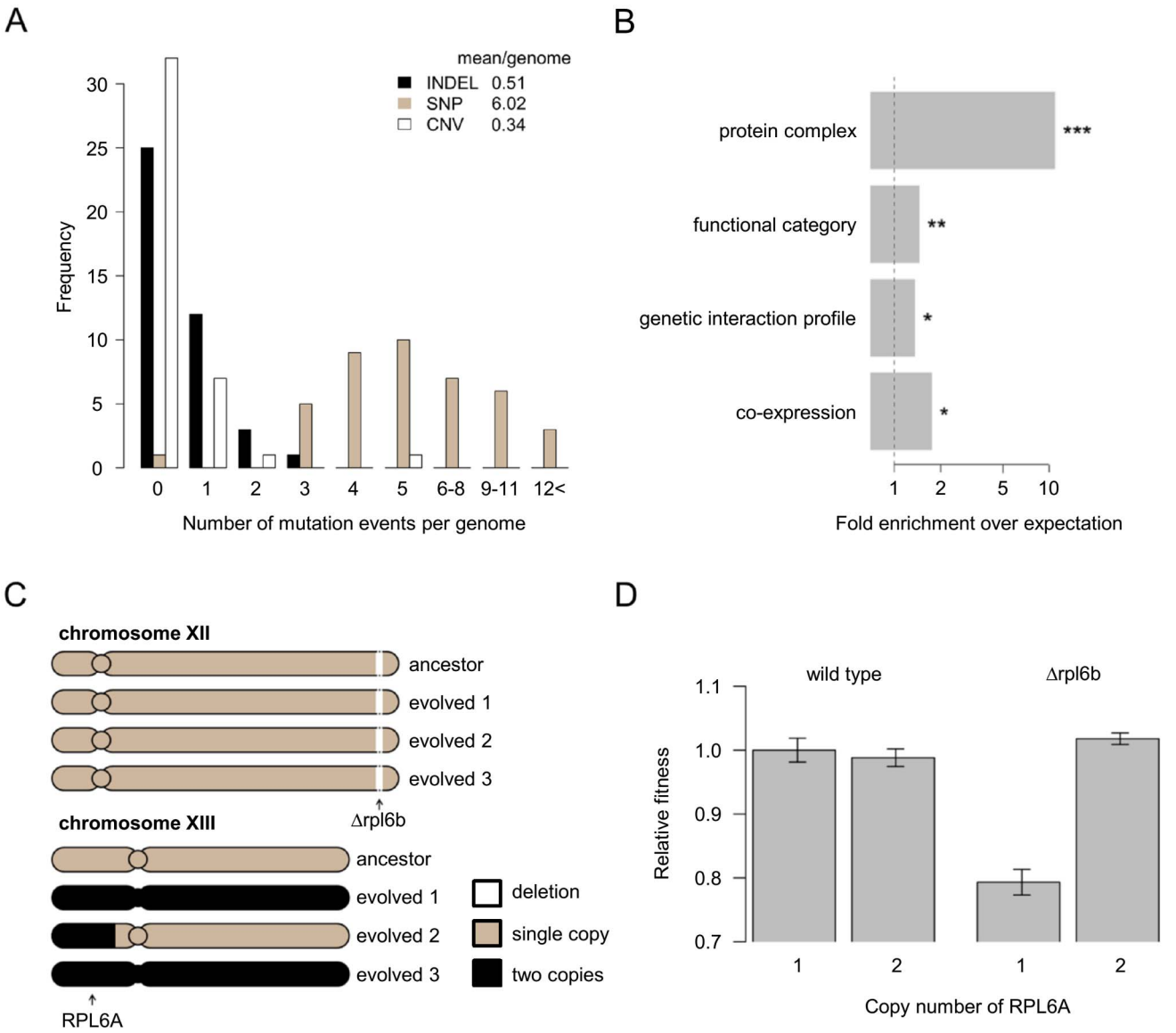


Figure 3. Genomic analyses of evolutionary compensation. (A) Distribution of different mutational events (Table S2). The inlet shows the color coding and the average value of total mutational events per genotype. (B) The originally deleted gene and the gene with identified *de novo* mutation participated more often in the same protein complex, were more often assigned to the same functional category and showed significantly more similar genetic interaction and expression profile than expected by random shuffling of the knock-out gene–mutated gene network. Dashed line represents no enrichment; */**/*** indicates *p*-value$<$0.05/0.01/0.001, respectively. The x axis is logarithmically scaled. (C) Δ*rpl6b* evolved lines showed duplication of the chromosomal region (or the complete chromosome) carrying a duplicate with redundant function (*RPL6A*). The gene positions are marked by arrows below the corresponding chromosome, copy numbers are shown by color codes. (D) Dosage compensation of Δrpl6b by increased copy number of RPL6A (Table S5). Copy number of *RPL6A* was increased by transforming the *RPL6A* bearing plasmid of the MoBY ORF Library. As the vector carries a selectable marker and a yeast centromere, the plasmid is present in one to three copies per cell. As a control, strains were transformed with the empty centromeric plasmid. Relative fitness was measured as colony sizes on agar plates, values were normalized to the wild-type control with a single genomic copy of RPL6A. All strains were grown on synthetic complete medium without uracil to select for the plasmids. Error bars show standard error.
doi:10.1371/journal.pbio.1001935.g003

Compensatory Evolution Generates Diverse Growth Phenotypes across Environments

Taken together, compensatory evolution following gene loss did not restore wild-type genomic expression and promoted genomic divergence across populations. Are these evolutionary outcomes phenotypically completely equivalent? This problem was first addressed by monitoring the fitness of 237 evolved populations in 14 environmental settings, including previously tested nutrients and stress factors [27]. Prior to evolution, genotypes carrying a gene deletion generally displayed slow growth in most environments (Table S1). The situation was far more complex following laboratory evolution. Considering all possible pairs of population-environment combinations, fitness improved in 52%, and declined in 8% of the cases (Figure 5A). Moreover, independently evolved populations carrying the same disrupted gene showed more fitness variation across the 14 tested conditions than in the environment they had been exposed to during laboratory evolution (Figure 5B, $p<10^{-7}$), while evolved wild-type populations did not show such a

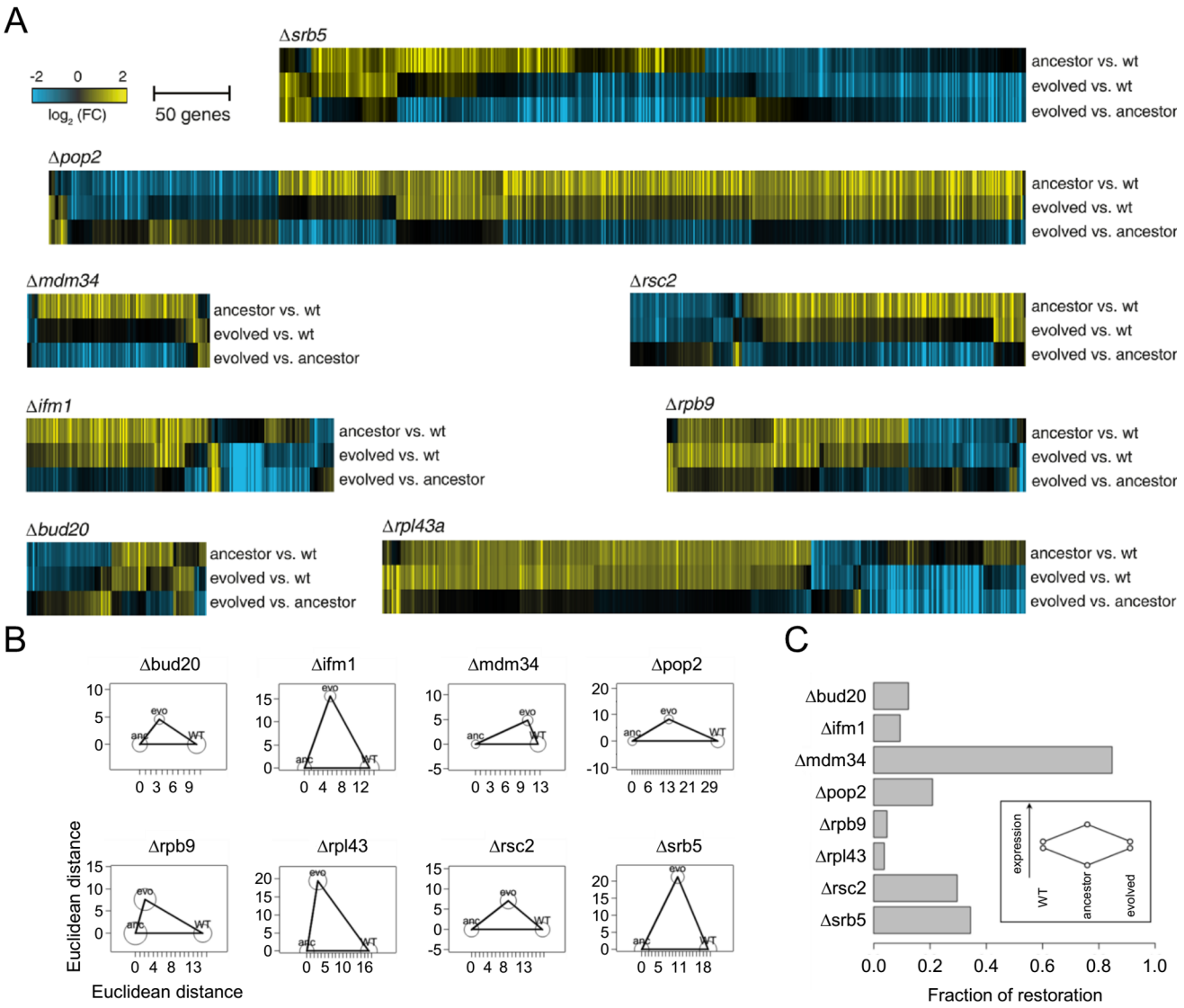


Figure 4. Comparisons of the transcriptome profiles of wild-type, ancestor, and evolved lines. (A) Heatmaps of transcriptome profiles of deletion mutants *Δrpl43a*, *Δpop2*, *Δmdm34*, *Δrsc2*, *Δifm1*, *Δrpb9*, and *Δbud20* and their corresponding evolved lines. For each deletion mutant, the fold-changes (FC) are shown for the ancestor strain versus the wild type, the evolved strain versus the wild type and the evolved strain versus the ancestor strain (Table S3). Color scales as indicated. Individual transcripts are depicted if they change significantly (FC>1.7, $p<0.05$) at least once in one of these comparisons. (B) The Euclidean distances of microarray profiles of the evolved evolutionary line from its ancestor and from wild type (WT) were calculated and normalized to the ancestor–wild type distance for each genotype. The distances of the points in the figure are proportional to the calculated profile distances. For each genotype triplet, distances were calculated on the basis of those genes that are differentially expressed in at least one of the pairwise comparisons. For each deletion strain, the edges of the triangle represent Euclidean distances of $\log_2$ mRNA expression fold-changes between the wild-type (WT), ancestor (anc), and evolved (evo) lines. To calculate these distances we used the average of four replicate expression measurements (two biological and two technical replicates). Circles around average values represent the Euclidean distance between the two biological replicates (calculated as the average based on the two technical replicates). For each genotype triplet, distances were calculated on the basis of those genes that are differentially expressed (FC>1.7, $p<0.05$) in at least one of the pairwise comparisons (Table S6). (C) Within the subset of genes that showed expression change upon gene deletion, the barplot shows the fraction of these genes that changed expression during evolution in the opposite direction (i.e., evolution towards restoration of wild-type expression level; see inset). With one major exception (lines disrupted in *mdm34*), only a small fraction of the expression changes were restored in the evolved lines (Table S6). The threshold for expression change was 1.7-fold-change and $p<0.05$, as in [62].
doi:10.1371/journal.pbio.1001935.g004

difference ($p = 0.93$, coefficient of variations compared by Z-test). Furthermore, the degree of fitness variation across conditions was especially high for gene deletions that showed large fitness gains during compensatory evolution (Spearman rho = 0.36, $p = 10^{-4}$) (Figure 5C). These results indicate that the level of discernible heterogeneity in fitness was relatively low in the evolved populations founded from the same genotype, but the variation can be uncovered upon environmental change.

Finally, our analysis revealed a few instances where the laboratory evolved lines displayed significantly higher than wild-type fitness in specific environments (Table S1). Most notably, the evolved *Δrpl6b* and *Δatp11* lines displayed 24%–26% fitness increase compared to that of the wild type in a medium containing sodium chloride (Table S1), a result that was confirmed by additional independent colony size assays with high replicate number ($n = 20$, Wilcoxon rank-sum test $p<10^{-4}$ in all cases).

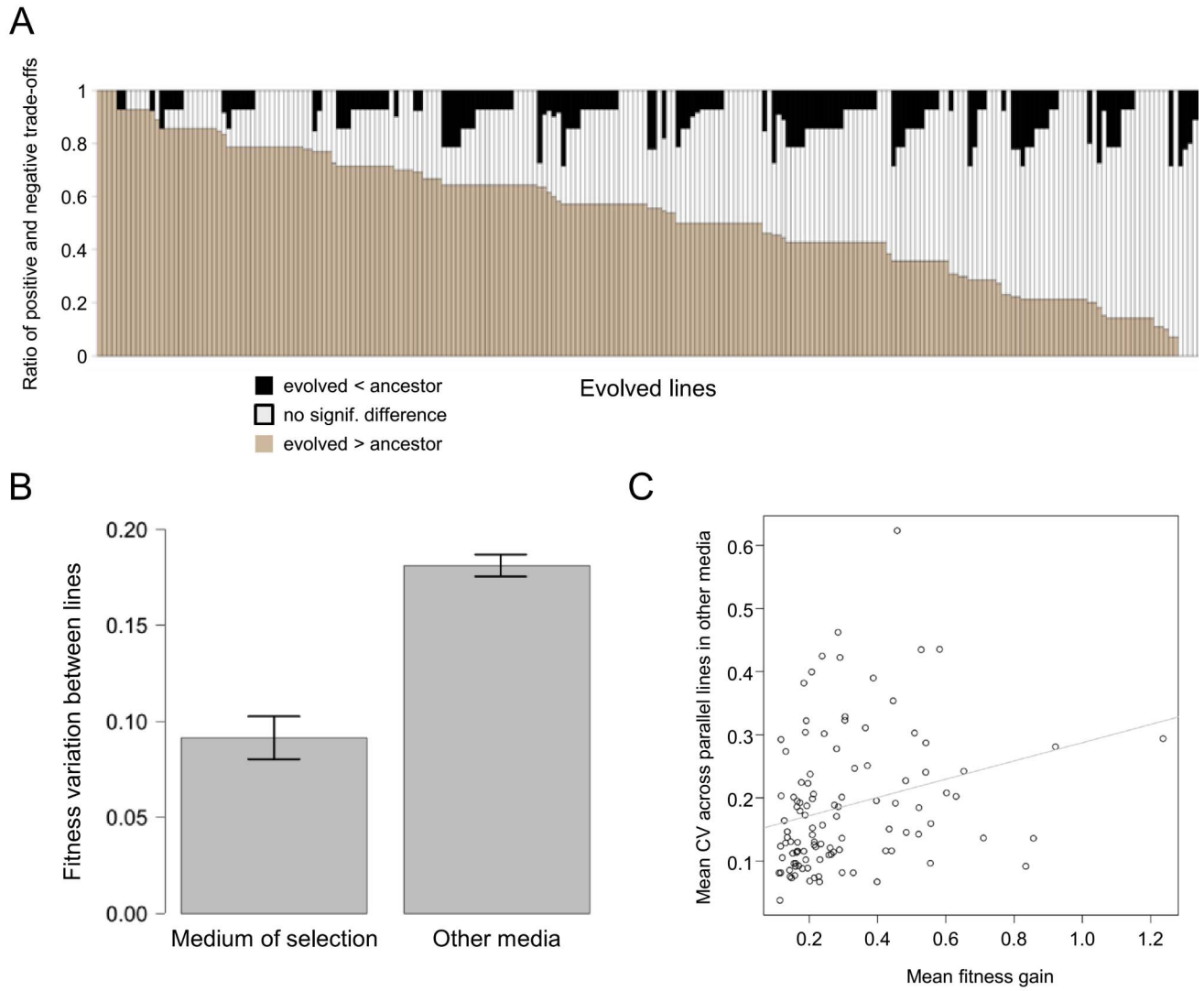


Figure 5. Large-scale phenotypic screen of evolved lines. (A) Fitness trade-offs in evolved lines carrying a deletion across 14 environments (Table S1). Lines are ranked according to the number of environments in which they display improved fitness (brown). Grey and black dots indicate conditions where the fitness of the line is statistically equal or lower, respectively, than that of the corresponding ancestor. (B) Fitness variation in independently evolving lines carrying the same gene deletion. The figure shows the coefficient of variation in the in the medium of selection (YPD) versus all other media (Table S1). The difference is highly significant (Wilcoxon rank sum test p-value$<10^{-7}$). The bars indicate mean of the coefficients of variations $\pm$ standard error. (C) Gene deletions showing larger fitness gains have higher variance of fitness between replicate lines across other environments (Spearman rank correlation, rho = 0.36, $p = 0.0001$). Each point represents a gene deletion genotype. The x-axis shows the mean of the fitness gains of the parallel evolving replicates of a given gene deletion, while the y-axis shows the mean of the coefficient of variations measured in each alternative media between the parallel evolving replicates after 104 days of lab evolution (Table S1). The gray line indicates fit by linear regression.
doi:10.1371/journal.pbio.1001935.g005

Moreover, the fitnesses of these lines in this medium surpassed all that of the 22 evolved wild-type controls. These results are all the more remarkable, as the corresponding ancestral *Δrpl6b* and *Δatp11* strains showed fitness values significantly lower than wild type under all environmental conditions considered. These preliminary results indicate that gene loss can promote adaptive evolution towards novel environments, a possibility that will be explored further in a future work.

A Case Study Reveals the Fitness Cost and Condition Dependence of Compensatory Evolution

Next, we conducted an in-depth genetic analysis with the *MDM34* deletion with the aim of deciphering the molecular mechanisms and/or potential fitness costs of compensatory mutations (Text S1). This gene codes for a component of the ERMES protein complex, and is involved in the exchange of phospholipids between mitochondria and the endoplasmatic reticulum (Figure 6A). Disruption of this gene yields impaired cardiolipin synthesis [28], as an insufficient amount of unsaturated fatty acids reaches the mitochondria (Figure 6A). Laboratory-evolved lines carrying deletion in this gene substantially improved fitness in the medium of selection (Table S1), but the putative cellular mechanisms of compensation were remarkably different across populations (Figures 6A and S4). The native copy of *MDM34* was reinserted into the ancestral line and four evolved lines carrying the same deletion (*Δmdm34*). The analysis revealed

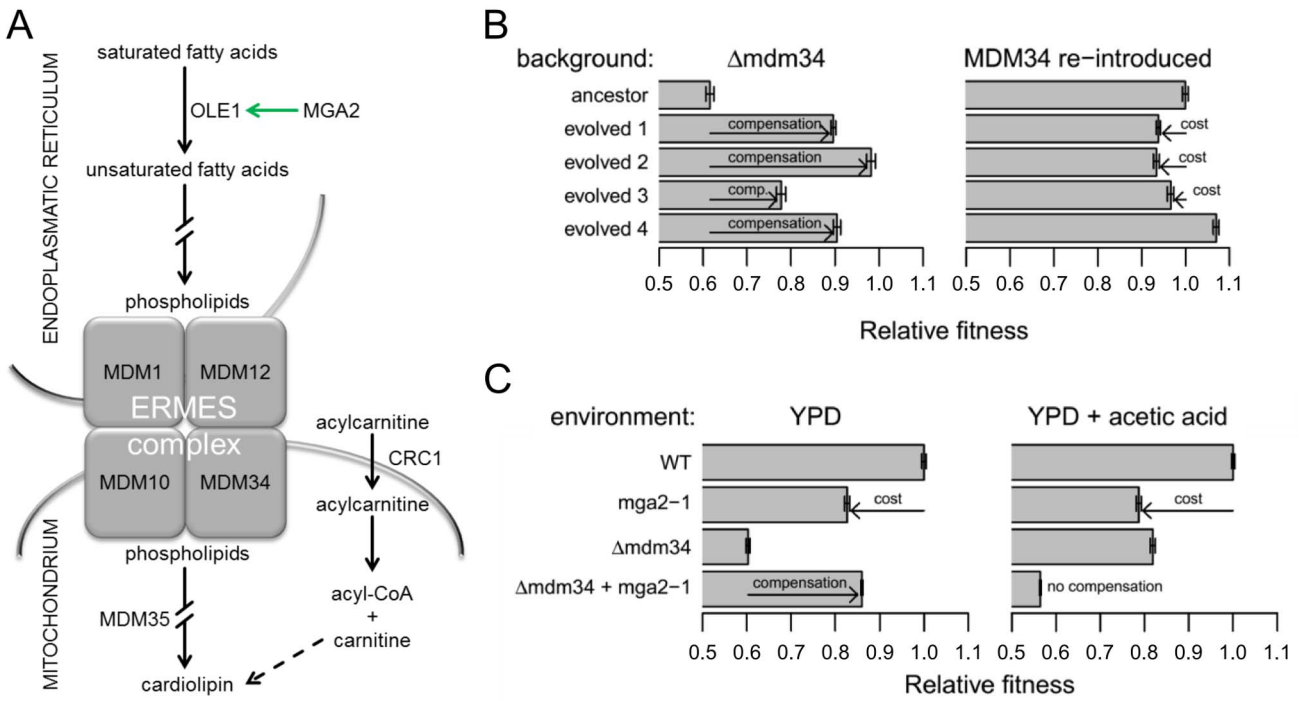


Figure 6. Compensation of the *MDM34* gene deletion. (A) The cardiolipin synthesis pathway with an emphasis on the ERMES complex. The complex tethers the endoplasmatic reticulum to the mitochondria, and is central for the transfer of phospholipids between the two compartments. *De novo* mutations in the independent evolutionary lines affected different, but related cellular subsystems, including upregulation of the unsaturated fatty acid synthesis (*MGA2*), another step of the cardiolipin synthesis pathway downstream of the ERMES complex (*MDM35*), and another mitochondrial transport process (*CRC1*), which most likely affects respiration by modulating the interaction between carnitine and cardiolipin. For further details on the underlying mechanisms see Text S1. The green arrow represents transcriptional upregulation; the dashed arrow indicates indirect positive effect. The mutations in *MGA2*, *MDM35*, and *CRC1* genes were found in Δ*mdm34* evolved lines 1, 3, and 4, respectively. (B) The cumulative fitness effects of the compensatory mutations in Δ*mdm34* and "wild type" (Δ*mdm34*+*MDM34* reintroduced) backgrounds (Table S7). (C) Epistatic interactions between mutations in two environments (Table S7). The bars in (B) and (C) indicate means ± standard error. Arrows indicate fitness costs and the extent of compensation.
doi:10.1371/journal.pbio.1001935.g006

that the net effect of mutations in three evolved lines were deleterious in the presence of *MDM34* (Figure 6B). Next, we concentrated on a specific mutation observed in *MGA2*, a gene involved in the regulation of unsaturated fatty acid biosynthesis (Figure 6A; Text S1). Inserting the observed mutations (*mga2-1*) into wild type and Δ*mdm34* resulted in very similar conclusions. *mga2-1* and Δ*mdm34* showed strong sign-epistasis [29]: they were independently deleterious but significantly less so when they occurred together (Figure 6C). Moreover, the capacity of *mga2-1* to compensate the loss of *MDM34* was restricted to non-acidic conditions (Figure 6C), probably because of the misregulation of the corresponding stress-induced pathway under low pH (Text S1).

Evolutionary Compensation by Loss-of-Function Mutation

Our dataset contains 21 independent point mutations that occurred during laboratory evolution and generated in-frame stop codons. Most notably, a mutation in *WHI2* emerged in an evolving *Δrpb9* line, which shortened the coding region from 480 to 133 codons, and hence most likely resulted in a non-functional protein.

To test the impact of loss of *WHI2* function on fitness and compensation, *Δwhi2* was introduced into *Δrpb9* cells using synthetic genetic array methodology (Figure 7A and 7B) [30]. In agreement with expectation, deletion of *WHI2* partly suppressed the harmful effect of the *RPB9* deletion (Figure 7B). *RPB9* is an RNA polymerase II subunit, and its deletion leads to elevated transcriptional error rate [31] and in turn, to proteotoxic stress [32], which can result in cell cycle arrest [33]. *WHI2* is known to be required for general stress response [34] and cell cycle arrest [35]. We speculate that less stringent cell cycle control due to *WHI2* deletion is favorable in *Δrpb9* (see also [36]).

Next, the fitness impact of *WHI2* deletion was evaluated across 14 environments. The fitnesses of the *Δrpb9 Δwhi2* strain varied strongly across conditions, and showed correlation with that of the evolved *Δrpb9* line, which carried the *WHI2* non-sense mutations (Spearman rho = 0.77, $p<0.005$) (see Figure 7C). Most notably, the compensation of *Δrpb9* by *Δwhi2* was completely abolished in the presence of cycloheximide (Figure 7B). We conclude that the compensatory effect of *WHI2* deletion is plastic across environments.

Discussion

Our work addresses one of the most long-standing debates in evolution. Since the early 1920s, Ronald Fisher pioneered the view that adaptation is by and large a hill climbing process: it proceeds through progressive accumulation of beneficial mutations [37,38]. However, as slightly deleterious mutations are far more abundant, they have a significant contribution to genetic variation in natural populations [2]. In the long run, the wealth of such detrimental mutations is expected to promote fixation of compensatory

A

MATa parent	MATalpha parent	offspring designation
Δhis3::KanMX4	Δho::NatMX4	WT
Δrpb9::KanMX4	Δho::NatMX4	Δrpb9
Δhis3::KanMX4	Δwhi2::NatMX4	Δwhi2
Δrpb9::KanMX4	Δwhi2::NatMX4	Δrpb9 Δwhi2

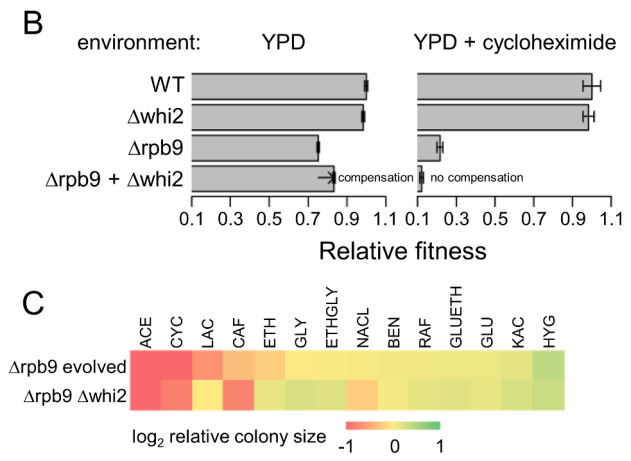


Figure 7. Environment-dependent compensation by a loss-of-function mutation. (A) *Δrpb9* and *Δwhi2* mutations were crossed by SGA using haploid parental strains as shown. To compare the double mutant *Δrpb9 Δwhi2* with the wild-type control and corresponding single mutants, the resistance cassettes required by the SGA method were introduced into wild-type and single mutants by crossing them with parental strains where the corresponding resistance cassettes reside at a non-functional locus (*Δhis3*::KanMX4 and *Δho*::NatMX4). (B) Relative fitness was measured as colony sizes on YPD and YPD supplemented with cycloheximide (CYC), values were normalized to WT. The arrow shows the extent of compensation of *Δrpb9* by *Δwhi2* on glucose medium (Wilcoxon rank sum test $p = 0.005$, error bars show standard error) (Table S8). (C) Relative fitness of *Δrpb9* replicate evolving line 2 and *Δrpb9 Δwhi2* double mutant were measured as colony sizes grown on different media. Genotypes are indicated on the left, the growth media are indicated above the heat map. For media composition and abbreviations, see Table S4. Values are normalized to *Δrpb9* ancestor. Log_2 values are shown according to the color coding (Table S8).
doi:10.1371/journal.pbio.1001935.g007

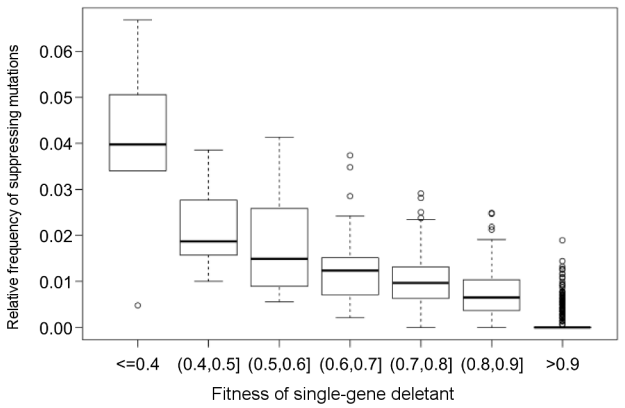


Figure 8. Strongly deleterious gene deletions can be suppressed by a large number of other null mutations according to a genome-wide genetic interaction study. The plot shows the relationship between the fitness of a given single-gene deletion strain and the fraction of other genes across the genome whose deletion suppresses the fitness effect of this mutation (Table S9). Boxplots present the median and first and third quartiles, with whiskers showing either the maximum (minimum) value or 1.5 times the interquartile range of the data. Spearman correlation on raw data: $\text{rho} = -0.69$, $p < 10^{-16}$, $n = 3880$. We note that using the fraction of suppressive interactions among all genetic interactions displayed by a given gene yields a very similar result ($\text{rho} = -0.69$, $p < 10^{-16}$), indicating that the relationship is not simply due to the fact that slow-growing strains generally display especially large numbers of both positive and negative interactions [21]. Information on suppression genetic interactions and single-deletion fitness comes from a global genetic interaction map of yeast [21]. Suppression interactions were defined as in previously [70]. In brief, deletion of gene *B* suppresses deletion of gene *A* if their fitness values obey the following rules: $F_A < F_B$ and $F_{AB} > F_A + \sigma_A$ (where F_A, F_B, and F_{AB} are the fitness measures of single deletants *A*, *B*, and the double deletant *AB*, respectively, and σ_A is the standard deviation of F_A). One important caveat is that as this simple analysis considers null mutations only, the results should be considered preliminary.
doi:10.1371/journal.pbio.1001935.g008

mutations elsewhere in the genome. This work focused on a specific aspect of this problem, and asked whether deleterious gene loss events promote adaptive genetic changes and what the side consequences of such a process might be. To systematically study compensatory evolution following gene loss, we initiated laboratory evolutionary experiments with over 180 haploid yeast genotypes, all of which initially displayed slow growth owing to the deletion of a single gene, and investigated the genomic and phenotypic capacities of the evolved lines in detail. Thanks to the exceptionally large-scale analysis of our study, the following major conclusions can be drawn.

First, compensatory evolution following gene loss was pervasive: 68% of the deleterious, but non-lethal gene disruptions were compensated through the accumulation of adaptive mutations elsewhere in the genome (Figure 2B). Furthermore, in agreement with prior bacterial studies [16,17], the process was strikingly rapid. As the set of disrupted genes are functionally very diverse (Table S1), it appears that defects in a broad range of molecular processes can readily be compensated during evolution.However, we and others [17] also found that not all genotypes are equally likely to be recovered during laboratory evolution. Therefore, future works should clarify the exact molecular, functional, and systems level gene properties that influence compensability. Second, our large-scale study indicates that the extent of fitness loss due to gene disruption is one if not the strongest predictor of compensatory evolution (Figure 2E). Although this relationship has been observed previously in small-scale studies [16], the reasons remained largely unknown. One may argue that the spread of compensatory mutations with mild beneficial effects would have taken many more than 400 generations to reach fixation [16]. Although this explanation cannot be excluded, there is another intriguing possibility [13]. Consistent with Fisher's geometric model [37,38], fitness improvement in populations close to an optimal state can only be achieved by relatively rare mutations with small effects. However, when a population with a gene defect is further away from a fitness peak, compensatory evolution may proceed through a wider range of mutations, including ones that have deleterious side effects. Two lines of evidence are consistent with this scenario. Compensatory evolution has associated pleiotropic effects (Figures 5 and 6C). Moreover, the theory predicts that compensatory mutations should be especially frequent in the case of strongly deleterious null mutations. An analysis based on data of a prior genome-wide genetic interaction study [21] suggests that it may indeed be so (Figure 8).

Third, genomic analysis of the evolved lines revealed that deletion of any single gene drives adaptive genetic changes specific to the functional defect incurred (Figure 3B), and consequently convergent evolution at the molecular level was extremely rare. In agreement with a prior bacterial evolutionary study [17], we found that gene duplication has only a minor role during compensatory evolution following gene loss. A more general issue is the extent to which mutations that affect gene expression could alone recover fitness [17,39]. Although genetic changes in putative promoter regions were not overrepresented in our dataset (Binomial test, $p = 0.87$), 21 observed point mutations generated in-frame stop codons, most likely yielding proteins with compromised or no activities (see also Figure 7). These results indicate that fitness recovery following gene loss can partly be achieved purely through inactivation of other genes.

Fourth, compensatory evolution promoted divergence of genomic diversification, and shifted the evolved population towards novel genomic expression states (Figure 4B). Despite substantial fitness improvements, the majority of the gene expression changes due to gene deletion remained unrestored during evolution. This finding is consistent with prior works arguing that no clear relationship exists between the change in mRNA expression of a gene and its requirement for growth in the same condition [40].

Fifth, independently evolved populations showed substantial fitness variation across environments that they had not been exposed to during laboratory evolution (Figure 5). These results suggest that accumulation of adaptive mutations during compensatory evolution generated substantial genetic differences between populations, and this variation can be uncovered upon environmental change.

Taken together, several lines of evidence indicate that fitness gains in the evolved lines reflect accumulation of gene specific compensatory mutations rather than a global adaptation: (i) evolving wild-type control populations showed only minor changes in fitness, (ii) the rate of adaptation was genotype specific, (ii) convergence at the molecular across genotypes was extremely rare, (iv) evolution preferentially affected genes that are functionally related to that of the disrupted gene, and (v) compensatory mutations had no beneficial impact in a wild-type genetic background.

The above results encouraged us to distinguish between two evolutionary scenarios. Organisms may attempt to restore the disrupted molecular function through mutations in genes with redundant functions (functional restoration). Alternatively, they may aim to minimize the cellular damage incurred by gene disruption (functional replacement). While the possibility of full functional restoration cannot be excluded, the rarity of compensation through mutations in gene duplicates and the plasticity of compensatory mutational effects across environments are consistent with the second scenario. Indeed, our work demonstrates that gene loss promotes genetic changes that have a large impact on evolutionary diversification, genomic expression, and viability upon environmental change. An important implication of our study is that the beneficial effects of compensatory mutations should frequently be conditional, and subsequent changes to the environment can reveal the hidden fitness effects (beneficial and detrimental alike). Lack of restoration of fitness across environments is broadly consistent with the emerging view that epistatic interactions are plastic across conditions [41,42].

The perspective offered in this work leads to the re-formulation of several fundamental questions. First, it sheds light on an evolutionary paradox: while core cellular processes are generally conserved during evolution [43], the constituent genes are partly different across related species with similar lifestyles. We propose that gene content variation across species is partly due to the action of compensatory evolution and may not need to reflect changes in environmental conditions and the consequent passive loss of genes. Although the exact population genetic conditions facilitating this process remain to be elucidated, several observations are consistent with this view. Most notably, the phylogenetic conservation of indispensable genes depends on how easily the gene can be functionally replaced through enhanced expression of other genes [44]. Second, it has been suggested that deleterious mutations may act as stepping stones in adaptive evolution by providing access to fitness peaks that are not otherwise accessible [45,46]. Indeed, our analysis revealed a few instances where the laboratory evolved lines displayed significantly higher than wild-type fitness in specific environments. Finally, given the prevalence of gene loss events during tumorigenesis, future work should elucidate whether similar processes drive the somatic evolution of cancer [47].

Materials and Methods

Yeast Strains and Media

All strains used in this study were derived from the BY4741 *S. cerevisiae* parental strain. Non-essential single-gene deletion strains from the haploid yeast deletion collection [40] (MATa; his3Δ 1; leu2Δ 0; met15Δ 0; ura3Δ 0; xxx::KanMX4) were used to systematically identify all gene disruptions with a significant growth defect. Slow-growing mutants were identified in two steps. An earlier study identified 671 gene deletants in diploid background, which showed a significant fitness defect on both rich and synthetic media [48]. We thus measured fitness of the corresponding MATa haploid strains by recording their growth curves in liquid media. We identified 187 deletants showing at least 10% growth rate defect, which constituted the set of ancestral strains subjected to laboratory evolution (for details of growth measurements see below).

The slow-growing yeast deletants used in this study are listed in Table S1. The evolutionary experiment was conducted using rich liquid medium (YPD, 1% yeast extract, 2% peptone, 2% glucose). Solid media were prepared using 2% agar, which were found to be optimal for reproducible colony size measurement. Details on the media used in the phenotypic profiling experiment can be found in Table S4. Oleic acid and stearic acid was dissolved in DMSO as a 100 mM stock and added to the medium after autoclaving to a final concentration of 0.1 mM.

Laboratory Evolution

Compensatory adaptation refers to fitness gains in a gene deletion strain that are greater than fitness gains occurring in an isogenic wild-type strain. We conducted a series of laboratory evolutionary experiments using four independent populations of each of the 187 slow-growing deletants along with 22 independent lineages of an isogenic wild-type strain (referred to as *evolving* wild types). The *YOR202W* deletion strain was used as evolving wild-type control because the fitness of this strain is indistinguishable from the BY4741 parental wild-type strain [19]. Moreover, this strain carries the KanMX4 cassette in the nonfunctional *his3Δ1* allele, thus it was possible to control for the reported mutation-generating effect of the KanMX4 cassette [36]. All strains were inoculated into randomly selected positions of 96-well plates. Four wells in different positions were not inoculated by cells to help plate identification and orientation. Cells were grown in standard laboratory rich media to minimize selection pressure originating from nutrient limitation. The presence of the KanXM4 cassette

was not selected for during the evolutionary experiment, since G418 was omitted from the medium for two reasons. First, using G418 at 200 mg/l concentration decreases the growth rate of the unevolved wild-type control strain (unpublished data) and might lead to selection for increased resistance. Second, the usage of the drug at a growth-limiting concentration may induce mutagenesis through environmental stress response. To provide optimal growth conditions, plates were covered with sandwich cover (Enzyscreeen BV), shaken at 350 rpm, and incubated at 30°C. Using a handheld replicator, $\sim 10^5$ cells ($\sim$0.5 µl sample volume) were transferred every second day to 100 µl of fresh medium in 96-well plates resulting in $\sim$7.6 generations between transfers. The experiment was run for 104 days ($\sim$400 generations total) and samples from days 0, 26, 52, 78, and 104 were frozen in 15% glycerol and kept at −80°C until fitness measurement. Cross-contamination events were regularly checked by PCR and visual inspection of empty wells (unpublished data).

High-Throughput Fitness Measurements

We used established protocols specifically designed to measure fitness in yeast populations [49]. Growth was assayed by monitoring the optical density (OD_{600}) of liquid cultures of each strain using 384-well microtiter plates containing YPD medium (as during the evolutionary experiments). We used relative growth rate as a proxy for relative fitness (see below). Compared to laborious competition based fitness assays, this protocol allows estimating growth rate on a relatively large scale in an environment that is nearly identical to the one used in the evolutionary experiments.

Growth Curve Recording

Starter cultures were inoculated from frozen samples using 96-well plates. The starter plates were grown for 48 hours under identical conditions to the evolutionary experiment. 384-well plates filled with 60 µl rich medium per well were inoculated for growth curve recording from the starter plates using pintool with 1.58 mm floating pins. The pintool was moved by a Microlab Starlet liquid handling workstation (Hamilton Bonaduz AG) to provide uniform inoculum across all samples. The median blank corrected initial OD_{600} of the wells was 0.027. Each 384-well plate were inoculated with four different starter plates: one plate having the unevolved wild-type control as a reference strain in all wells in order to estimate various within-plate measurement biases, and three plates containing the same set of mutants from three of the five time points of the evolutionary experiment. The 384-well plates were incubated at 30°C in an STX44 (LiCONiC AG) automated incubator with alternating shaking speed every minute between 1,000 rpm and 1,200 rpm. Plates were transferred by a Microlab Swap 420 robotic arm (Hamilton Bonaduz AG) to Powerwave XS2 plate readers (BioTek Instruments Inc) every 20 minutes and cell growth was followed by recording the optical density at 600 nm. Six technical replicate measurements were executed on all strains sampled from each time-point of the evolutionary experiment. Measurements with growth curve irregularities were automatically removed. Only those strains were further analyzed where at least four technical replicate measurements remained after this quality control step.

Growth Curve Analysis

Growth rate was calculated from the obtained growth curves following an established procedure [49,50]. To eliminate potential within-plate effects that might cause measurement bias, growth rates were normalized by the growth rate of neighboring reference wells that contained the wild-type controls. For each strain and each evolutionary time point, relative fitness was calculated as the median of the normalized growth rates of the technical replicates divided by the median growth rate of the wild-type controls. At day 0, the technical replicate measurements of the isogenic independently evolving lines were combined to calculate median ancestral fitness since by that time these populations had no independent evolutionary history. Stringent criteria were used to define the set of ancestor strains with substantial growth rate defect: a minimum of 10% fitness drop was required compared to the wild-type controls; significance was determined by one-tailed Wilcoxon rank sum test, p-value was corrected with a false discovery rate of 0.05.

Identifying Lines Showing a Significant Compensatory Adaptation

To determine whether the fitness defect of a given knock-out strain became compensated during the evolutionary experiment two criteria must have been met: First, the growth rate improvement had to be significant (one-tailed Wilcoxon rank sum test, p-value corrected with a false discovery rate of 0.05). Second, the growth rate increment of the knock-out strain had to be disproportionally larger than that of the evolving wild-type control strains. To test whether fitness gain in a knockout is higher than those occurring in the evolving control lines, we first fitted a normal distribution to the fitness improvement values of the evolving control lines. Next, we defined a fitness improvement cutoff, so that the probability that an evolving control line would show an improvement at least that high is less than 0.05.

To evaluate the extent of evolutionary compensation, a relative compensation index was calculated according to the following formula:

$$\frac{\Delta_{end} - \Delta_{start}}{WT_{end} - \Delta_{start}}$$

where WT and Δ means median normalized growth rate of the evolving wild-type control and the knock-out strain, respectively, measured before (start) and after (end) the evolutionary experiment. Thus, a relative compensation of 1 indicates that the knock-out strain reached the same fitness after evolution as the evolving wild-type control strains. See Table S1 for the whole dataset.

Phenotypic Profiling across Environmental Conditions

To study the pleiotropic effects of compensatory adaptation, we measured the fitnesses of 237 evolved lines carrying a single gene deletion, all evolved wild-type control lines along with the corresponding ancestors across various environmental conditions. As this experiment demands high-throughput analyses (over 14,000 data points), fitness was estimated by colony size on solid agar media. Moreover, it allowed direct comparison of the reliability of our measurements to results of a previous study (Figure S5).

We prepared solid agar media of 14 different compositions to expose the strains to fundamentally diverse environments and to obtain sufficient throughput. Our list of 14 growth media was primarily based on a previous study [27] and included various carbon sources and stress conditions (Table S4). A robotized replicating system was set up for colony size based fitness measurement. The system consists of a Microlab Starlet liquid handling workstation (Hamilton Bonaduz AG) equipped with a pintool with 768 pins (S&P Robotics Inc) and a custom-made pintool sterilization station. Several aspects of the replication procedure had been experimentally customized to achieve uniform, reproducible inoculation of yeast cells.

Fitness of the ancestor (day 0) and evolved strains (day 104) was approximated by measuring colony sizes of ordered arrays of strains at 768 density. First, four different 96-well plates of the evolutionary experiment were scaled up to arrays of 384 colonies: one having the unevolved wild-type control in all positions, and three different plates of the mutant set from the same time point. Then pairs of 384 arrays with corresponding strains from day 0 and 104 were combined to reach 768 density. With this set up, all evolving replicate lines derived from the same ancestral genotype from both day 0 and day 104 were grown on the same 768 plate to exclude potential plate-to-plate variations when comparing colony growth of ancestor and evolved lines. Four technical replicates of these 768 arrays were transferred into each of the 14 different media.

After acclimatization to the media at 30°C for 48 hours the plates were replicated again onto the same type of media and photographed after 48 hours of incubation at 30°C. Digital images were processed to calculate colony sizes, and potential systematic biases in colony growth were eliminated (Text S1). For each growth environment, fitness of each original knock-out genotype at day zero and each independently evolving line at day 104 was determined as the median of the size of replicate colonies. The reliability of our experimental setup and data processing was confirmed by comparing the fitness measurements of ancestral knock-out strains with the published data of Dudley and colleagues (Figure S5) [27].

To determine whether an ancestor genotype shows a significantly altered fitness compared to the wild-type control in a given environment, we used a Wilcoxon rank sum test (with p-value corrected for each condition with a false discovery rate of 0.05). The same statistical test was used to determine whether the fitness of an evolved line is different from that of its ancestor in a given environment. See result in Table S1.

Genome Sequencing

To reveal the underlying molecular mechanisms of compensation, we subjected 41 strains to whole-genome re-sequencing. Our list of sequenced strains primarily included genotypes with large initial fitness defect, substantial fitness improvement and gradual fitness increase over the course of evolution. To be able to detect parallel evolution at the molecular level, we selected two to four independently evolving lines of each ancestor genotype for sequencing. Overall, 41 evolved lines from 14 deletion strains were chosen along with their corresponding ancestor strains. Candidates were re-streaked and single clones were isolated and their fitness increase was confirmed by growth curve recording.

Genomic DNA was prepared using a glass bead lysis protocol: clones were inoculated into 5 ml YPD+G418 (200 mg/l) and grown to saturation at 30°C. Cells were pelleted and resuspended in 500 µl of lyis buffer (1% SDS, 50 mM EDTA, 100 mM Tris [pH 8]). Cells were mechanically disrupted by vortexing for 3 minutes at high speed with 500 µl glass bead (500 µm, acid washed). After adding 275 µl 7 M ammonium acetate, samples were incubated at 65°C for 5 minutes, followed by a second incubation on ice for 5 minutes. The samples were extracted with chloroform:isoamyl alcohol (24:1) and centrifuged for 10 minutes. The aqueous layer was transferred into a new tube and precipitated with 1 ml isopropanol, pelleted and washed with 70% ethanol, and resuspended in 500 µl RNaseA solution (50 ng/ml). After 30 minutes RNaseA treatment at room temperature, samples were chloroform:isoamyl alcohol (24:1) extracted, precipitated with 50 µl sodium acetate (3 M [pH 5.2]) and 1,250 µl ethanol, pelleted and washed with 70% ethanol. Finally, the genomic DNA was dissolved in water. The steps of re-sequencing was done by the UD-GenoMed Medical Genomic Technologies Ltd: amplified genomic shotgun libraries were run on the Illumina HighScan SC with 1×100 bp single read module resulting in an average coverage of about 80×. Reads were aligned to the *S. cerevisiae* EF4 genome assembly using the BWA software package [51] having the genomic repeats masked using RepeatMasking [52]. Variant calling was performed using the GATK software package [53]. Genomic single-nucleotide polymorphisms with less than 200 phred-scaled quality score or lower than 0.3 mutant/reference ratio were ignored. Duplications of large chromosomal segments or whole chromosomes were identified as increased read coverage of certain regions. Elevated read coverage of regions with a minimum of 25 kb length were accepted as duplications if both the Control-FREEC [54] (Wilcoxon rank-sum test, $p<0.01$) and the CNV-seq [55] ($p<0.0001$) software predicted significant alteration from the read coverage of the reference genome.

Our primary aim was to analyze *de novo* mutational events. *De novo* mutations were identified as alterations from the reference genome specifically found in the evolved lines but not present in the ancestral strains. Mutations, which occurred before our evolutionary experiment but after the gene knock-out, are referred to as secondary ancestor mutations. These mutations were identified in the ancestral strains as SNPs and indels present only in the corresponding ancestor strain, not in any other ancestral strains. The rationale behind this consideration is not to classify mutations accumulated in the parental strain of the mutant library prior to the generation of the knock-out strain as a secondary ancestor mutation. The list of identified mutations can be found in Table S2.

Ratio of Non-Synonymous to Synonymous SNPs

Whole-genome re-sequencing revealed that 86% of SNPs in the coding regions were non-synonymous. To statistically test whether the ratio of non-synonymous to synonymous SNPs was higher than expected based on a neutral model of evolution, we employed the method of Barrick and colleagues [56]. Briefly, we took all different point mutations observed in protein coding regions and calculated the probability that 86% or more substitutions would result in a non-synonymous substitution if it occurred in a random coding position. The excess of non-synonymous substitution observed in the evolved genomes was significant ($p = 0.003$).

Datasets Used for Bioinformatic Analysis

To test whether the extent of evolutionary compensation is influenced by the disrupted gene's pleiotropy, we used three complementary measures of gene pleiotropy. Environmental pleiotropy of a non-essential gene was defined as the number of unique conditions in which the removal of the gene resulted in a fitness defect according to Dudley and colleagues [27]. Network pleiotropy was measured as the total number of protein-protein interactions reported in the BioGRID database [57]. Finally, multifunctionality of a gene was calculated on the basis of a set of GO terms considered to be specific by yeast geneticists, as previously described [58].

To investigate whether mutations accumulated during compensatory evolution preferentially affected genes that are functionally related to the disrupted gene, we used different measures of functional relatedness: co-membership within stable protein complexes, shared functional category, genetic interaction profile similarity, co-expression, and paralogy. For protein complexes we used the manually curated dataset based on tandem affinity purification/mass spectrometry studies (YHTP2008) from the Wodak lab [59]. For functional categories, the MIPS Functional Catalogue Database was downloaded [60]. Genetic interaction

profile similarities were obtained from a large-scale genetic interaction screen study [21]. The authors calculated the genetic interaction profile for a given gene deletion genotype as the list of genetic interaction scores detected across all other genes in their dataset. The genetic interaction profile similarity between two genes was defined as the Pearson correlation value of the two genetic interaction profiles [21]. For calculating co-expression data, 247 normalized microarray datasets from the M3D database [61] were used to create an expression profile for each gene. In case of multiple replicates per experiment, the average normalized values were calculated, and employed further. For each gene pair, co-expression value was calculated as the Pearson correlation coefficient between the two expression profiles.

Paralog gene pairs were identified by performing all-against-all BLASTP similarity searches of yeast open reading frames. We defined two genes as paralogs if (i) the BLAST score had an expected value $E<10^{-8}$, (ii) alignment length exceeded 100 residues, (iii) sequence similarity was >30%, and (iv) they were not parts of transposons.

Gene Expression Analysis

Eight evolved lines were selected for microarray analysis, all of them showing high fitness following evolution (at least 20% initial fitness defect compared to the wild-type control and at least 20% fitness improvement as a result of the evolutionary process). The corresponding ancestral strains and the wild-type control were also subjected to gene expression profiling. Table S3 contains the list of strains. Candidates were re-streaked and single clones were isolated and their fitness increase was confirmed by growth curve recording.

Two independent colonies of the wild-type control, evolved, and corresponding ancestor knock-out strains were inoculated into 15 ml YPD and grown overnight at 30°C. The saturated populations were diluted to an OD_{600} of 0.15 in 60 ml YPD and grown to early mid-log phase (OD_{600} 0.6 ± 0.05) in 250 ml Erlenmeyer flasks with 220 rpm shaking at 30°C. Cells were harvested by centrifugation (4,000 rpm, 3 min, 30°C) and immediately frozen in liquid nitrogen after removal of supernatant. Total RNA was prepared by hot acidic phenol extraction and cleaned up using the QIAGEN's RNAeasy kit.

All steps after RNA isolation were automated using robotic liquid handlers as described previously [62]. Dual-channel 70-mer oligonucleotide arrays were used with a common reference pool of wild-type RNA. Quality control, normalization, and dye-bias correction was performed as described earlier [62]. The reported fold change is the average of the four replicate mutant profiles versus the average of all wild-type controls. A total of 58 transcripts showed stochastic changes in wild-type profiles and were excluded from the analyses. Differentially expressed genes were defined as those showing a 1.7-fold abundance change and a p-value<0.05 when comparing two strains. The raw dataset is available online at ArrayExpress (http://www.ebi.ac.uk/arrayexpress/, accession number E-MTAB-2352).

Robustness of Results of the Transcriptome Analysis to Growth Rate Related Genes and Copy Number Variations

All transcriptome comparisons of the wild-type, knockout, and evolved strains were repeated on a dataset where CNVs, genes showing expression response to aneuploidy, and growth rate related genes were excluded. CNVs were identified on the basis of the read coverage of the genome sequence data (Table S2) with the exception of one strain (Δ*rpl43a*), which was not sequenced. In the case of Δ*rpl43a*, whole chromosome duplication was predicted on the basis of visual inspection of expression profiles. The position of partial chromosome duplication was predicted by the Charm algorithm [63]. In evolved strains carrying aneuploid chromosomes, genes showing expression response to that particular aneuploidy were excluded from the transcriptome comparisons (data on the transcriptome effects of aneuploidy were obtained from [64]). Genes showing significant expression response to changes in growth rate were also excluded, as defined previously [65] on the basis of the growth rate measurements of Brauer and colleagues [66].

Strain Modifications to Investigate the Fitness Costs and Epistatic Effects of Compensatory Mutations

The evolved lines of Δ*mdm34* were chosen for in-depth genetic analysis. The fitness cost of the set of compensatory mutations accumulated in the evolved Δ*mdm34* lineages was measured in wild-type genetic background. To this end, the *MDM34* gene was re-introduced into the ancestor and evolved Δ*mdm34* lineages according to the *delitto perfetto* method [67]. First, the KanMX4 cassette in the ancestor and evolved Δ*mdm34* lineages was swapped with the CORE-UH cassette, containing the *KlURA3* and *hyg* markers. Then the *MDM34* open reading frame with longer than 0.3 kb flanking regions on both sides was amplified from the unevolved wild-type control strain and transformed into the cells to replace the CORE-UH cassette. The replacement of the *KlURA3* marker was counter-selected using 5-FOA containing medium. The loss of hyg^r was confirmed, the site and orientation of gene replacement was verified by PCR and the sequence of the *MDM34* gene was determined by capillary sequencing.

In a second analysis, a point mutation identified in the *MGA2* gene in one of the evolved Δ*mdm34* lineages was reinserted into both the wild-type and ancestor Δ*mdm34* background. This specific point mutation changes the 750th codon of *MGA2* from GAT to TAT resulting in the incorporation of tyrosine instead of aspartic acid. We refer to the mutant allele as *mga2-1*. Using the *delitto perfetto* method [67], we introduced this point mutation into the unevolved wild-type control strain. First, the CORE-UH cassette was inserted into the genome at the desired position of the SNP. Then, two complementary oligonucleotides of 81 bp length with the sequence of the region of interest and the SNP in the 41st position were transformed. The replacement of the *KlURA3* marker with the missense SNP was counter-selected using 5-FOA containing medium, loss of hyg^r was confirmed, and the result of the site-directed mutagenesis was verified by capillary sequencing. Attempts to introduce the *mga2-1* mutation into the ancestor Δ*mdm34* strain in this way were not successful, presumably due to the severe slow growth of the intermediate strain that lacks both *MDM34* and *MGA2* gene in a functional form. To complement this, a helper plasmid with *MDM34* gene (MoBY ORF Library [68]) was transformed into the cells prior to the site directed mutagenesis [69]. Because of the presence of the *URA3* marker on the helper plasmid, the CORE-Hp53 cassette was used in this experiment. The steps of mutagenesis were similar as without the helper plasmid, which was removed by passaging cells through 5-FOA afterwards.

qPCR Method

Yeast samples were grown in 20 ml YPD medium to mid-log phase (0.8 OD600 value). RNA was extracted from 10^7 yeast cells by acidic phenol method using TRI Reagent Protocol (Sigma-Aldrich Co). The RNA samples were concentrated by the NucleoSpin RNA Plant Kit (Macherey-Nagel), according to the manufacturer's instructions. A total of 500 ng RNA was used as a template to prepare cDNA using the Maxima First Strand cDNA Synthesis kit (Thermo Scientific). Reactions without template were

set up to detect contaminations of the reagents used in the cDNA synthesis. qPCR reactions were set up in 20 µl volume, using the following templates: no template control, 10 ng non-transcribed RNA and cDNA transcribed from 10 ng RNA. The qPCR reactions were run in a Bioer LineK Gene device, using 2× Maxima SYBR Green qPCR Master Mix (Thermo Scientific). All samples had three technical replicates. Gene expression was determined in arbitrary units using a standard curve fitted on triplicates of a four-step 10-fold dilution series. *OLE1* expression level was determined relative to *TUB1* expression level as an internal control. All control reactions, not treated with reverse transcriptase or not having template, gave Ct values at least 10 cycles higher than the corresponding samples.

Supporting Information

Figure S1 Fitness trajectories often show a saturating trend by day 104 of the evolution experiment. Fitness was measured at five time points during laboratory evolution (at day 0, 26, 52, 78, and 104), and fitness improvements were tested for each line and time interval (Wilcoxon rank-sum test, with a p-value cut-off of 0.05, see Methods and Table S10). (A) focuses on lines that showed one significant fitness improvement during the four 26-day time intervals. There is a strong (5-fold) depletion of lines that showed a fitness improvement in the last time step of the evolutionary experiment (eight out 159 cases, 40 expected, Chi-square test, $p<10^{-8}$), indicating saturating compensatory evolution. (B) Representative examples of fitness trajectories showing a saturating trend (replicate lines of six genotypes are depicted).
(TIF)

Figure S2 The extent of compensatory evolution in knock-outs is genotype-specific. Here, we tested whether there are inherent differences in the propensity for compensation across genotypes (i.e., lines carrying different gene deletions). We defined compensatory evolution as a fitness increase that is disproportionally large relative to that in the evolving wild-type lines (Table S1). Accordingly, genotypes can be classified into three major categories on the basis of the fraction of corresponding lines fulfilling the above criteria (none, mixed, all). To assess the degree of departure from random expectation a randomization protocol was used. It generated a distribution of the above three categories under the assumption that all genotypes are equally likely to gain high fitness during the course of laboratory evolution. Specifically, the matrix of lines was shuffled one thousand times (gray bars) and the above categories were recalculated. The analysis revealed a strong enrichment of genotypes where all lines were compensated (''all'') and genotypes where none of the lines were compensated (''none''), while the ''mixed'' category was relatively rare (a). This result is not simply due to the fact that null mutations with more severe defects are especially likely to be compensated for. When only genotypes with similar initial fitness defects were considered, the trend remained (b,c,d). The four plots show the observed and randomly expected distributions a, for the whole dataset; b, c, d, for initial fitness ranges <0.7, 0.7–0.8, >0.8, respectively. Genotypes where either all or none of the evolutionary lines showed compensation are significantly enriched in all four cases, the corresponding Chi-square test p-values for a, b, c, and d are $<10^{-20}$, 0.013, 7×10^{-6} and 10^{-8}, respectively.
(TIF)

Figure S3 Global transcriptome changes following compensatory evolution. (A and B) were prepared by reproducing the main results of Figure 4, after excluding genes from the transcriptome profiles that (i) show copy number changes in the evolved lines, (ii) change expression level in aneuploid lines [13], or (iii) whose expression level depends on cellular growth rate (for details see Materials and Method). (A) The Euclidean distances of microarray profiles of the evolved evolutionary line from its ancestor and from wild type (WT) were calculated and normalized to the ancestor–wild type distance for each genotype (Table S11). The distances of the points on the figure are proportional to the calculated profile distances. For each genotype triplet, distances were calculated on the basis of those genes that are differentially expressed in at least one of the pairwise comparisons. (B) The figure focuses on the subset of genes that showed expression change upon gene deletion, and shows the fraction of these genes that changed expression during evolution in the opposite direction (i.e., evolution towards restoration of wild-type expression level; see inset). With one major exception (*Δmdm34*), only a small fraction of the expression changes were restored in the evolved lines (Table S11). The threshold for expression change was 1.7-fold-change and $p<0.05$, as previously described [14].
(TIF)

Figure S4 Pleiotropic effects and mechanism of compensation of *Δmdm34*. (A) Diversity of pleiotropic effects in independently evolved lines. Relative fitness across environments of isolated clones of independently evolving lines founded from the same *Δmdm34* genotype were measured as colony sizes grown on different media (Table S12). Genotypes are indicated on the left, the growth media are indicated above the heat map. For media composition and abbreviations, see Table S4. Values were normalized to that of the ancestral *Δmdm34* strain in the corresponding environment. In (A) and (D) log2 values are shown according to the color coding. (B) Quantitative PCR confirmation of upregulation of *OLE1* in both the evolved line carrying the *mga2-1* mutation and in the *Δmdm34 mga2-1* double mutant strain (Table S12). *OLE1* expression was measured relative to *TUB1* as an internal control and expression values were normalized to *Δmdm34* ancestor. Error bars show standard error. (C) Addition of oleic acid to the medium suppresses the fitness defect of *Δmdm34*, but does not affect the fitness of the evolved line carrying the *mga2-1* mutation or the strain carrying both *Δmdm34* and *mga2-1* mutations. Fitness was measured as colony sizes relative to unevolved wild-type control on rich media supplemented with DMSO as solvent control (non-treated), 0.1 mM oleic acid and 0.1 mM stearic acid (Table S12). For each genotype relative fitness change compared to the corresponding non-treated strain is shown. Error bars show standard error. (D) A specific point mutation in *MGA2* recapitulates the pleiotropic effects of compensatory evolution observed in evolved line 1. Relative fitnesses of *Δmdm34* evolving line 1, and *Δmdm34 mga2-1* double mutant were measured as colony sizes grown on different media (Table S12). Genotypes are indicated on the left, the growth media are indicated above the heat map. For media composition and abbreviations, see Table S4. Values were normalized to that of the ancestral *Δmdm34* strain in the corresponding environment.
(TIF)

Figure S5 Validation of the phenotypic profiling experiment. We compared our colony size measurements (Table S1) of the ancestral knockout strains to a published fitness profiling of the yeast deletion collection [4]. In the environments that match the published study, we find a good agreement between our data and the classification of Dudley and colleagues [4]. In each environment, knockouts present in our dataset were labeled as ''no defect'' versus ''no/slow growth'' based on Dudley and colleague's data. A significant difference was found between the

two groups in our continuous fitness measurement (y-axis) in each of the environments (one-tailed Wilcoxon rank-sum test; */**/*** indicates p-value<0.05/0.01/0.001, respectively).
(TIF)

Table S1 Fitness of strains in various environments. The table includes fitness values of ancestor and evolved strains as measured in liquid YPD and in different agar media. Pleiotropy measures and GO process terms of the deleted genes are also presented.
(XLSX)

Table S2 Mutations identified by Illumina next generation sequencing. The table contains all the identified de novo and ancestral mutations in the sequenced genomes.
(XLS)

Table S3 Microarray analysis results. The table contains microarray data on all ancestral and evolved lines subjected to microarray analysis.
(XLS)

Table S4 Composition of media used for phenotypic profiling. Each media contained 1% yeast extract, 2% pepton, 2% agar, and different carbon sources. Some media also contained growth inhibitors as indicated. Concentration of drug inhibitors were set to have a minor but detectable growth inhibitory effect on the evolving wild-type control (unpublished data). The list of 14 growth media was primarily based on a previous study [4].
(XLSX)

Table S5 Data supporting Figure 3D. The table contains fitness measurements supporting dosage compensation of Δrpl6b by increased copy number of RPL6A.
(XLSX)

Table S6 Data supporting Figure 4B and 4C. (4B) Euclidean distances between pairs of wild-type evolved and ancestor knock-out strains, and also between the corresponding biological replicates. (4C) Categories of expression changes for each gene in the eight evolved knockout strains. Genes, which show initial expression change in the knockout can be categorized as restored, if expression during evolution goes in the opposite direction or unrestored if not. The category "other" includes genes not showing initial expression change.
(XLSX)

Table S7 Data supporting Figure 6. The table contains fitness (colony size) data employed for epistasis analysis between the mdm34 gene deletion and the mutations accumulated in the evolving strains (Figure 6B) and between the mdm34 gene deletion and one particular compensatory mutation ("mga2-1") (Figure 6C) in both YPD and acetic acid, respectively.
(XLSX)

Table S8 Data supporting Figure 7. The table contains colony size measurement data on the environment-dependent compensation of the deletion of rpb9 by a loss-of-function mutation of whi2.
(XLSX)

Table S9 Data supporting Figure 8. Table includes single-gene knockout fitness and relative frequency of suppressing mutations for 3880 non-essential yeast genes.
(XLSX)

Table S10 Data supporting Figure S1. The table contains data on the fitness trajectories of the evolving strains. Fitness was measured at day 0, 26, 52, 78, and 104. The columns "improved day x-y" show whether there is a statistically significant fitness improvement between day x and y, as assessed by one-sided Wilcoxon tests (with false discovery rate correction, p<0.05 cutoff).
(XLSX)

Table S11 Data supporting Figure S3. (S3A) Euclidean distances between pairs of wild-type, evolved, and ancestor knock-outs, after excluding genes from the transcriptome profiles that (i) show copy number changes in the evolved lines, (ii) change expression level in aneuploid lines, or (iii) whose expression level depends on cellular growth rate. (S3B) Table includes categories of expression changes for each gene in the eight evolved knockout strains, excluding genes from the transcriptome profiles that (i) show copy number changes in the evolved lines, (ii) change expression level in aneuploid lines, or (iii) whose expression level depends on cellular growth rate. Genes displaying an initial expression change in the knockout can be categorized as restored, if its expression level changes in the opposite direction during evolution, or unrestored. The category "other" includes genes that did not display an initial expression change.
(XLSX)

Table S12 Data supporting Figure S4. The table contains data on the pleiotropic effects and mechanism of compensation of the deletion strain Δmdm34.
(XLSX)

Text S1 Additional analyses supporting the prevalence of and mechanisms underlying compensatory evolution following gene loss. The text includes a bioinformatic analyses of deleterious loss-of-function variants in natural yeast populations, a case study on compensatory mutations, and a brief description of image analysis used for measuring the extent of compensatory evolution.
(DOC)

Acknowledgments

We thank Charles Boone and Francesca Storici for sharing plasmids and yeast strains and Andrásné Borka for technical assistance. We are also grateful to Laurence D. Hurst for his insightful comments.

Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: CP BP BS FCPH. Performed the experiments: BS DK ZF VL MH PK MJAGK ER. Analyzed the data: GB BS CP BP KK GF. Contributed reagents/materials/analysis tools: GF PK MJAGK. Contributed to the writing of the manuscript: CP BP BS GB FCPH.

References

1. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, et al. (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500: 571–574.
2. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, et al. (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 4: e1000183.
3. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, et al. (2012) A systematic survey of loss-of-function variants in human protein-coding genes. Science 335: 823–828.
4. Kimura M (1985) The role of compensatory neutral mutations in molecular evolution. J Genet 64: 7–19.
5. Weinreich DM, Chao L (2005) Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution 59: 1175–1182.
6. Andersson DI, Levin BR (1999) The biological cost of antibiotic resistance. Curr Opin Microbiol 2: 489–493.
7. Estes S, Lynch M (2003) Rapid fitness recovery in mutationally degraded lines of *Caenorhabditis elegans*. Evolution 57: 1022–1030.

8. Burch CL, Chao L (1999) Evolution by small steps and rugged landscapes in the RNA virus *phi6*. Genetics 151: 921–927.
9. Estes S, Phillips PC, Denver DR (2011) Fitness recovery and compensatory evolution in natural mutant lines of *C. elegans*. Evolution 65: 2335–2344.
10. Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, et al. (2012) Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Res 22: 2376–2284.
11. Kuo D, Licon K, Bandyopadhyay S, Chuang R, Luo C, et al. (2010) Coevolution within a transcriptional network by compensatory trans and cis mutations. Genome Res 20: 1672–1678.
12. Kenigsberg E, Bar A, Segal E, Tanay A (2010) Widespread compensatory evolution conserves DNA-encoded nucleosome organization in yeast. PLoS Comput Biol 6: e1001039.
13. Poon A, Chao L (2005) The rate of compensatory mutation in the DNA bacteriophage *φX174*. Genetics 170: 989–999.
14. Poon AF, Chao L (2006) Functional origins of fitness effect-sizes of compensatory mutations in the DNA bacteriophage *phiX174*. Evolution 60: 2032–2043.
15. Harcombe WR, Springman R, Bull JJ (2009) Compensatory evolution for a gene deletion is not limited to its immediate functional network. BMC Evol Biol 9: 106.
16. Moore FB, Rozen DE, Lenski RE (2000) Pervasive compensatory adaptation in *Escherichia coli*. Proc Biol Sci 267: 515–522.
17. Blank D, Wolf L, Ackermann M, Silander OK (2014) The predictability of molecular evolution during functional innovation. Proc Natl Acad Sci U S A 111: 3044–3049.
18. Andersson DI, Hughes D (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? Nat Rev Microbiol 8: 260–271.
19. Qian W, Ma D, Xiao C, Wang Z, Zhang J (2012) The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep 2: 1399–1410.
20. Huang ME, Rio AG, Nicolas A, Kolodner RD (2003) A genomewide screen in *Saccharomyces cerevisiae* for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A 100: 11529–11534.
21. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, et al. (2010) The genetic landscape of a cell. Science 327: 425–431.
22. Wang Z, Liao BY, Zhang J (2010) Genomic patterns of pleiotropy and the evolution of complexity. Proc Natl Acad Sci U S A 107: 18034–18039.
23. Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, et al. (2012) The molecular diversity of adaptive convergence. Science 335: 457–461.
24. Steffen KK, McCormick MA, Pham KM, MacKay VL, Delaney JR, et al. (2012) Ribosome deficiency protects against ER stress in *Saccharomyces cerevisiae*. Genetics 191: 107–118.
25. Carroll SM, Marx CJ (2013) Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. PLoS Genet 9: e1003427.
26. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, et al. (2000) Functional discovery via a compendium of expression profiles. Cell 102: 109–126.
27. Dudley AM, Janse DM, Tanay A, Shamir R, Church GM (2005) A global view of pleiotropy and phenotypically derived gene function in yeast. Mol Syst Biol 1: 2005.0001.
28. Kornmann B, Currie E, Collins SR, Schuldiner M, Nunnari J et al. (2009) Tethering complex revealed by a synthetic biology screen. Science 325:477–81.
29. Weinreich DM, Watson RA, Chao L (2005) Perspective: sign epistasis and genetic costraint on evolutionary trajectories. Evolution 59: 1165–1174.
30. Baryshnikova A, Costanzo M, Dixon S, Vizeacoumar FJ, Myers CL, et al. (2010) Synthetic genetic array (SGA) analysis in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Methods Enzymol 470: 145–179.
31. Nesser NK, Peterson DO, Hawley DK (2006) RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. Proc Natl Acad Sci U S A 103: 3268–3273.
32. Paredes JA, Carreto L, Simoes J, Bezerra AR, Gomes AC, et al. (2012) Low level genome mistranslations deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC Biology 10: 55.
33. Jonas K, Liu J, Chien P, Laub MT (2013) Proteotoxic stress induces a cell-cycle arrest by stimulating Lon to degrade the replication initiator DnaA. Cell 154: 623–636.
34. Kaida D, Yashiroda H, Toh-e A, Kikuchi Y (2002) Yeast Whi2 and Psr1-phosphatase form a complex and regulate STRE-mediated gene expression. Genes Cells 7: 543–552.
35. Sudbery PE, Goodey AR, Carter BL (1980) Genes which control cell proliferation in the yeast *Saccharomyces cerevisiae*. Nature 288: 401–404.
36. Teng X, Dayhoff-Brannigan M, Cheng WC, Gilbert CE, Sing CN, et al. (2013) Genome-wide consequences of deleting any single gene. Mol Cell 52: 485–494.
37. Fisher RA (1930) The genetical theory of natural selection: a complete variorum edition. Oxford: Oxford University Press.
38. Orr HA (2005h) The genetic theory of adaptation: a brief history. Nat Rev Genet 6: 119–127.
39. Magtanong L, Ho CH, Barker SL, Jiao W, Baryshnikova A, et al. (2011) Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. Nat Biotechnol 29: 505–511.
40. Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387–391.
41. Harrison R, Papp B, Pál C, Oliver SG, Delneri D (2007) Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci U S A 104: 2307–2312.
42. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, et al. (2010) Rewiring of genetic networks in response to DNA damage. Science 330: 1385–1389.
43. Kirschner M, Gerhart J (1998) Evolvability. Proc Natl Acad Sci U S A 95: 8420–8427.
44. Bergmiller T, Ackermann M, Silander OK (2012) Patterns of evolutionary conservation of essential genes correlate with their compensability. PLoS Genet 8: e1002803.
45. Covert AW, Lenski RE, Wilke CO, Ofria C (2013) Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. Proc Natl Acad Sci U S A 110: E3171–E3178.
46. Lenski RE, Ofria C, Pennock RT, Adami C (2003) The evolutionary origin of complex features. Nature 423: 139–144.
47. Iwasa Y, Michor F, Nowak MA (2004) Stochastic tunnels in evolutionary dynamics. Genetics 166: 1571–1579.
48. Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, et al. (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics 169: 1915–1925.
49. Warringer J, Blomberg A (2003) Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. Yeast 20: 53–67.
50. Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A (2003) High-resolution yeast phenomics resolves different physiological features in the saline response. Proc Natl Acad Sci U S A 100: 15724–15729.
51. Li H, Durbin R (2010q) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589–595.
52. Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7: 474.
53. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297–1303.
54. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, et al. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28: 423–425.
55. Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10: 80.
56. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, et al. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461: 1243–1247.
57. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34: D535–D539.
58. Myers CL, Barrett DR, Hibbs MA, Huttenhower C, Troyanskaya OG (2006) Finding function: evaluation methods for functional genomic data. BMC Genomics 7: 187.
59. Pu S, Wong J, Turner B, Cho E, Wodak SJ (2009) Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res 37: 825–831.
60. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, et al. (2004t) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res 32: 5539–5545.
61. Faith JJ, Driscoll ME, Fusaro VA, Cosgrove EJ, Hayete B, et al. (2008) Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata. Nucleic Acids Res 36: D866–D870.
62. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, et al. (2010) Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell 143: 991–1004.
63. Myers CL, Dunham MJ, Kung SY, Troyanskaya OG (2004) Accurate detection of aneuploidies in array CGH and gene expression microarray data. Bioinformatics 20: 3533–3543.
64. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, et al. (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science 317: 916–924.
65. Airoldi EM, Huttenhower C, Gresham D, Lu C, Caudy AA, et al. (2009) Predicting cellular growth from gene expression signatures. PLoS Comput Biol 5: e1000257.
66. Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, et al. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Mol Biol Cell 19: 352–367.
67. Moqtaderi Z, Geisberg JV (2013) Construction of mutant alleles in *Saccharomyces cerevisiae* without cloning: overview and the delitto perfetto method. Curr Protoc Mol Biol 104: Unit 13.10C.
68. Ho CH, Magtanong L, Barker SL, Gresham D, Nishimura S, et al. (2009) A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat Biotechnol 27: 369–377.
69. Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2: 31–34.
70. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, et al. (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Methods 7: 1017–1024.

Molecular Systems Biology 9; Article number 700; doi:10.1038/msb.2013.57
Citation: ***Molecular Systems Biology*** **9:700**
www.molecularsystemsbiology.com

OPEN ACCESS TRANSPARENT PROCESS molecular systems biology

Bacterial evolution of antibiotic hypersensitivity

Viktória Lázár[1], Gajinder Pal Singh[1], Réka Spohn[1], István Nagy[2], Balázs Horváth[2], Mónika Hrtyan[1], Róbert Busa-Fekete[3], Balázs Bogos[1], Orsolya Méhi[1], Bálint Csörgő[1], György Pósfai[1], Gergely Fekete[1], Balázs Szappanos[1], Balázs Kégl[3], Balázs Papp[1,*] and Csaba Pál[1,*]

[1] Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Center, Szeged, Hungary, [2] Genomics Unit, Institute of Biochemistry, Biological Research Center, Szeged, Hungary and [3] Linear Accelerator Laboratory, University of Paris-Sud, CNRS, Orsay, France
* Corresponding authors. C Pál or B Papp, Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Center, Temesvari krt 62, Szeged 6726, Hungary. Tel.: +36 62 599 661; Fax: +36 62 433 506; E-mail: cpal@brc.hu or pappb@brc.hu

Received 27.5.13; accepted 25.9.13

The evolution of resistance to a single antibiotic is frequently accompanied by increased resistance to multiple other antimicrobial agents. In sharp contrast, very little is known about the frequency and mechanisms underlying collateral sensitivity. In this case, genetic adaptation under antibiotic stress yields enhanced sensitivity to other antibiotics. Using large-scale laboratory evolutionary experiments with *Escherichia coli*, we demonstrate that collateral sensitivity occurs frequently during the evolution of antibiotic resistance. Specifically, populations adapted to aminoglycosides have an especially low fitness in the presence of several other antibiotics. Whole-genome sequencing of laboratory-evolved strains revealed multiple mechanisms underlying aminoglycoside resistance, including a reduction in the proton-motive force (PMF) across the inner membrane. We propose that as a side effect, these mutations diminish the activity of PMF-dependent major efflux pumps (including the *AcrAB* transporter), leading to hypersensitivity to several other antibiotics. More generally, our work offers an insight into the mechanisms that drive the evolution of negative trade-offs under antibiotic selection.

Molecular Systems Biology **9**: 700; published online 29 October 2013; doi:10.1038/msb.2013.57
Subject Categories: microbiology & pathogens
Keywords: antibiotic resistance; collateral sensitivity network; evolutionary experiment; trade off

Introduction

Evolutionary adaptation to an environment may be accompanied by a decline or an increase in fitness in other environments. Although such trade-offs are frequently observed in nature, the governing rules and the corresponding molecular mechanisms are generally unclear. The evolution of antibiotic resistance offers an ideal model system to systematically investigate this issue. Enhanced level of resistance can be achieved by mutations in the genome or by acquisition of resistance-conferring genes through horizontal gene transfer. The relative contribution of these mechanisms depends both on the antibiotic employed and on the bacterial species considered (Alekshun and Levy, 2007). It has been suggested that the progressive accumulation of mutations can simultaneously change an organism's sensitivity to many different antimicrobial agents and can serve as the first step in the evolution of clinically significant resistance by more specific and effective mechanisms (Baquero, 2001; Goldstein, 2007; Gullberg *et al*, 2011). A recent review argued that the evolution of multidrug resistance and hypersensitivity are among the central issues of the field (Palmer and Kishony, 2013). Better understanding of these phenomena is important as they could potentially inform future therapeutic strategies to mitigate resistance evolution. For example, the choice of optimal antibiotic combinations depends on both the presence of physiological drug interactions and the frequency of mutations with pleiotropic fitness effects (Chait *et al*, 2007; Palmer and Kishony, 2013; Pena-Miller *et al*, 2013).

Specifically, it remains unclear how frequently genetic adaptation to a single antibiotic increases the sensitivity to others and what the underlying molecular mechanisms of hypersensitivity are. No large-scale, systematic laboratory evolution study has been devoted to investigate this problem under controlled environmental settings. To our best knowledge, the only prior work with similar aims was published 60 years ago and was limited to phenomenological descriptions (Szybalski and Bryson, 1952).

Here, for the first time, we apply an integrated approach to decipher collateral-sensitivity interactions between antibiotics. We initiated the laboratory evolution of *E. coli* populations in the presence of one of the several different antimicrobial agents. These antibiotics are well characterized, widely employed in the clinic, and have diverse modes of actions (Table I). Our list also includes antibiotics that are typically used against Gram-positive bacteria. Consistent with previous studies (Curtiss *et al*, 1965; Vuorio and Vaara, 1992; Elkins and Nikaido, 2002), we found that these antibiotics inhibited the growth of wild-type *E. coli* at high concentrations and that resistance readily evolved against these compounds

Table I Employed antibiotics and their modes of actions

Antibiotic name	Abbreviation	Mode of action	Bactericidal or Bacteriostatic
Ampicillin	AMP*	Cell wall	Bactericidal
Pipericallin	PIP	Cell wall	Bactericidal
Cefoxitin	FOX*	Cell wall	Bactericidal
Fosfomycin	FOS	Cell wall	Bactericidal
Lomefloxacin	LOM	Gyrase	Bactericidal
Ciprofloxacin	CPR*	Gyrase	Bactericidal
Nalidixic acid	NAL*	Gyrase	Bactericidal
Fosmidomycin	FSM	Lipid	Bactericidal
Nitrofurantoin	NIT*	Multiple mechanisms	Bactericidal
Amikacin	AMK	Aminoglycoside	Bactericidal
Gentamicin	GEN	Aminoglycoside	Bactericidal
Kanamycin	KAN*	Aminoglycoside	Bactericidal
Tobramycin	TOB*	Aminoglycoside	Bactericidal
Streptomycin	STR	Aminoglycoside	Bactericidal
Tetracycline	TET*	Protein synthesis, 30S	Bacteriostatic
Doxycycline	DOX*	Protein synthesis, 30S	Bacteriostatic
Chloramphenicol	CHL*	Protein synthesis, 50S	Bacteriostatic
Erythromycin	ERY*	Protein synthesis, 50S	Bacteriostatic
Fusidic acid	FUS	Protein synthesis, 50S	Bacteriostatic
Sulfamonomethoxine	SLF	Folic acid biosynthesis	Bacteriostatic
Trimethoprim	TRM*	Folic acid biosynthesis	Bacteriostatic
Muporicin	MUP	Gram positive	NA
Cycloserine	CYC	Gram positive	NA
Vancomycin	VAN	Gram positive	NA

The functional classification is based on previous studies (Girgis *et al*, 2009; Yeh *et al*, 2006). For an overlapping set of 12 selected antibiotics (indicated by stars), populations were allowed to evolve in the presence of successively increased antibiotic concentrations.

(see below). Next, we charted the network of collateral-sensitivity interactions by measuring the susceptibility of each evolved line against all the other antibiotics. Our analysis revealed a strikingly dense network of collateral-sensitivity interactions. Many of these interactions involved aminoglycosides. Finally, laboratory-evolved lines were subjected to whole-genome sequence analysis and biochemical assays to decipher the underlying molecular mechanisms of these interactions.

Results

Parallel evolution of antibiotic susceptibility patterns in the laboratory

We followed established protocols with minor modifications to evolve bacterial populations under controlled laboratory conditions (Hegreness *et al*, 2008). Starting from a single ancestral clone, populations were propagated in batch culture (minimal glucose medium containing a single antibiotic), whereby 1% of each culture was diluted into fresh medium on a daily basis.

Microbes frequently encounter low or varying antibiotic concentrations (Baquero, 2001). For example, the limited accessibility of antibiotics to certain tissues or incomplete treatment can lead to the formation of concentration gradients within the body (Kohanski *et al*, 2010a). Similarly, antibiotic-polluted natural environments generally form such gradients radiating from the source. To mimic these natural conditions, we employed two selection regimes. In the first set of experiments, a fixed sublethal antibiotic concentration (i.e., sufficient to reduce the growth of the starting population by 50%) was employed. Using this set-up, we propagated 10 independent populations in the presence of each antibiotic for ~140 generations, resulting in 240 parallel-evolved lines. As selection pressure and resistance-conferring mutations can differ substantially between low and high antibiotic concentrations, we also employed a complementary laboratory evolutionary protocol. For an overlapping set of 12 selected antibiotics (Table I), populations were allowed to evolve to successively higher antibiotic concentrations (96 replicate populations per antibiotic). Starting with subinhibitory antibiotic concentrations, the antibiotic concentration was increased every 4 days over a period of 240–384 generations. Despite the short evolutionary timescale, many of the evolved populations reached very high resistance levels (20- to 328-fold increases in the minimum inhibitory concentrations (MICs); Supplementary Table S1). For each antibiotic, we selected 10 independently evolved resistant populations for further analysis (Materials and methods). In addition, to control for potential adaptive changes that are not specific to the employed antibiotics, we also established 10 parallel populations that were grown in an environment devoid of antibiotics, referred to as adapted control lines.

Next, we measured the corresponding changes in the sensitivities of all laboratory-evolved populations to other antibiotics. Fitness was measured by monitoring the optical density of liquid cultures of all evolved and control lines in the presence and absence of sublethal concentrations of antibiotics. Our protocol was highly sensitive and could efficiently detect both weak negative and positive trade-offs, which may be overlooked in other assays (Materials and methods; Supplementary Text S1). Furthermore, by measuring fitness in antibiotic-free medium, we could distinguish between general growth defects and genuine collateral-sensitivity interactions. Specifically, we employed a rigorous statistical procedure to identify those collateral-sensitivity interactions that are not expected based on the generally weak growth defect observed in the absence of antibiotics (see Supplementary Text S2 and Supplementary Figure S3). The reliability of the method was confirmed by comparing its results with sensitivity estimates based on the colony size (Supplementary Text S2; Supplementary Figure S1).

We noticed that parallel-evolving populations exposed to the same antibiotic displayed very similar antibiotic susceptibility patterns (Supplementary Figure S2). Thus, we developed a data analysis pipeline to infer evolutionary interactions at the level of antibiotic pairs based on the growth patterns of the antibiotic-adapted and control populations. The analysis ultimately led to a map of evolutionary interactions between antibiotics (Figures 1A and B; Supplementary Table S2). In this study, we concentrated on antibiotic pairs showing collateral sensitivity; cross-resistance interactions will be described elsewhere.

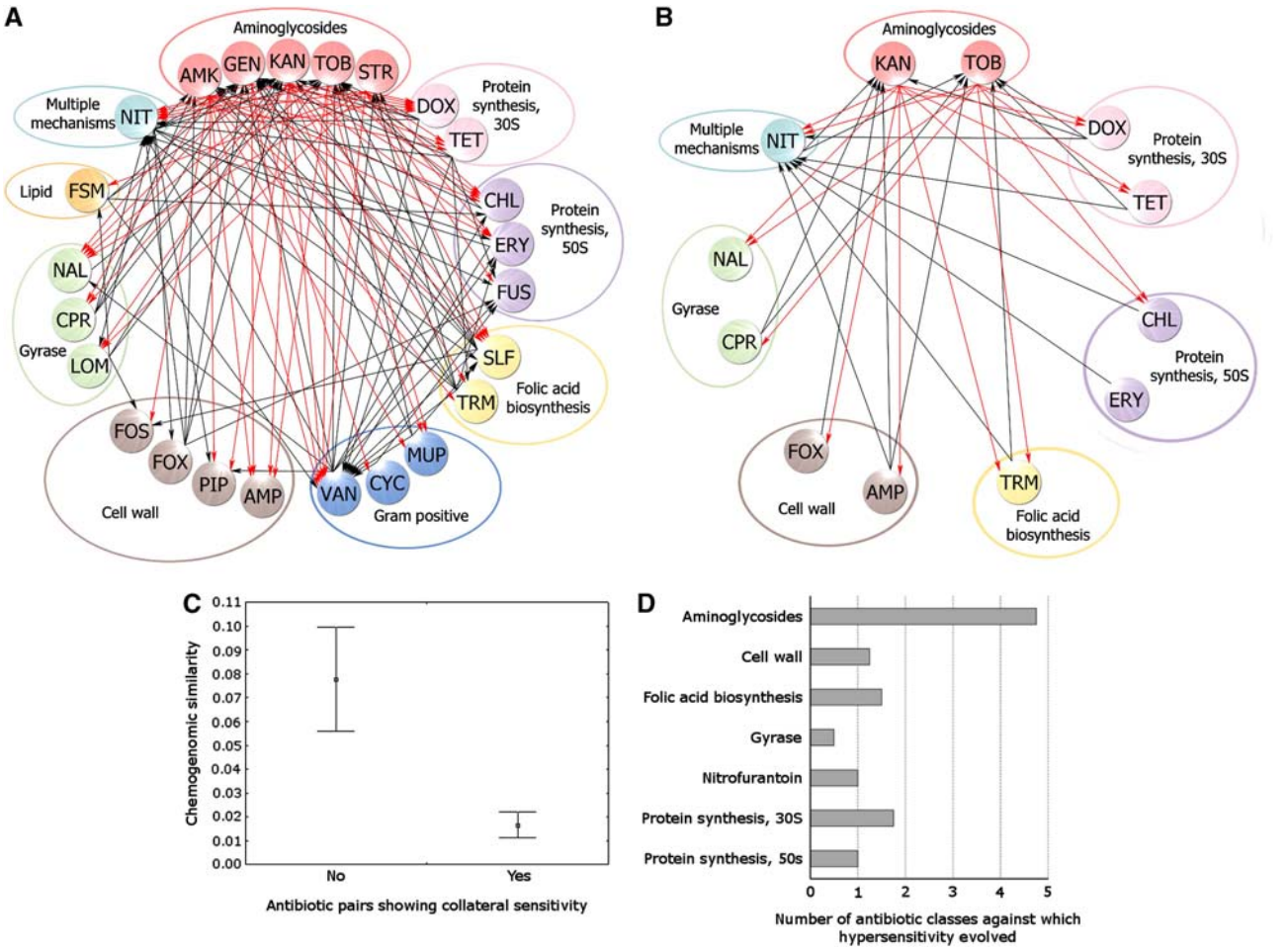


Figure 1 Networks of collateral-sensitivity interactions. Collateral-sensitivity interaction networks inferred from the adaptation to (**A**) low antibiotic concentrations and (**B**) increasing concentrations of antibiotics. Antibiotics are grouped according to their mode of action. An arrow from antibiotic A to antibiotic B indicates that adaptation to A increased the sensitivity to B. Aminoglycosides dominate the collateral-sensitivity network, with numerous links to other classes of antibiotics (red arrows). (**C**) Collateral-sensitivity antibiotic pairs show relatively low overlap in their chemogenomic profiles ($N = 120$, Mann–Whitney U-test $P < 10^{-5}$). Chemogenomic distance was calculated as pairwise Jaccard distance between sets of genes that influence antibiotic susceptibility (Girgis *et al*, 2009). Error bars indicate 95% confidence intervals. (**D**) Collateral-sensitivity interaction degrees of antibiotic classes (i.e., average number of antibiotic classes against which a population evolves hypersensitivity if exposed to the antibiotic class shown on the vertical axis. Degrees are weighted by the number of antibiotics representing each class).

Uneven distribution of collateral sensitivity across antibiotic classes

The maps based on the evolutionary experiments performed with constant and gradually increasing antibiotic concentrations were similar. In all, 85% of the interactions between antibiotics overlapped ($P < 10^{-5}$, randomization test). Three main patterns emerge from our map. First, these interactions occurred frequently: at least 35% of all investigated antibiotic pairs showed collateral sensitivity in at least one direction. Second, the mode of antibiotic action has a strong influence on the distribution of interactions. Collateral sensitivity never occurred between antibiotic pairs that target the same cellular subsystem (Fisher's exact test, $P = 0.013$). Thanks to systematic chemogenomic studies, the mode of antibiotic action can be defined and compared in a more quantitative manner. Specifically, a previous study exposed a nearly complete mutagenized *E. coli* library to several antibiotics and determined the fitness contribution of individual genes (Girgis *et al*, 2009). Using this data set, we calculated the sets of genes that influence susceptibility for each antibiotic used in our study (chemogenomic profile). Collateral sensitivity was depleted between antibiotic pairs with substantial overlap in their chemogenomic profiles (Figure 1C). Third, most antibiotic classes displayed collateral sensitivity with relatively few other classes (Figure 1D). There was one major exception: 44% of the collateral-sensitivity interactions involved aminoglycosides. Genetic adaptation to aminoglycosides increased the sensitivity to many other classes of antibiotics, including inhibitors of DNA synthesis, cell-wall synthesis, and other classes of protein synthesis inhibitors. The observed interactions generally represented 2- to 10-fold decreases in the MICs (Figure 2; Supplementary Table S3), a result that is consistent with an earlier report on antibiotic hypersensitivity (Szybalski and Bryson, 1952). This rate is also rather similar to the 2- to 8-fold increases in MIC typically observed in different efflux pump mutants (Piddock, 2006).

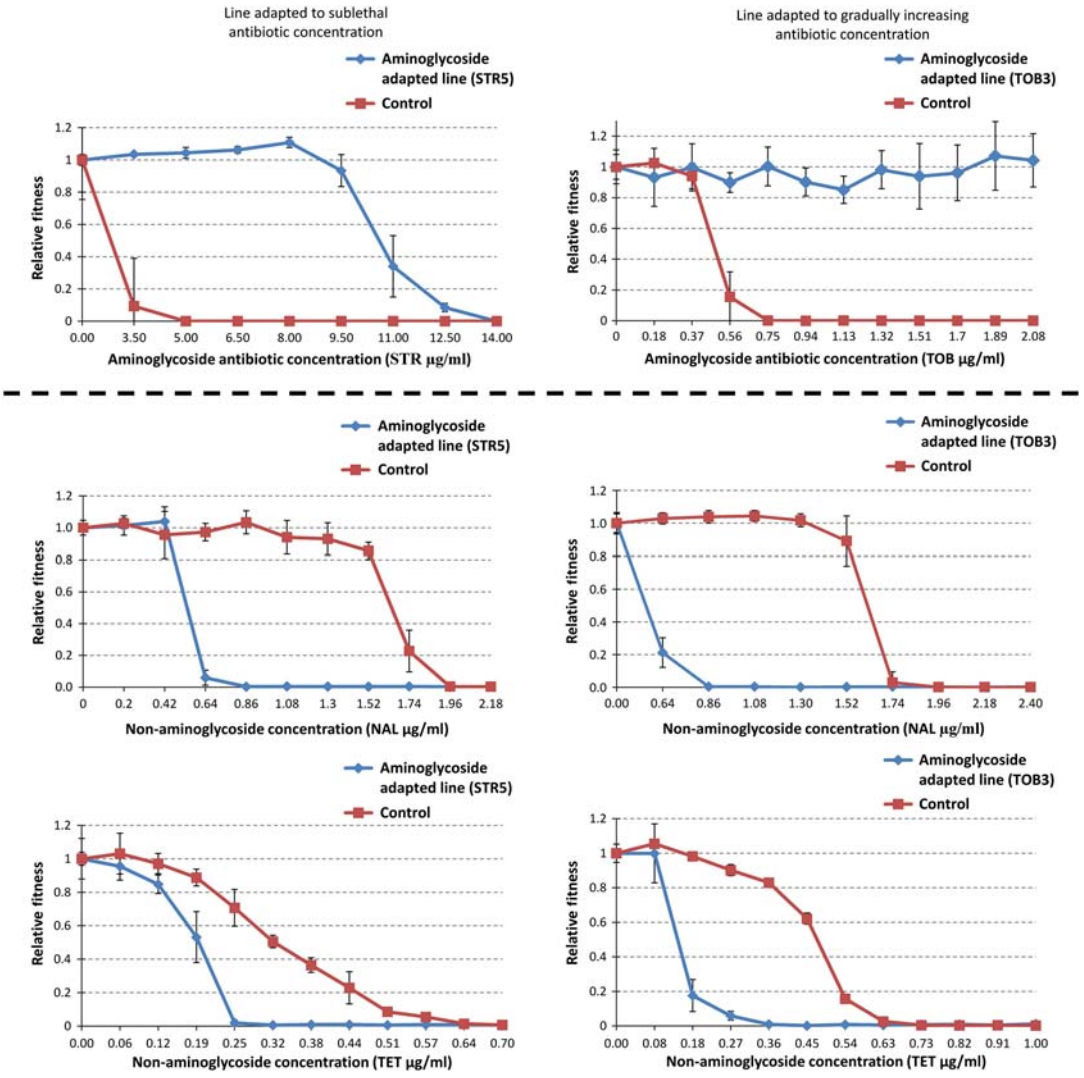


Figure 2 Dose–response curve of selected aminoglycoside-adapted lineages exhibiting collateral sensitivity. Error bars indicate 95% confidence intervals.

Multiple mechanisms underlying aminoglycoside resistance

Three major mechanisms of aminoglycoside resistance have been recognized: inactivation of the drugs by aminoglycoside modifying enzymes, modification of ribosome, and decreased membrane permeability (partly through changes in a membrane potential). To gain insight into the molecular mechanisms underlying aminoglycoside resistance and collateral sensitivity in our laboratory-evolved strains, we selected 14 clones evolved in the presence of a single aminoglycoside (kanamycin, tobramycin, or streptomycin) for whole-genome resequencing. All of these clones exhibited hypersensitivity to other classes of antibiotics (Supplementary Tables S2 and S3). The genomes of these independently evolved clones were resequenced using the Applied Biosystems SOLiD platform, and the identified single-nucleotide polymorphisms (SNPs) were confirmed using capillary sequencing.

In total, we identified 100 mutations (SNPs and indels) affecting 44 protein-coding genes. On average, we observed eight mutations per clone in lines adapted to increasing concentrations and two mutations in those adapted to a fixed sublethal concentration (Supplementary Table S4). Three lines of evidence indicated that these substitutions were driven by adaptive evolution. First, 89% of the mutations were in protein-coding regions and were non-synonymous. Second, convergent evolution was prevalent at multiple levels, as a total of 6.7% of the mutations at the single nucleotide level were shared by two or more clones (Supplementary Table S4). Evolutionary convergence was even more apparent at the level of genes and functional units, as a total of 29.5% of the affected 44 genes were mutated repeatedly (Supplementary Tables S4 and S5). Third, comparison with the results of available chemogenomic screens revealed that 36% of the mutated genes influence aminoglycoside susceptibility when inactivated (Supplementary Table S4, $P = 0.013$, Fisher's exact test).

Aminoglycosides directly target the ribosome. Mistranslation and the consequent misfolding of membrane proteins have an important role in aminoglycoside-induced oxidative stress and cell death (Kohanski *et al*, 2010b). Aminoglycosides

generally require respiration for uptake (Taber *et al*, 1987) and enter the cell in a membrane potential-dependent manner. This process relies on cytochromes and the maintenance of the proton-motive force (PMF) through the quinone pool (Kohanski *et al*, 2010b).

Pathway enrichment analyses (Carbon *et al*, 2009) revealed the overrepresentation of several biological processes in the set of accumulated mutations (Supplementary Table S6). In agreement with our expectations, one major target of selection was the translational machinery, including several ribosomal proteins, elongation factors (fusA, rpsL), and tRNA synthetases. Second, several genes involved in membrane transport, phospholipid synthesis, and cell envelope homeostasis were mutated. Remarkably, this list included an oligopeptide transporter (OppF) with a key role in the recycling of cell-wall peptides and the two-component stress-response sensor CpxA (Kohanski *et al*, 2008; Supplementary Table S4). The biosynthesis of polyamines (including putrescine and spermidine) was also affected. These molecules reduce the intracellular production of reactive oxygen species during aminoglycoside stress (Tkachenko *et al*, 2012) and thereby diminish the levels of protein and DNA damage (Kohanski *et al*, 2010b). Third, we identified a broad class of genes expected to influence the membrane electrochemical potential (Supplementary Tables S4 and S5; Figure 2B). These genes are involved in oxidative phosphorylation, proton-potassium symport (*trkH*), oxygen-binding heme biosynthesis (*hemA*), while others are members of the cytochrome terminal oxidase complex (*cyoB*, *cyoC*). They also frequently affect the quinone pool, which serve as electron carriers in the respiratory electron transport chain (IspA and the Nuo protein complex). This third class most likely has a central contribution to the collateral-sensitivity patterns observed, not least because all sequenced clones had at least one mutation in this subsystem (Supplementary Table S4).

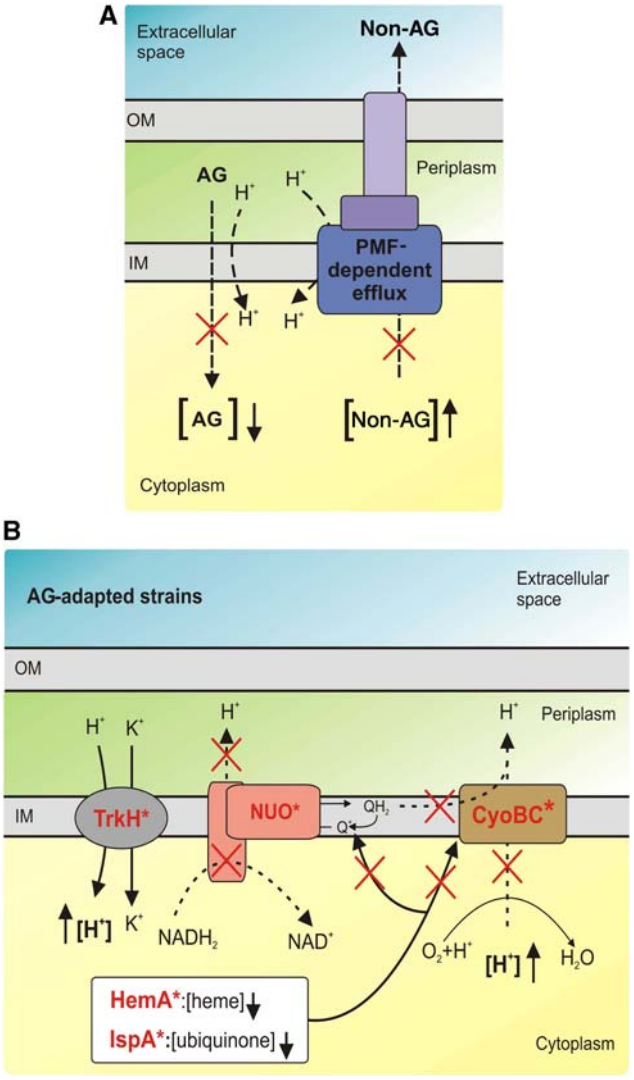


Figure 3 A putative mechanism underlying collateral sensitivity. (**A**) The theory. Altering the membrane potential across the inner bacterial membrane has two opposing effects: it reduces the uptake of many aminoglycoside-related antibiotics but simultaneously may lead to the reduced activity of PMF-dependent efflux pumps. For more details, see the main text. (**B**) Mutations supporting the theory. Whole-genome sequencing revealed that adaptation to aminoglycosides frequently proceeds through mutations that most likely diminish the generation of the PMF. Mutations are indicated by red, bolded protein names (TrkH, CyoB, HemA, IspA). The observed mutations in TrkH most likely increase the proton influx, whereas the mutations in CyoB and HemA (resulting in the inhibition of proton translocation and heme biosynthesis, respectively) interfere with the proper functioning of the cytochrome terminal oxidase complex. Furthermore, decreased IspA activity reduces the levels of membrane-bound quinones and therefore the level of oxidative phosphorylation. Altogether, these mutations likely reduce the PMF and thus aminoglycoside uptake. Simultaneously, the activity of the PMF-dependent efflux system is expected to decrease, resulting in greater sensitivity to antibiotics transported by these pumps. AG, aminoglycoside; OM, outer membrane; IM, inner membrane; NUO, NADH-Ubiquinone-oxidoreductase; PMF, proton-motive force.

Evidence for antagonistic mutational effects on membrane permeability

Why should membrane potential affecting mutations alter the susceptibility to multiple different antibiotics? These genes are expected to influence aminoglycoside-induced oxidative stress and/or aminoglycoside uptake. Indeed, aminoglycosides uniquely require the PMF for active cellular uptake (Taber *et al*, 1987; Allison *et al*, 2011). In sharp contrast, the efflux of many other antibiotics depends on PMF-dependent pumps (Paulsen *et al*, 1996). On the basis of these observations, we propose a model in which low-level aminoglycoside resistance is achieved by altering the membrane potential across the inner bacterial membrane (Figure 3). As a secondary consequence, these mutations diminish the activity of PMF-dependent major efflux pumps. Indeed, it has been previously shown that CCCP, a chemical inhibitor of oxidative phosphorylation, decreases the intracellular accumulation of aminoglycosides (Allison *et al*, 2011), but most likely increases the intracellular accumulation of several other antibiotics (Coldham *et al*, 2010).

Along with the observed mutations, biochemical assays provided further support for this model. First, we investigated changes in the membrane potential in aminoglycoside-resistant strains. The membrane potential was monitored using the carbocyanine dye diethyloxacarbocyanine (DiOC2) (3) (Novo *et al*, 2000). In agreement with our expectations, the membrane potential was reduced in aminoglycoside-adapted populations (Figure 4A; Supplementary Figure S4;

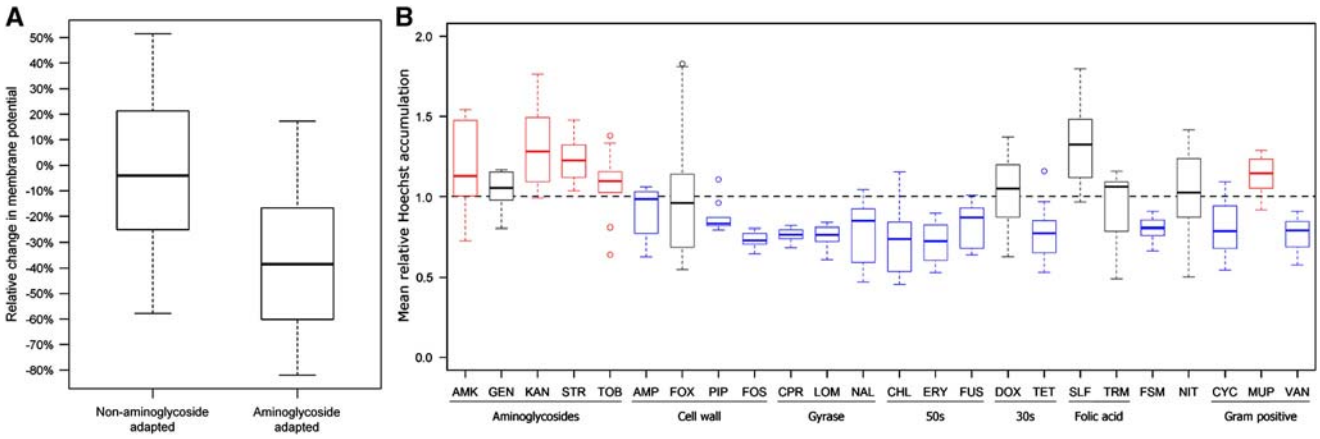


Figure 4 Membrane permeability (Hoechst dye) and membrane potential changes in evolved lines. (**A**) Membrane potential changes in antibiotic-adapted populations. Changes in the membrane potential were monitored using the carbocyanine dye DiOC2(3). The red/green fluorescence values for a representative set of aminoglycoside- and non-aminoglycoside-adapted populations were determined relative to the average of those of three wild-type controls. The membrane potential was significantly reduced in aminoglycoside-resistant populations (Wilcoxon rank-sum test $P = 0.002$). Boxplots present the median and first and third quartiles, with whiskers showing either the maximum (minimum) value or 1.5 times the interquartile range of the data. The data are based on 10 and 22 measurements for aminoglycoside- and non-aminoglycoside-adapted populations, respectively. (**B**) Substantial differences in the accumulation of the fluorescent probe Hoechst 33342 across populations adapted to different classes of antibiotics (10 evolved populations each) relative to wild-type controls. Boxplots present the median and first and third quartiles, with whiskers showing either the maximum (minimum) value or 1.5 times the interquartile range of the data, whichever is smaller (higher). The above figures are based on the results for lineages evolved in the presence of constant sublethal antibiotic concentrations. For further results, see Supplementary Figures S4 and S5.

Supplementary Table S5). Simultaneously, these populations showed elevated intracellular levels of the fluorescent probe Hoechst 33342 (Figure 4B; Supplementary Figure S5), indicating either increased porin activity or diminished efflux pump activity (Coldham *et al*, 2010). This result contrasts with the results for populations adapted to other antibiotic classes, as these populations frequently exhibited reduced intracellular levels of Hoechst 33342 (Figure 4B; Supplementary Figure S5).

The most direct evidence for antagonistic mutational effects comes from a gene involved in K^+ uptake (*trkH*). Mutations in *trkH* were observed in 64% of the sequenced aminoglycoside-adapted populations. The amino-acid residue affected by one of the observed mutations (T350L) is close to the ion channel and therefore was chosen for further analysis. This mutation was inserted into wild-type *E. coli*, and the inserted mutation conferred mild resistance to aminoglycosides and, simultaneously, increased the susceptibility to other classes of antibiotics, as expected (Figure 5A). Consistent with a causal role of the PMF in this negative trade-off, this mutation resulted in a diminished membrane potential and enhanced the accumulation of Hoechst dye (Figures 5B and C). In further support of the involvement of the PMF, a related regulator of K^+ uptake has been shown to control both the membrane potential and the multidrug susceptibility (Castaneda-Garcia *et al*, 2011).

There are further examples supporting the scenario. The list of mutated genes entails four genes (*cyoB*, *ispA*, *nuoF*, and *nuoE*) with the following remarkable combination of properties (Supplementary Table S4). First, functional connection to electron transport can be established based on the literature data, strongly suggesting that these genes influence PMF. Second, 57% of the observed mutations in these genes generate frame-shift or in frame stop-codons, and hence most likely yield proteins with compromised or no activities. Third, null mutations in these genes reduce the aminoglycoside susceptibility but enhance the sensitivity to other antibiotics.

Collateral sensitivity is partly linked to the AcrAB efflux system

Recent studies systematically investigated the substrate specificities of all major drug transporters through deletion and overexpression over a wide range of drugs (Nishino and Yamaguchi, 2001; Girgis *et al*, 2009; Liu *et al*, 2010; Nichols *et al*, 2011). Comparison of results of chemogenomic screens and our study revealed that as high as 75% of the antibiotics showing collateral sensitivity with aminoglycosides are also substrates of the AcrAB efflux pump system. This system is member of the resistance nodulation family, and a major multidrug resistance mechanism in *E. coli*. Overexpression of this system confers resistance to a wide range of drugs and detergents, but not to aminoglycosides (Okusu *et al*, 1996; Nishino and Yamaguchi, 2001; Alekshun and Levy, 2007). A proton electrochemical potential gradient across cell membranes is the driving force for drug efflux by this system.

On the basis of these facts, we suggest that the AcrAB efflux system has a key role in the collateral-sensitivity patterns observed. More specifically, activity of this system is assumed to be impaired in aminoglycoside-resistant lines due to the presence of mutations diminishing the membrane potential. To test this hypothesis, we examined drug resistance phenotypes conferred by the AcrAB efflux system in the presence/absence of mutations in *trkH* and *cyoB*. As shown above, mutations in these genes were frequently observed in aminoglycoside-resistant lines, and we could confirm that the corresponding strains have diminished the membrane potential (Figures 4 and 5). We took advantage of the availability of a multicopy plasmid that encodes the AcrAB transporter genes of *E. coli* with the corresponding native promoters. Following protocols of a prior study (Nishino and Yamaguchi, 2001), the plasmid was transformed into wild-type and aminoglycoside-resistant mutants. We tested the corresponding changes in susceptibilities to four representative antibiotics (all of which are known

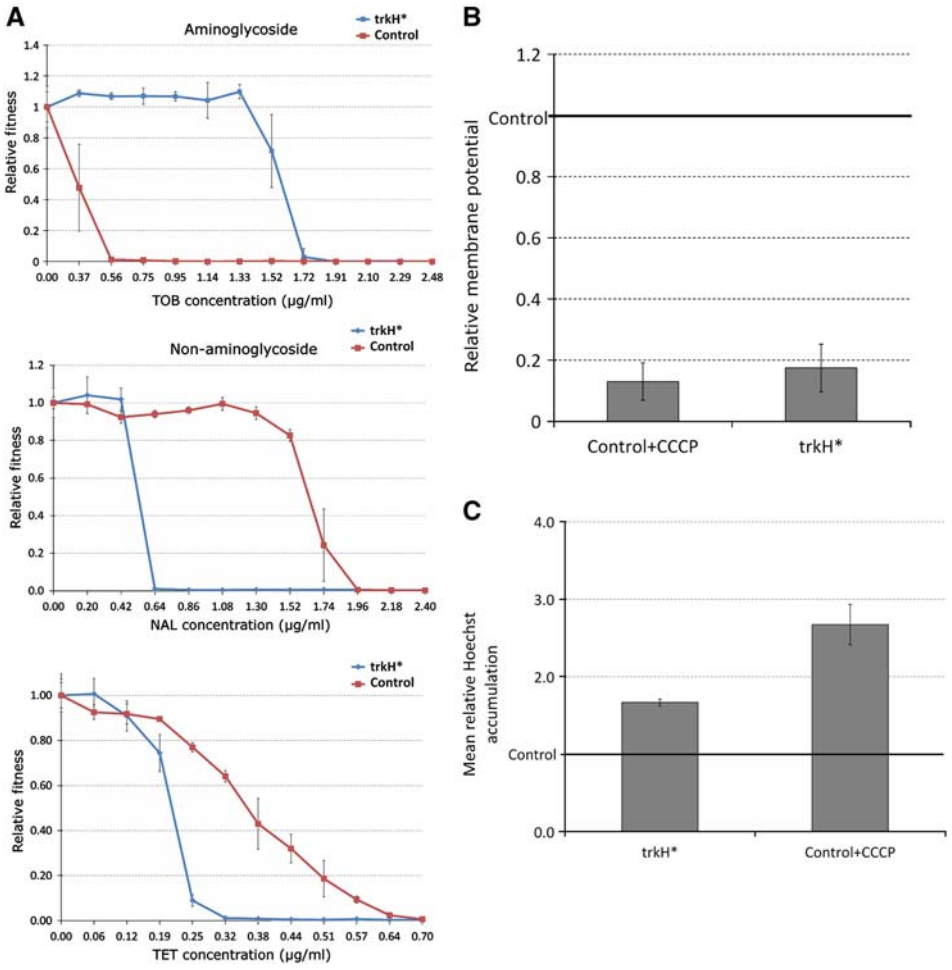


Figure 5 Pleiotropic effects of a single mutation in trkH Individual antibiotic dose–response curves for growth inhibition were constructed for a trkH mutant strain. The red line denotes the trkH mutant strain and the blue line indicates the corresponding wild-type control. Error bars indicate the standard errors based on four technical replicates. A mutation in the *trkH* gene originally identified in a streptomycin-adapted population reduced the susceptibility to aminoglycosides but inhibited growth in the presence of several non-aminoglycoside antibiotic stresses. For more details on minimum inhibitory changes, see Supplementary Table S7. This mutation also (**B**) reduced the membrane potential (Wilcoxon rank-sum test $P = 0.02$, based on four replicate measurements) and (**C**) the enhanced accumulation of Hoechst dye (Wilcoxon rank-sum test $P = 0.0005$, based on eight replicate measurements). Control populations treated with a chemical inhibitor of the PMF (CCCP) showed similar patterns.

substrates of the AcrAB efflux pump). First, strains with deficient AcrAB efflux system were sensitive to all four antibiotics, regardless of the presence of mutations affecting the membrane electrochemical potential (Figure 6). Second, the AcrAB overexpression plasmid conferred a significant resistance to all four antibiotics in genotypes with wild-type membrane potential. Third, and most strikingly, the same plasmid conferred substantially weaker resistance when introduced into the mutant lines (Figure 6). Taken together, these results confirm that mutations conferring aminoglycoside resistance via diminishing the membrane potential increase the sensitivity to other agents by interfering with the AcrAB efflux system.

Discussion

By combining experimental evolution, whole-genome sequencing of laboratory-evolved bacteria and biochemical assays, this work charted a map of collateral-sensitivity interactions between antibiotics in *E. coli*, and aimed to understand these negative evolutionary trade-offs. We demonstrated that collateral-sensitivity interactions occurred at high rates. Strikingly, laboratory evolution to different aminoglycoside antibiotics frequently enhanced the sensitivity to many other antimicrobial agents (2–10 fold MIC change). Whole-genome sequencing of laboratory-evolved strains revealed multiple mechanisms underlying aminoglycoside resistance. As expected, the major targets of selection were the translational machinery, membrane transport, phospholipid synthesis, and cell envelope homeostasis. Strikingly, we also identified a broad class of mutated genes involved in maintenance of the membrane electrochemical potential. Notably, similar sets of mutations have been observed in clinical and experimental settings (Supplementary Table S4).

Aminoglycoside resistance can be achieved through reduction in the PMF across the inner membrane (Proctor *et al*, 2006; Pranting and Andersson, 2010). In this paper, we demonstrated that these changes underlie the hypersensitivity of aminoglycoside-resistant bacteria to several other antimicrobial agents,

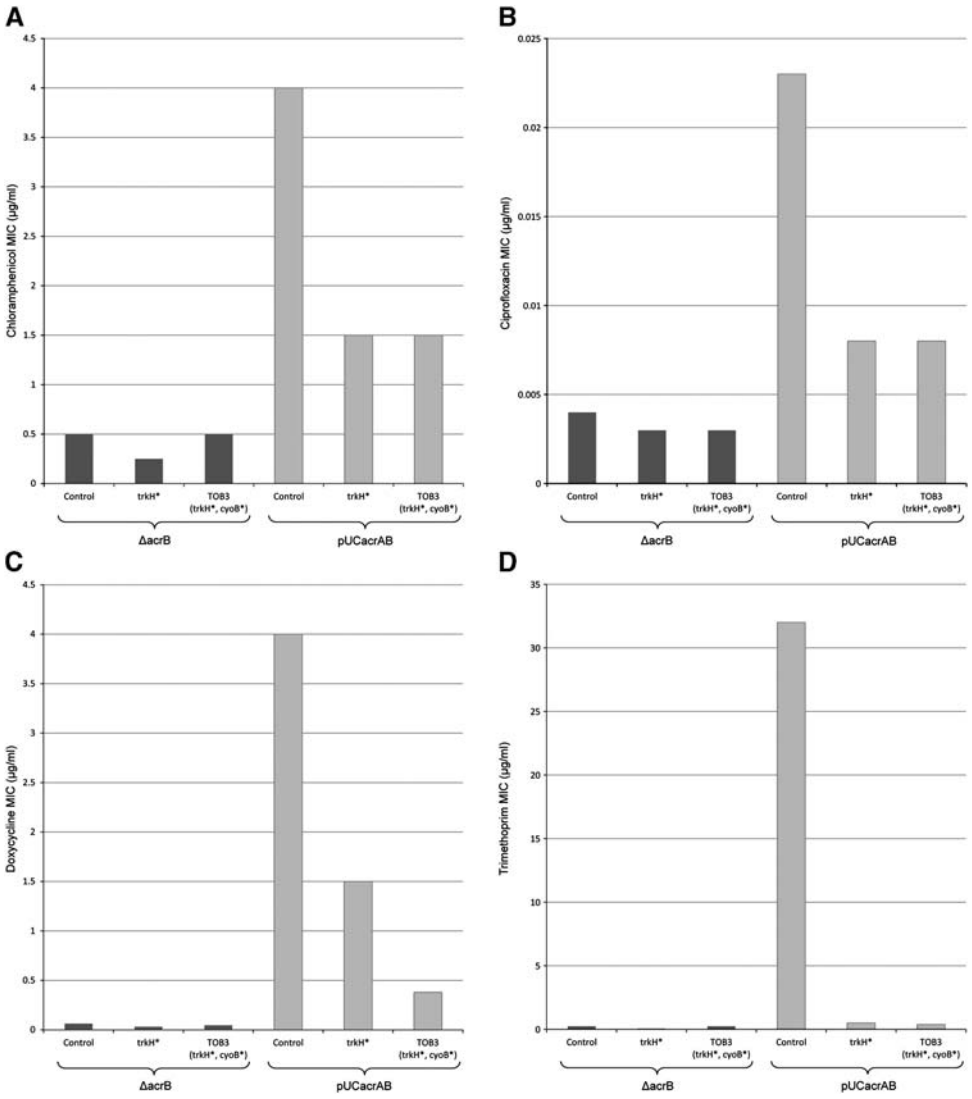


Figure 6 Link between the copy number of the major drug efflux system AcrAB and the extent of collateral-sensitivity interactions. The AcrAB efflux system confers resistance to a variety of drugs, but not to aminoglycosides. Two aminoglycoside-resistant strains (trkH* and TOB3) and a wild-type strain (control) were modified either by deleting the *acrB* gene (ΔacrB) or by harbouring a multicopy plasmid carrying the *acrAB* genes (pUCacrAB). Change of MIC in modified strains was measured using *E*-test stripes containing one of four antibiotics (**A**) Chloramphenicol, (**B**) Ciprofloxacin, (**C**) Doxycycline, (**D**) Trimethoprim, representing different classes of modes of action (Table I). The plasmid conferred a significant resistance to all four antibiotics in control strain, but resistance levels were substantially reduced when the same plasmid was associated with membrane potential affecting mutations in either trkH (trkH*) or both *trkH* and *cyoB* genes (TOB3).

partly through diminishing the activity of PMF-dependent major efflux pumps. Taken together, these results indicate the existence of an antagonistic mechanism by which bacteria modulate intracellular antibiotic concentrations. Evolutionary experiments performed with constant and gradually increasing antibiotic concentrations yielded similar broad-scale collateral-sensitivity patterns (see Figures 1A and B), and one of the membrane potential affecting genes (*trkH*) was repeatedly mutated in response to both treatments (Supplementary Table S4). On the basis of these findings, we speculate that the membrane potential affecting mutations may arise at an early stage of resistance evolution.

Overexpression of a major multidrug transport (AcrAB) conferred only a relatively low level of resistance in association with PMF-affecting mutations. This result could have a broad significance. PMF-dependent efflux pumps are frequently delivered by horizontal gene transfer, and have crucial contribution to the evolution of multidrug resistance patterns in a broad range of bacterial species (Paulsen *et al*, 1996; Mine *et al*, 1999; Norman *et al*, 2008). Thus, resistance to one antibiotic may not only confer changed sensitivity to another antimicrobial agent, but also affect its further evolution (see also Palmer and Kishony, 2013).

We emphasize that we do not consider our explanation exclusive. However, we failed to find evidence for other mechanisms playing a role in the observed collateral-sensitivity patterns of aminoglycoside-resistant populations. Existing chemogenomic, literature, and experimental data are fully consistent with our scenario (see Results). There is at least one potentially interesting case that should be explored in

a future work. The elongation factor fusA was regularly associated with aminoglycoside resistance in our experiments. Previous works suggest that resistance conferred by fusA mutations in Salmonella caused enhanced sensitivity to other classes of antimicrobial agents (Macvanin and Hughes, 2005). Strikingly, these mutations also caused low levels of heme biosynthesis and reduced respiratory activity.

More generally, it will be important to determine how conserved these networks of collateral-sensitivity interactions are between bacterial species. Due to the similarities in the cellular uptake mechanisms of aminoglycosides and cationic antimicrobial peptides (Moore and Hancock, 1986), our results may be more general. At least three mutated proteins (ArnC, SbmA, and TrK) in our experiments influence resistance to antimicrobial peptides (Supplementary Table S4), and more generally, mutations in the heme biosynthesis pathway in Salmonella provide resistance to several small peptides and aminoglycosides and simultaneously increase the susceptibility to other antibiotics (Pranting and Andersson, 2010).

Last, we need to emphasize the limitations of our work. First, a substantial fraction of the collateral-sensitivity interactions unrelated to aminoglycosides require mechanistic explanation (Figure 1). Second, we neglected the evolution of resistance through the acquisition of genes by horizontal transfer. Third, due to the lack of systematic studies, the frequency of PMF-altering mutations in clinical isolates is largely unknown. Therefore, discussing the direct therapeutic implications of our collateral-sensitivity network is beyond the scope of this paper. It remains controversial whether the temporal rotation or simultaneous use of two antibiotics can select against the development of resistance (Bonhoeffer *et al*, 1997; Chait *et al*, 2007). The success of such strategies may depend on the choice of antibiotics, as treatment with a single antibiotic followed by a switch to a cross-sensitive partner may represent a viable strategy. More generally, our work established the prevalence of antagonistic pleiotropy during the evolution of antibiotic resistance.

Materials and methods

Laboratory evolutionary experiments

We followed established protocols with minor modifications to evolve resistant bacteria by propagating them in batch cultures in the presence of antibiotics (Orlen and Hughes, 2006; Hegreness *et al*, 2008). Populations of *Escherichia coli* K12 (BW25113) were grown in autoclaved MS-minimal medium supplemented with 0.2% glucose and 0.1% casamino acids. Antibiotic solutions were prepared from powder stocks and filter sterilized before use and diluted in the growth media. Fresh antibiotic stocks were used on a weekly basis. Parallel cultures were propagated in 96-well microtiter plates, continuously shaken at ~320 r.p.m. (30°C). Plates were covered with special sandwich covers (Enzyscreen) to ensure an optimal oxygen exchange rate and limit evaporation. Every 24 h, bacterial cells were transferred by inoculating ~1–1.2 μl of stationary phase culture to 100 μl fresh medium using a 96-pin replicator (VP407) to give a daily dilution of ~100 or about 6–7 doublings. The evolved and control strains were preserved at −80°C in 20% (v/v) glycerol solution.

Constant, sublethal antibiotic dosage

Using sublethal antibiotic concentrations (approximately half maximal inhibitory concentrations, IC50), we propagated 10 independent populations in the presence of each of the 24 antibiotics for ~140 generations (Table I), resulting in 240 parallel-evolved lines. In addition, to control for potential adaptive changes that are not specific to the employed antibiotics, we also established 10 parallel populations growing in an environment devoid of antibiotics for ~140 generations, referred to as adapted control lines. As expected, these lines showed no major changes in antibiotic susceptibility. Plates contained 36 bacteria-free wells to monitor potential contamination events: all remained uncontaminated during the entire course of the experiment. On average, the employed antibiotic had a mere 7% inhibitory effect on the growth of laboratory-adapted populations.

Gradually increasing antibiotic dosage

In this experimental setting, populations were allowed to evolve to successively higher antibiotic concentrations. Starting with a subinhibitory (IC50) antibiotic concentration, antibiotic dosage was increased gradually (1.5 times the previous dosage) at every fourth transfer. The optical density at 600 nm (OD600) of each well was measured in a Biotek Synergy plate reader before each transfer. As expected, during the course of laboratory evolution, populations grew to different densities, reflecting independent evolutionary trajectories. Population extinction was defined as the failure to obtain growth ($OD600 < 0.05$). The experiments ended once only 10 populations had showed growth or antibiotic concentration had reached its upper solubility limit.

As this laboratory evolutionary protocol frequently leads to extinction of bacterial populations, 96 independent parallel populations were propagated in the presence of each antibiotic. Due to the large number of replicate lineages required, we concentrated on 12 selected antibiotics out of the 24 listed in Table I. This set still covers diverse modes of actions, but includes only 1–2 members of each major antibiotic class.

Most surviving bacterial populations from the final day of the experiments reached a very high resistance level, comparable to that found in clinical isolates (Supplementary Table S1). Depending on the antibiotics employed (and the corresponding extinction dynamics of parallel evolving populations), the experiments lasted for ~240–384 generations (Supplementary Table S1). For each antibiotic, 10 populations with the highest cell densities were selected for further analysis. We also established 10 parallel populations growing in an environment devoid of antibiotics for the same number of transfers, referred to as adapted control lines. Subsequently, we determined the antibiotic susceptibilities of these selected populations against all other antibiotics.

Systematic measurement of antibiotic susceptibilities

Given two panels of laboratory-evolved strains, our next goal was to detect changes in their sensitivities towards other antimicrobial agents. To this end, we developed a high-throughput screening and robust statistical analysis methodology to systematically detect collateral-sensitivity interactions in *E. coli*.

Growth measurement

Bacterial growth was monitored by measuring optical density (OD600) of the liquid cultures at a single time point. Preliminary experiments showed that a single reading of optical density after 14 h of incubation shows very strong linear correlation ($R^2 > 0.99$) with the area under the growth curve, a descriptor of overall inhibitory effect that covers the entire growth period (Supplementary Figure S6). We used a robotic liquid handling system (Hamilton Star Workstation) to improve reproducibility and thereby allowing us to perform hundreds of growth measurements in parallel on 384-well microtiter plates. Slight variations in temperature or humidity within the plate during incubation may lead to local trends of altered growth (within-plate effects). To overcome any measurement bias caused by the inhomogeneous environment and to convert raw OD values into relative fitness values that are comparable across plates, we employed a normalization procedure as described in Supplementary Text S1.

Estimating collateral sensitivity

We tested the sensitivity of each evolved line against the entire set of antibiotics by measuring the growth in liquid cultures of all antibiotic-adapted lines and adapted control lines at sublethal doses of antibiotics (i.e., at around half-maximal effective concentration) in four technical replicates (i.e., strains were cultivated in quadruplicate on the same 384-well plate). In addition, we also measured the growth of evolved lines in a medium devoid of antibiotics to discern condition-specific fitness defects from general costs of resistance.

Because growth media with half-maximal effective concentrations of antibiotics are difficult to prepare in a reproducible manner, we conducted four independent experimental runs for each combination of strains and antibiotic conditions. Next, to filter out unreliable measurements and those where the antibiotic dosage was too high to detect collateral-sensitivity interactions, we excluded cases where (i) cross-contamination might have occurred on the plate during susceptibility measurements (based on the growth in non-inoculated wells), (ii) the control wells devoid of antibiotics showed large variations (coefficient of variation was above 20%), (iii) the applied antibiotic dosage was too high which strongly inhibited the growth of the adapted control populations (>90% effective concentration). This quality control procedure resulted in 2–3 replicates for each combination of strains and antibiotics.

To identify statistically significant collateral-sensitivity interactions, we compared normalized growth values of evolved lines with that of adapted control lines under the same treatment condition. Specifically, using growth data on evolved lines in each antibiotic condition, we tested whether growth of the 10 evolved lines, as a group, showed at least 10% difference from that of the 10 adapted control lines, as a group, under the same treatment condition. Statistical significance was assessed using a non-parametric bootstrap method (Efron and Tibshirani, 1994) and growth of each line was represented by the median value of the four technical replicates. The *P*-values resulting from independent experimental runs were combined using Fisher's combined probability test (data can be found in '*P*_values1' sheets of Supplementary Tables S8 and S9). As a final step, we set up a rigorous statistical procedure to ensure that the collateral-sensitive interactions detected above are not due to general fitness costs of resistance (see Supplementary Text S2, data can be found in 'RF_values' and '*P*_values2' sheets of Supplementary Tables S8 and S9). This yielded a matrix of evolutionary interactions between antibiotics (Supplementary Table S2). For more details on the accuracy of high-throughput interaction measurements and control for potential confounding factors, see Supplementary Text S2.

Whole-genome resequencing

Fourteen independently evolved clones were subjected to next-generation sequencing to identify mutations responsible for the resistant phenotype. All of them showed hypersensitivity towards other classes of antibiotics, and two of them had evolved to constant, sublethal antibiotic dosage. Briefly, genomic DNA (gDNA) was extracted from selected *E. coli* isolates (SIGMA GenElute Bacterial Genomic DNA kit, standard procedures), fragmented, and SOLiD sequencing adaptors were ligated. Subsequently, sequencing beads were prepared and sequenced on the SOLiD System. As a result, millions of short reads (50 or 75 bp) were generated along with data indicating the sequencing quality of each nucleotide. Finally, variants, such as SNP or multi-nucleotide polymorphism (MNP), insertions and deletions (InDels), were identified compared with the reference *E. coli* genome using standard bioinformatics analysis.

Preparation of the libraries and sequencings were performed by cycled ligation sequencing on a SOLiD 5500xl System (Life Technologies; LT) using reagents and protocols provided by LT. Briefly, 3 μg of purified bacterial gDNA was fragmented by the Covaris S2 System to 150–350 bp. The fragmented DNA was end-repaired and ligated to P1 (5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3′) and P2 (5′CTGCCCCGGGTTCCTCATTCTCTGTGTAAGAGGCTGCTGACGGC CAAGGCG-3′) adapters, which provide the primary sequences for both amplification and sequencing of the sample library fragments. The P2 adapter contains a 10-bp barcode sequence which provided the basis for multiplex sequencing. The templates were clonally amplified by emulsion PCR (ePCR) with P1 primer covalently attached to the bead surface. Emulsions were broken with butanol, and ePCR beads enriched for template-positive beads by hybridization with P2-coated capture beads. Template-enriched beads were extended at the 3′ end in the presence of terminal transferase and 3′ bead linker. Beads with clonally amplified DNA were then deposited onto a SOLiD Flowchip (Ondov *et al*, 2010). About 250 million beads with clonally amplified DNA were deposited onto one lane of the flowchip. The slide was then loaded onto a SOLiD 5500xl instrument and the 50-base sequences were obtained according to the manufacturer's protocol.

Further analysis of genome sequences

The obtained sequences were aligned to the *E. coli* str. K-12 substr. MG1655chromosome (Accession NC_000913; Version NC_000913.2 GI: 49175990). Alignment was performed using Genomics Workbench (Floratos *et al*, 2010) 4.9 and the Omixon Gapped SOLiD Alignment 1.3.2 plugin, provided by CLC Bio and Omixon, respectively. A minimum average coverage of 50-fold was accomplished for each strain. The maximum gap and mismatch count within a single read was set to 2 with a minimum of 4 reads to call a potential variation before further analysis. Selected putative variants (SNPs and indels) detected by whole-genome resequencing were verified by PCR followed by Sanger sequencing on a 3500 Series Genetic Analyzer (LT). The primers were designed using Genomics Workbench and are available on request.

Hoechst dye (H33342 bisbenzimide) accumulation assay

To estimate changes in cellular permeability, we implemented a recently developed and scalable fluorescence assay (Coldham *et al*, 2010). The method is based on accumulation of the fluorescent probe Hoechst (H) 33342 (bisbenzimide). All laboratory-evolved populations were cultured overnight in MS-minimal medium supplemented with 0.2% glucose and 0.1% casamino acid. Optical densities of evolved bacterial populations were adjusted to OD600 = 0.3. In all, 180 μl aliquots of bacterial cultures were transferred to 96-well microtiter plates (8 technical replicates per evolved line). Plates were incubated in a Synergy 2 microplate reader at 30°C, and 25 μM Hoechst dye (SIGMA) was added to each well using an auto-injection device (BioTek dispenser box). The OD and fluorescence curves were measured for 1 h with 75-s delays between readings. The first 15 data points were excluded from further analysis due to the high standard deviation between replicates. Blank normalized OD values were calibrated as described in Supplementary Text S1. Data curves were smoothed and fluorescence per OD ratio curves were calculated. Next, areas under these ratio curves were determined. Finally, we calculated changes in Hoechst dye accumulations relative to the appropriate wild-type controls derived from the same experiment.

Measurement of bacterial membrane potential

Two adapted strains per each antibiotic selection regime were selected randomly, and were subjected to membrane potential measurement. The BacLight BacterialMembrane Potential Kit (B34950, Invitrogen) was used to assess changes in PMF in the evolved strains. Briefly, DiOC2 exhibits green fluorescence in all bacterial cells, but the fluorescence shifts towards red emission in cells maintaining high membrane potential. The ratio of red to green fluorescence provides a measure of membrane potential that is largely independent of cell size.

Overnight bacterial cultures were diluted to $\sim 10^6$ cells/ml in filtered buffer (PBS). Aliquots of 200 μl bacterial suspension were added to 96-well microtiter plates for staining treatments. The DiOC2 dye was added to each sample in a 0.03-mM concentration (no antibiotic was added to the medium). After 30 min of incubation, samples were diluted 10-fold, and were analysed by a GUAVA EasyCyte 8HT Capillary Flow Cytometer. The instrument settings were adjusted according to the BacLight kit manual. In all, 15 000 events were recorded and gated

out by visual inspection using the forward versus side scatter before data acquisition. The red/green fluorescence values for a representative set of aminoglycoside and non-aminoglycoside evolved populations were calculated relative to the average of three control wild-type populations.

MIC and dose–response curve measurements

MICs were determined using a standard linear broth dilution technique (Wiegand *et al*, 2008). In order to maximize reproducibility and accuracy, we used a robotic liquid handling system (Hamilton Star Workstation) to prepare 12 linear dilution steps automatically in 96-well microtiter plates. Approximately 10^6 bacteria/ml were inoculated into each well with a 96-pin replicator, and were propagated at 30°C shaken at 300 r.p.m. (4 replicates per strain/antibiotic concentration). After 24 h of incubation, raw OD values were measured in a Biotek Synergy 2 microplate reader. MIC was defined by a cutoff OD value (i.e., mean + 2 standard deviations of OD values of bacteria-free wells containing only growth medium).

Allele replacements

Allele replacement in trkH was constructed by a suicide plasmid-based method in a markerless allele replacement, which can be distinguished by sequencing of the given chromosomal region. For details on primer sequences, see Feher *et al* (2008). Standard steps and plasmids (pST76-A, pSTKST) of the procedure have been described (Feher *et al*, 2008). In brief, an ~800-bp long targeting DNA fragment carrying the desired point mutation in the middle was synthesized by PCR, then cloned into a thermosensitive suicide plasmid. The plasmid construct was then transformed into the cell, where it was able to integrate into the chromosome via a single crossover between the mutant allele and the corresponding chromosomal region. The desired cointegrates were selected by the antibiotic resistance carried on the plasmid at a non-permissive temperature for plasmid replication. Next, the pSTKST helper plasmid was transformed, then induced within the cells, resulting in the expression of the I-SceI meganuclease enzyme, which cleaves the chromosome at the 18-bp recognition site present on the integrated plasmid. The resulting chromosomal gap is repaired by way of RecA-mediated intramolecular recombination between the homologous segments in the vicinity of the broken ends.

AcrAB efflux system and collateral sensitivity

To investigate the drug resistance phenotype conferred by the AcrAB efflux system, three different strains were modified either by deleting the *acrB* gene or by transforming a multicopy plasmid carrying the AcrAB efflux pump (pUCacrAB) into various strains. The three selected strains were the following: (i) the ancestral BW25113 strain (control), (ii) an aminoglycoside-evolved line carrying mutations in the *trkH* and *cyoB* genes (TOB3, see Supplementary Table S4), and (iii) the ancestral strain carrying a mutation in trkH (T350L; the corresponding strain will be referred to as trkH*). The appropriate acrB deletion strains (ΔacrB/control, ΔacrB/trkH*, ΔacrB/TOB3) were constructed using standard protocols by P1 transduction (Baba *et al*, 2006; Miller, 1972).

To increase the copy number of the efflux pump, the plasmid pUCacrAB with the corresponding native promoters was transformed into the appropriate strains (pUCacrAB/control, pUCacrAB/trkH*, pUCacrAB/TOB3) following the protocols of a prior study (Nishino and Yamaguchi, 2001). The pUCacrAB plasmid was constructed and provided us by Kunihiko Nishino and Akihito Yamaguchi (Osaka University, Osaka, Japan).

Changes in susceptibility towards four representative antimicrobial agents (chloramphenicol, ciprofloxacin, trimethoprim, and doxycycline) were tested applying E-test stripes (bioMerieux). E-test inoculum preparation and plating, strip application, and subsequent MIC determinations were carried out in accordance with the manufacturer's instructions. The applied antibiotics represent different classes of mode of action; however, the AcrAB efflux system is known to cause resistance towards all of them.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We would like to thank Stephen G Oliver and Olivier Tenaillon for their insightful comments on an earlier version of this manuscript. We are also grateful to Dr Nishino for sending us plasmids used in this study. We also acknowledge the use of the NGS platform established by the ERC Advanced Grant 'SymBiotics' of Eva Kondorosi. This work was supported by grants from the European Research Council (202591) and the EMBO Young Investigator Programme (to CP), the Wellcome Trust, and the 'Lendület Program' of the Hungarian Academy of Sciences (to CP and BP), and the International Human Frontier Science Program Organization (to BP).

Author contributions: BP and CP conceived and supervised the project; VL designed the experiments, VL, RS, IN, BH, MH, BB, OM, and BC performed the experiments; GPS, VL, RB-F, GF, BS, and BK developed data analysis procedures and interpreted the data; GP gave technical support; CP and PB wrote the manuscript with contributions of all other authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alekshun MN, Levy SB (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128:** 1037–1050

Allison KR, Brynildsen MP, Collins JJ (2011) Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nature* **473:** 216–220

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

Baquero F (2001) Low-level antibacterial resistance: a gateway to clinical resistance. *Drug Resist Updat* **4:** 93–105

Bonhoeffer S, Lipsitch M, Levin BR (1997) Evaluating treatment protocols to prevent antibiotic resistance. *Proc Natl Acad Sci USA* **94:** 12106–12111

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* **25:** 288–289

Castaneda-Garcia A, Do TT, Blazquez J (2011) The K+ uptake regulator TrkA controls membrane potential, pH homeostasis and multidrug susceptibility in Mycobacterium smegmatis. *J Antimicrob Chemother* **66:** 1489–1498

Chait R, Craney A, Kishony R (2007) Antibiotic interactions that select against resistance. *Nature* **446:** 668–671

Coldham NG, Webber M, Woodward MJ, Piddock LJ (2010) A 96-well plate fluorescence assay for assessment of cellular permeability and active efflux in Salmonella enterica serovar Typhimurium and Escherichia coli. *J Antimicrob Chemother* **65:** 1655–1663

Curtiss 3rd R, Charamella LJ, Berg CM, Harris PE (1965) Kinetic and genetic analyses of D-cycloserine inhibition and resistance in Escherichia coli. *J Bacteriol* **90:** 1238–1250

Efron B, Tibshirani R (1994) An Introduction to the Bootstrap Chapman & Hall, New York

Elkins CA, Nikaido H (2002) Substrate specificity of the RND-type multidrug efflux pumps AcrB and AcrD of Escherichia coli is determined predominantly by two large periplasmic loops. *J Bacteriol* **184:** 6490–6498

Feher T, Karcagi I, Gyorfy Z, Umenhoffer K, Csorgo B, Posfai G (2008) Scarless engineering of the Escherichia coli genome. *Methods Mol Biol* **416:** 251–259

Floratos A, Smith K, Ji Z, Watkinson J, Califano A (2010) geWorkbench: an open source platform for integrative genomics. *Bioinformatics* **26:** 1779–1780

Girgis HS, Hottes AK, Tavazoie S (2009) Genetic architecture of intrinsic antibiotic susceptibility. *PLoS One* **4:** e5629

Goldstein F (2007) The potential clinical impact of low-level antibiotic resistance in Staphylococcus aureus. *J Antimicrob Chemother* **59:** 1–4

Gullberg E, Cao S, Berg OG, Ilback C, Sandegren L, Hughes D, Andersson DI (2011) Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathog* **7:** e1002158

Hegreness M, Shoresh N, Damian D, Hartl D, Kishony R (2008) Accelerated evolution of resistance in multidrug environments. *Proc Natl Acad Sci USA* **105:** 13977–13981

Kohanski MA, DePristo MA, Collins JJ (2010a) Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol Cell* **37:** 311–320

Kohanski MA, Dwyer DJ, Collins JJ (2010b) How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol* **8:** 423–435

Kohanski MA, Dwyer DJ, Wierzbowski J, Cottarel G, Collins JJ (2008) Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* **135:** 679–690

Liu A, Tran L, Becket E, Lee K, Chinn L, Park E, Tran K, Miller JH (2010) Antibiotic sensitivity profiles determined with an Escherichia coli gene knockout collection: generating an antibiotic bar code. *Antimicrob Agents Chemother* **54:** 1393–1403

Macvanin M, Hughes D (2005) Hyper-susceptibility of a fusidic acid-resistant mutant of Salmonella to different classes of antibiotics. *FEMS Microbiol Lett* **15:** 215–220

Miller JH (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Press, Cold Spring Harbor, NY, USA

Mine T, Morita Y, Kataoka A, Mizushima T, Tsuchiya T (1999) Expression in Escherichia coli of a new multidrug efflux pump, MexXY, from Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **43:** 415–417

Moore RA, Hancock RE (1986) Involvement of outer membrane of Pseudomonas cepacia in aminoglycoside and polymyxin resistance. *Antimicrob Agents Chemother* **30:** 923–926

Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, Lee S, Kazmierczak KM, Lee KJ, Wong A, Shales M, Lovett S, Winkler ME, Krogan NJ, Typas A, Gross CA (2011) Phenotypic landscape of a bacterial cell. *Cell* **144:** 143–156

Nishino K, Yamaguchi A (2001) Analysis of a complete library of putative drug transporter genes in Escherichia coli. *J Bacteriol* **183:** 5803–5812

Norman A, Hansen LH, She Q, Sorensen SJ (2008) Nucleotide sequence of pOLA52: a conjugative IncX1 plasmid from Escherichia coli which enables biofilm formation and multidrug efflux. *Plasmid* **60:** 59–74

Novo DJ, Perlmutter NG, Hunt RH, Shapiro HM (2000) Multiparameter flow cytometric analysis of antibiotic effects on membrane potential, membrane permeability, and bacterial counts of Staphylococcus aureus and Micrococcus luteus. *Antimicrob Agents Chemother* **44:** 827–834

Okusu H, Ma D, Nikaido H (1996) AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol* **178:** 306–308

Ondov BD, Cochran C, Landers M, Meredith GD, Dudas M, Bergman NH (2010) An alignment algorithm for bisulfite sequencing using the Applied Biosystems SOLiD System. *Bioinformatics* **26:** 1901–1902

Orlen H, Hughes D (2006) Weak mutators can drive the evolution of fluoroquinolone resistance in Escherichia coli. *Antimicrob Agents Chemother* **50:** 3454–3456

Palmer AC, Kishony R (2013) Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet* **14:** 243–248

Paulsen IT, Brown MH, Skurray RA (1996) Proton-dependent multidrug efflux systems. *Microbiol Rev* **60:** 575–608

Pena-Miller R, Laehnemann D, Jansen G, Fuentes-Hernandez A, Rosenstiel P, Schulenburg H, Beardmore R (2013) When the most potent combination of antibiotics selects for the greatest bacterial load: the smile-frown transition. *PLoS Biol* **11:** e1001540

Piddock LJ (2006) Multidrug-resistance efflux pumps—not just for resistance. *Nat Rev Microbiol* **4:** 629–636

Pranting M, Andersson DI (2010) Mechanisms and physiological effects of protamine resistance in Salmonella enterica serovar Typhimurium LT2. *J Antimicrob Chemother* **65:** 876–887

Proctor RA, von Eiff C, Kahl BC, Becker K, McNamara P, Herrmann M, Peters G (2006) Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Microbiol* **4:** 295–305

Szybalski W, Bryson V (1952) Genetic studies on microbial cross resistance to toxic agents. I. Cross resistance of Escherichia coli to fifteen antibiotics. *J Bacteriol* **64:** 489–499

Taber HW, Mueller JP, Miller PF, Arrow AS (1987) Bacterial uptake of aminoglycoside antibiotics. *Microbiol Rev* **51:** 439–457

Tkachenko AG, Akhova AV, Shumkov MS, Nesterova LY (2012) Polyamines reduce oxidative stress in Escherichia coli cells exposed to bactericidal antibiotics. *Res Microbiol* **163:** 83–91

Vuorio R, Vaara M (1992) The lipid A biosynthesis mutation lpxA2 of Escherichia coli results in drastic antibiotic supersusceptibility. *Antimicrob Agents Chemother* **36:** 826–829

Wiegand I, Hilpert K, Hancock RE (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3:** 163–175

Yeh P, Tschumi AI, Kishony R (2006) Functional classification of drugs by properties of their pairwise interactions. *Nat Genet* **38:** 489–494

nature COMMUNICATIONS

ARTICLE

Received 13 Nov 2013 | Accepted 9 Jun 2014 | Published 8 Jul 2014

DOI: 10.1038/ncomms5352 OPEN

Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network

Viktória Lázár[1,*], István Nagy[2,*], Réka Spohn[1,*], Bálint Csörgő[1], Ádám Györkei[1], Ákos Nyerges[1], Balázs Horváth[2], Andrea Vörös[2], Róbert Busa-Fekete[3], Mónika Hrtyan[1], Balázs Bogos[1], Orsolya Méhi[1], Gergely Fekete[1], Balázs Szappanos[1], Balázs Kégl[4], Balázs Papp[1] & Csaba Pál[1]

Understanding how evolution of antimicrobial resistance increases resistance to other drugs is a challenge of profound importance. By combining experimental evolution and genome sequencing of 63 laboratory-evolved lines, we charted a map of cross-resistance interactions between antibiotics in *Escherichia coli*, and explored the driving evolutionary principles. Here, we show that (1) convergent molecular evolution is prevalent across antibiotic treatments, (2) resistance conferring mutations simultaneously enhance sensitivity to many other drugs and (3) 27% of the accumulated mutations generate proteins with compromised activities, suggesting that antibiotic adaptation can partly be achieved without gain of novel function. By using knowledge on antibiotic properties, we examined the determinants of cross-resistance and identified chemogenomic profile similarity between antibiotics as the strongest predictor. In contrast, cross-resistance between two antibiotics is independent of whether they show synergistic effects in combination. These results have important implications on the development of novel antimicrobial strategies.

[1] Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Temesvari krt 62, Szeged 6726, Hungary. [2] Sequencing Platform, Institute of Biochemistry, Biological Research Centre, Temesvari krt 62, Szeged 6726, Hungary. [3] MTA-SZTE Research Group on Artificial Intelligence, Tisza Lajos krt 103., H-6720 Szeged, Hungary. [4] Linear Accelerator Laboratory, University of Paris-Sud, CNRS, Orsay 91898, France. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to B.P. (email: pappb@brc.hu) or to C.P. (email: cpal@brc.hu).

Evolutionary adaptation to a specific environment may result in correlated fitness changes in other environments[1,2]. While such evolutionary interactions are widespread in nature, the general principles and underlying molecular mechanisms remain poorly understood[3]. Antibiotic resistance in bacteria offers a platform to systematically investigate evolutionary adaptations. The evolution of antibiotic resistance is frequently mediated by the accumulation of mutations across the genome during therapy[4]. The accumulation of such mutations can potentially change the sensitivity to many antibiotics simultaneously[5]. Despite their clinical relevance, the altered sensitivity profiles of antibiotic resistant strains have not been investigated systematically, except for a pioneering but largely phenomenological study published 60 years ago[6]. Recent works[7,8] investigated the frequency and mechanisms underlying collateral sensitivity (that is, when genetic adaptation under antibiotic stress yields enhanced sensitivity to other antibiotics). The aim of the current paper is to provide insights into the general principles driving cross-resistance interactions. Here, we (i) chart the network of such evolutionary cross-resistance interactions, (ii) explore the underlying molecular mechanisms and (iii) investigate the extent to which cross-resistance is predictable based on the knowledge of antibiotic properties and the set of accumulated mutations.

To accomplish these goals, we initiated parallel laboratory evolutionary experiments with *Escherichia coli* to adapt to increasing dosages of one of 12 antibiotics, and inferred a network of cross-resistance interactions. Laboratory-evolved lines were subjected to whole-genome sequence analysis and biochemical assays to decipher the underlying molecular mechanisms of these interactions.

The following main conclusions were reached. First, the cross-resistance network was dense, indicating that exposure to a single antibiotic frequently yields multidrug resistance. Cross-resistance between two antibiotics is largely independent of whether they show synergistic effects in combination. Second, evolution of resistance is partly achieved through the accumulation of genomic rearrangements and loss-of-function mutations. Third, as parallel evolution at the molecular level is prevalent, cross-resistance patterns are predicable based on the set of accumulated mutations and chemogenomic profile similarities between antibiotics. Taken together, resistance evolution is governed by mutations with highly pleiotropic, but predictable side-effects.

Results

High-throughput laboratory evolutionary experiments. In a previous work[7], we initiated high-throughput laboratory evolutionary experiments starting with *E. coli* K12. Parallel evolving bacterial populations were exposed to 1 of 12 antibiotics (Table 1). Starting from a single ancestral clone, populations were allowed to evolve to successively higher antibiotic concentrations. Evolved populations reached up to 328-fold increases in the minimum inhibitory concentrations relative to the ancestor (Supplementary Table 1). For each antibiotic, 10 independently evolved, resistant populations were subjected to further analysis.

Using an established high-throughput and highly sensitive protocol[7], we previously measured the corresponding changes in susceptibilities of the 120 laboratory-evolved populations to all other 11 antibiotics (Supplementary Data 1). The reliability of the detected cross-resistance interactions was confirmed by measuring changes in minimum inhibitory concentrations using standard *E*-tests (Fig. 1b): the rates of false positives and negatives were around 5 and 16%, respectively (Supplementary Data 2). This allowed us to calculate the frequency of cross-resistance (FCR) interactions for each antibiotic pair (see Methods) and ultimately chart a map of cross-resistance between antibiotics (Fig. 1a).

Properties of the cross-resistance network. Three main patterns emerged from our map (Fig. 1a). First, the evolution of multidrug resistance was frequent under a single antibiotic pressure: on average, 52% of all investigated antibiotic pairs showed cross-resistance in at least one direction. However, the strength of cross-resistance interactions in the data set was highly variable and caused 2 to 128-fold increases in minimum inhibitory concentrations (Fig. 1b). Antibiotic pairs belonging to different functional classes also showed evidence of cross-resistance (Supplementary Data 2). For example, lines adapted to the gyrase inhibitor ciprofloxacin displayed 48 to 68-fold enhancements in resistance to a cell wall inhibitor (cefoxitin).

Table 1 | Antibiotics employed and their modes of actions.

Antibiotic name	Abbreviation	Mode of Action	Bactericidal or Bacteriostatic
Ampicillin	AMP	Cell wall	Bactericidal
Cefoxitin	FOX	Cell wall	Bactericidal
Ciprofloxacin	CPR	Gyrase	Bactericidal
Nalidixic Acid	NAL	Gyrase	Bactericidal
Nitrofurantoin	NIT	Multiple mechanisms	Bactericidal
Kanamycin	KAN	Protein synthesis, 30S, Aminoglycosides	Bactericidal
Tobramycin	TOB	Protein synthesis, 30S, Aminoglycosides	Bactericidal
Tetracycline	TET	Protein synthesis, 30S	Bacteriostatic
Doxycycline	DOX	Protein synthesis, 30S	Bacteriostatic
Chloramphenicol	CHL	Protein synthesis, 50S	Bacteriostatic
Erythromycin	ERY	Protein synthesis, 50S	Bacteriostatic
Trimethoprim	TRM	Folic acid biosynthesis	Bacteriostatic

Functional classification is based on refs 12,20. These antibiotics are widely deployed in the clinic, well characterized, cover a wide range of modes of actions and were subjects of chemogenomic studies in this species[20].

Second, antibiotics differed in their numbers of cross-resistance interactions (Fig. 1c). For instance, adaptation to doxycycline or fluoroquinolones generally led to multidrug resistance. As expected, the corresponding evolved lines frequently accumulated mutations in putative multidrug resistance genes (see below). In sharp contrast, lines adapted to aminoglycosides had few if any cross-resistance interactions, reflecting unusual resistance mechanisms and a unique pathway for cellular uptake[9]. Next, we investigated the other side of the coin: the extent to which resistance to a given antibiotic was achieved by selection to other antibiotics. For each antibiotic, we calculated the number of different antibiotic treatments that select for increased resistance against a given antibiotic (see in-degree on Fig. 1c). In this case, nitrofurantoin was an interesting outlier: nitrofurantoin resistance was reached in only 3% of the populations adapted to other antibiotics (Supplementary Data 1).

Third, prior works indicated that concurrent application of two antibiotics could be used to counter resistance evolution[10]. The efficiency of such combination treatment is determined by at least two factors. It depends on whether the two antibiotics show a synergistic or antagonistic effect on bacterial growth when used in combination (that is, their combined effect is above or below the sum of their individual effects)[11]. Furthermore, it depends on the availability of mutations that confer resistance to both antibiotics. Therefore, it is important to establish whether the antibiotic cross-resistance map overlaps with results of a previous antibiotic combination screen[12]. Aminoglycosides displayed an especially large number of synergistic interactions on growth when used in combination with other antibiotics and, as noted above, were also depleted of cross-resistance with other antibiotic classes ($P = 0.008$, $N = 55$, Kruskal–Wallis test). After excluding this antibiotic class, neither synergistic nor antagonistic antibiotic pairs were enriched in cross-resistance interactions ($P = 0.35$, $N = 45$, Kruskal–Wallis test; Fig. 1d). Thus, networks based on evolutionary and physiological antibiotic interactions show little overlap.

Adaptive mutations dominate in the laboratory-evolved lines. To gain insights into the underlying molecular mechanisms, we

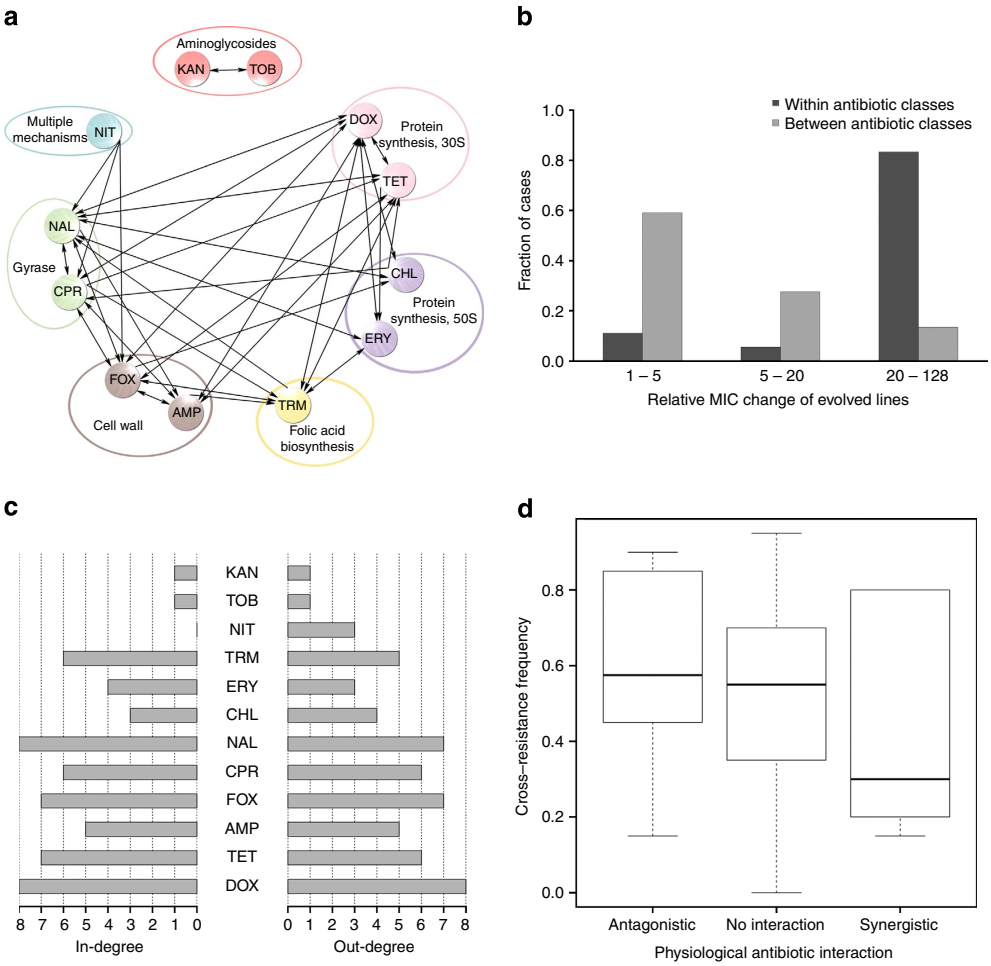


Figure 1 | Cross-resistance interactions and their general properties. (**a**) Network of cross-resistance interactions. Antibiotics are grouped according to their mode of action. An arrow from antibiotic A to antibiotic B indicates that adaptation to A decreased sensitivity to B in at least 50% of the evolved populations. (**b**) Distribution of the strength of cross-resistance interactions, as estimated by *E*-tests. (**c**) Cross-resistance interaction degrees of antibiotics. In-degree measures the number of antibiotic treatments which select for increased resistance against a given antibiotic while out-degree is defined as the number of antibiotics to which cross-resistance evolves when adapting to a given drug. The data are based on that of **a**. (**d**) The frequency of cross-resistance interactions between antibiotics is independent of whether they show physiological interactions (that is, synergy or antagonism), $P = 0.35$, $N = 45$, Kruskal–Wallis test. Aminoglycosides are excluded from the analysis as they show an especially large number of synergistic interactions and are strongly depleted in cross-resistance interactions with other antibiotics). Box plot presents the median and first and third quartiles, with whiskers showing either the maximum (minimum) value or 1.5 times the interquartile range of the data, whichever is smaller (larger).

selected 63 independently evolved lines from the final day of experiments (5–6 lines per antibiotic). These lines were subjected to whole-genome sequencing using the Applied Biosystems SOLiD platform. We implemented an established computational pipeline to identify mutations relative to the ancestral genome (see Methods). To ensure that our pipeline correctly identified true mutations, a set of randomly chosen structural variants, such as point mutations, deletions, inversions and duplications, were validated by independent methods, that is, Sanger sequencing and qPCR. Altogether 16 validations were performed and the results are in complete agreement with the whole-genome sequencing data (Supplementary Table 2). Mutator bacterial populations have frequently been associated with decreased antibiotic susceptibility in clinics[13,14] and laboratory evolution[15]. In agreement with this general trend, two evolved lines exerted elevated genomic mutation rates due to mutations in methyl-directed mismatch repair and in the DNA proof-reading subunit of DNA polymerase III (Supplementary Fig. 1). As a consequence, these lines accumulated exceptionally large numbers of mutations (synonymous and non-synonymous alike), many of which were unlikely to be functionally relevant (Fig. 2a and Supplementary Data 3). Therefore, these lines were excluded from all further analyses.

For the remaining 61 lines, we identified 402 independent mutational events (SNPs, insertions, small and large genomic rearrangements). On average, we detected 4.2 point mutations, 1.2 deletions, 0.26 insertions and 0.07 duplications per clone (Fig. 2a,b). Deletions were generally short (1–100 bp), with 19 major exceptions that span over 0.3–58 kb and eliminated 1–61 genes (Fig. 2c and Supplementary Data 3). Insertion sequences (IS) initiated large-scale genomic rearrangements (inversion, transposition or duplication) and were observed in 59% of the laboratory-evolved lines (Supplementary Data 3).

Several lines of evidence indicate that the accumulation of the mutations in protein-coding regions was largely driven by selection towards increased resistance. First, 87% of point mutations were non-synonymous. Second, at least 19% of the mutated genes showed significant sequence similarity to known antibiotic resistance genes[16] (Fig. 2d and Supplementary Data 4), and several observed substitutions were previously found in natural or clinical isolates (Supplementary Data 5).

Consistent with prior studies[17], antibiotic resistance generally conferred a measurable fitness cost: at least 41% of the laboratory-evolved lines showed a significantly reduced growth in antibiotic-free medium compared to the wild-type. As expected, lines with especially low fitness values in antibiotic-free medium have

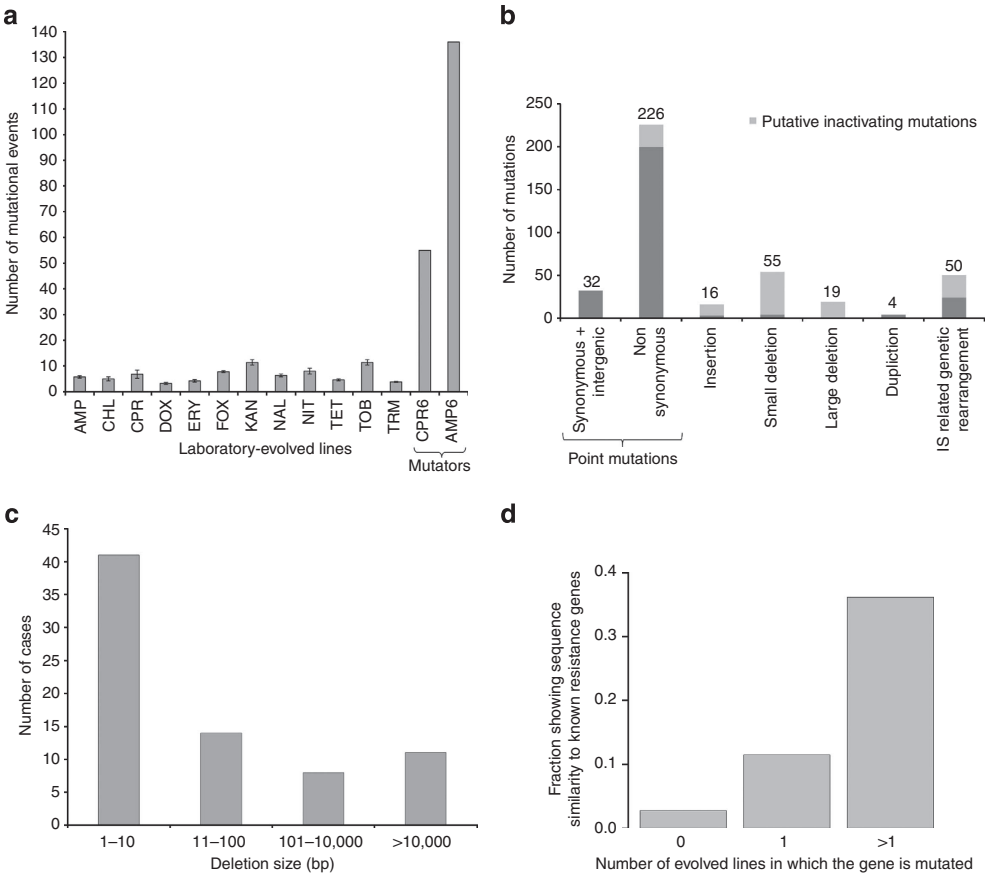


Figure 2 | Mutations identified in independently evolved lines. Distribution of mutational events according to antibiotic (**a**), type (**b**) and size of DNA deletions (**c**). Laboratory-evolved mutator lines have accumulated exceptionally large numbers of mutations. The total number of putative loss-of-function mutations among point mutations, insertions and small deletions is 27% (**b**). (**d**) Observed mutations and known antibiotic resistance genes. Genes mutated in evolved lines are more likely to show significant sequence similarity to known antibiotic resistance genes[16] than non-mutated ones (28 out of 143 versus 120 out of 4,358, $P<10^{-14}$, Fisher's exact test). Furthermore, genes showing sequence similarity to known resistance genes are enriched among genes mutated in multiple lines compared with those mutated in a single line (17 out of 47 versus 11 out of 96, $P<0.005$, Fisher's exact test). We identified genes showing significant sequence similarity to a set of genes curated in the Comprehensive Antibiotic Resistance Database[16] using BLASTP search. In brief, we used the standalone NCBI BLASTP+ tool to identify *E. coli* genes that show highly significant similarity to any of the curated resistance or target genes (a conservative *E*-value cutoff of 10^{-30} was applied).

accumulated large numbers of mutations, including deletions of large genomic segments (Supplementary Fig. 2).

Loss-of-function mutations are prevalent. Over 27% of the observed point mutations, small deletions and insertions generated in-frame stop codons, frameshifts or disruption of the start codon. These mutations were most likely to yield proteins with compromised or no activities (Fig. 2b and Supplementary Data 3). This figure is significantly higher than that observed in a previous large-scale laboratory evolutionary experiment towards high temperature[18] (90 out of 329 versus 145 out of 1,030, Fisher's exact test, $P = 1.017 \times 10^{-7}$). Furthermore, the frequency of nonsense mutations among point mutations is three-fold higher than expected, based on the spontaneous mutation rate inferred from whole-genome sequencing of mutation-accumulation lines[19] (26 out of 258 versus 8 out of 233, Fisher's exact test, $P < 0.005$). This result indicates widespread positive selection on inactivating mutations in our data set. Comparison with chemogenomic data[20] indicated that inactivation of the corresponding genes tends to reduce antibiotic susceptibility compared with that of all other genes in the *E. coli* genome (14 out of 43 versus 321 out of 3,933, Fisher's exact test, $P < 10^{-5}$). In many cases, the null mutations enhanced resistance to multiple drugs (Supplementary Data 6). For example, loss-of-function mutations occurred repeatedly in transcriptional repressors of antibiotic stress response (for example, *acrR*, *marR* and *mprA*). Similarly, IS-related inversions and transpositions frequently disrupted genes with known influence on antibiotic susceptibility. For instance, loss-of-function mutations in the NADPH nitroreductase genes (*nfsA* and *nfsB*) cause resistance to nitrofurantoin and related agents[21]. These genes were disrupted four times independently in nitrofurantoin-evolved lines (for other examples, see Supplementary Data 3).

Evidence for parallel evolution. A strong pattern of parallel evolution emerged at the level of amino-acid sites, genes and functional modules. Eight per cent of the point mutations were shared by at least two lines, and some were shared extensively (Supplementary Data 3). For example, a specific mutation (Val1127Gly) in a subunit (*acrB*) of the AcrAB/TolC efflux system was shared by four lines adapted to three different antibiotics (CHL, AMP and FOX). A total 35% of the affected genes were mutated repeatedly (Fig. 3a and Supplementary Table 3). These repeatedly mutated genes were especially likely to show significant sequence similarity to known antibiotic resistance genes[16] (Fig. 2d and Supplementary Data 4), and some were frequently found in clinical multidrug-resistant strains[22–28]. Similarly, 2% of the observed small deletion events (1–82 bp) and 75% of the large deletion events (0.3–58 kbp) were at identical or nearly identical positions in two or more lines (Supplementary Data 3). These large deletions were generally flanked by homologous IS elements, suggesting that these deletions were mediated by recombination events between IS elements (Supplementary Data 3).

The above figures are all the more surprising as 66% of all parallel mutated genes occurred in lines adapted to different antibiotics. These results indicate that despite substantial differences in antibiotic treatments, the ultimate targets of antibiotic selection are overlapping functional modules. To investigate this issue further, we grouped 88% of the mutations into several major resistance mechanisms based on literature data (Table 2). The following major conclusions can be drawn.

First, mutations in the subsystem targeted by the antibiotic were only found in 49% of the laboratory-evolved lines. The absence of target mutations in the remaining lines may reflect unusually high associated fitness costs[5], rarity of appropriate mutations and/or the efficiency of alternative resistance mechanisms (such as modification of efflux mechanisms, see Table 2). Mutations putatively affecting enzymatic modification of the antibiotic were observed in nitrofurantoin-adapted lines only (Table 2).

Second, genes involved in membrane transport, porin biosynthesis and membrane permeability were repeatedly mutated (Table 2), especially in lines adapted to protein synthesis inhibitors and quinolones. In sharp contrast, such mutations were conspicuously absent in aminoglycoside-resistant populations (Table 2, see also ref. 7).

Third, transcriptional regulatory genes were highly enriched in the set of accumulated mutations (Table 2). Many of them belong to specific two-component regulatory systems, and mediate general cellular defence against stressful conditions. These conditions include osmotic (OmpR/EnvZ, AcrR), acidic (PhoQ), metal (ComR), membrane (CpxR), antibiotic and oxidative stresses (MarA/SoxS/Rob regulon). Consistent with their roles in antibiotic tolerance[29], global transcriptional regulatory proteins (RpoC, Crp and Fis) were also occasionally mutated.

Fourth and more generally, nutrient and oxidative stress response pathways were mutated in response to several different antibiotic stresses (Table 2). Consistent with prior studies on antibiotic tolerance[30,31], central components of the stringent response (SpoT and SspA) were occasional targets of selection. Antioxidant stress response (SoxR and AhpF)[32] and production of antioxidant molecules[33], such as putrescine and spermidine, were also selected under antibiotic selection (Supplementary Data 3). In response to DNA-damaging antibiotic stress, populations mutated members of the SOS regulon (*dinB*, *yafO* and *yafP*) and cryptic prophages (cryptic prophage CP4-44). Indeed, prophages provide enhanced survival of the bacterial host in times of antibiotic stress[34].

Cross-resistance and parallel molecular evolution are linked. Despite differences in antibiotic selection pressure, parallel evolution was prevalent at multiple levels. This pattern is very unlikely to reflect adaptation unrelated to antibiotic treatment, as such parallel mutations generally incurred a fitness cost in antibiotic-free medium (see below). We hypothesized that parallel evolving mutations have an important contribution to the observed cross-resistance interactions. To investigate this issue, we calculated the average fraction of mutated genes shared by two strains for each pair of antibiotics (Fig. 3b).

Adaptation to certain antibiotics proceeds through diverse combinations of mutations (for example, on average, pairs of nitrofurantoin-adapted strains show 16.5% overlap in their sets of mutated genes), while the number of evolutionary trajectories appear to be more limited in other cases (for example, the same figure for chloramphenicol is 38%). Antibiotic pairs that have an especially low overlap in the corresponding sets of their mutated genes rarely displayed cross-resistance (Fig. 3c; $P < 10^{-10}$, $N = 66$, Wilcoxon rank-sum test when pairs with a mutation profile similarity of < 0.01 were compared with the rest). This pattern can be largely, but not exclusively, attributed to aminoglycosides: the sets of genes mutated under aminoglycoside selection pressure displayed practically no overlap with those detected in other laboratory-evolved lines (Fig. 3b), and cross-resistance was also absent. However, the association between low mutational overlap and scarcity of cross-resistance remains even when aminoglycosides are excluded from the analysis ($P < 0.005$, $N = 45$, Wilcoxon rank-sum test).

To investigate the role of parallel evolving mutations in cross-resistance further, we selected seven genes for further characterization, all of which were mutated in multiple laboratory-evolved

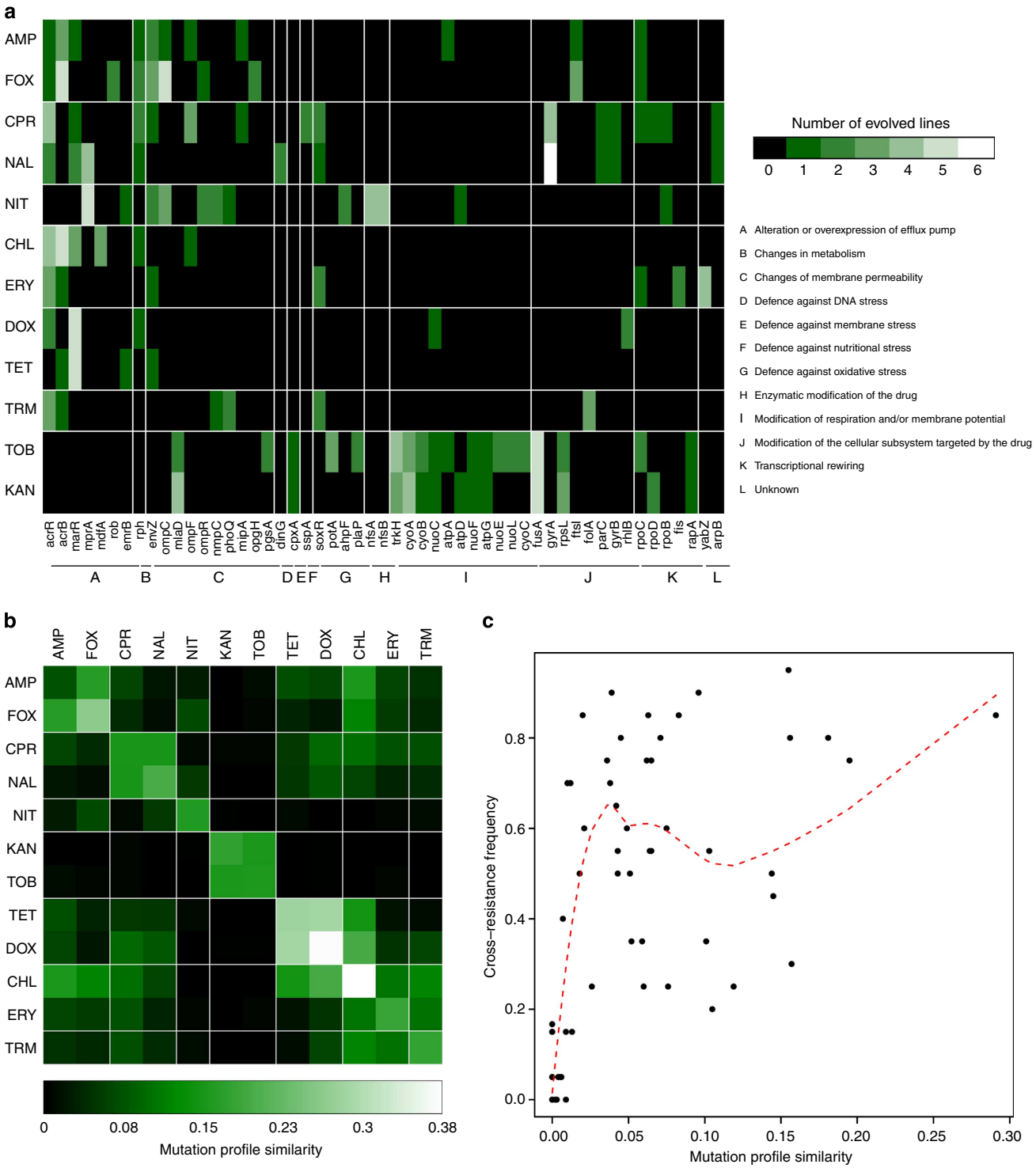


Figure 3 | Parallel evolution and cross-resistance. (**a**) Mutational profiles of the 12 antibiotic selection regimes. Only those genes are shown that mutated in two or more of the 61 sequenced non-mutator laboratory-evolved lines. Mutations in promoters of multi-genic operons were associated with all genes encoded by the operon. The colour code indicates the number of cases when the same gene was independently mutated in different lines evolved under the same antibiotic pressure. (**b**) Heatmap of the average mutation profile similarity of two strains adapted to different (off-diagonal elements) and identical (diagonal elements) antibiotics. Mutation profile similarity between each pair of evolved lines was estimated by the Jaccard's coefficient between their sets of mutated genes. Note that the map is symmetric. (**c**) Very-low average mutation profile similarities between strains adapted to different antibiotics are associated with low cross-resistance frequencies between antibiotic pairs. Mutation profile similarity was calculated as in **b**. Antibiotic pairs with mutation profile similarities <0.01 show significantly lower cross-resistance frequencies than the rest of the pairs ($P<10^{-10}$, $N=66$, Wilcoxon rank-sum test), even when aminoglycosides are excluded ($P<0.005$, $N=45$). Dashed red curve indicates a smooth curve fitted by Loess regression[56] (using the local polynomial regression fitting function of R).

Table 2 | Map of repeatedly mutating functional units across antibiotic treatments.

Functional category	Cell wall		Gyrase		Multiple	50s		30s		Folic acid	Aminoglycoside	
	AMP	FOX	CPR	NAL	NIT	CHL	ERY	DOX	TET	TRM	TOB	KAN
Alteration or overexpression of efflux pump	6	10	6	9	9	16	5	7	7	4	1	3
Changes in metabolism	3	2	3	3	5	1	0	1	1	0	2	6
Changes of membrane permeability	6	12	6	0	13	1	3	0	1	3	6	4
Defence against DNA stress	0	1	1	2	0	0	0	0	0	1	0	0
Defence against Membrane stress	0	0	0	0	1	0	0	0	0	0	1	1
Defence against Nutritional stress	0	0	2	0	1	0	0	0	0	0	0	0
Defence against Oxidative stress	0	0	3	1	2	0	2	0	0	2	7	3
Enzymatic modification of the drug	1	0	0	0	8	0	0	0	0	0	0	0
Modification of respiration and/or membrane potential	2	1	0	1	3	0	1	1	0	0	32	15
Modification of the cellular subsystem targeted by the drug	1	4	7	14	0	0	0	2	0	5	14	10
Prophage activation	0	0	1	1	0	1	0	0	0	0	1	1
Transcriptional rewiring	1	1	3	1	1	1	3	0	1	1	3	5

The numbers indicate the total sum of independent mutational events found in lines adapted to a given antibiotic.

Table 3 | Selected individual mutations and their sensitivity profiles across antibiotics.

Gene	Amino acid. change	Relative fitness in antibiotic free medium (± s.e.)	Relative MIC change											
			Cell wall		Gyrase		Multiple	50s		30s		Folic acid	Aminoglycoside	
			AMP	FOX	CPR	NAL	NIT	CHL	ERY	DOX	TET	TRM	TOB	KAN
mprA	Arg110Leu	0.99±0.016	1.0	1.0	0.8	3.1	1.4	1.0	1.5	0.9	1.0	1.0	0.8	0.9
marR	Val84Glu	0.95±0.008*	2.0	3.3	1.9	2.1	1.0	2.2	2.0	1.9	1.8	1.3	1.0	1.0
envZ	Ala396Thr	0.90±0.007*	1.7	2.7	2.2	1.0	0.9	1.0	1.2	0.8	1.3	0.8	0.7	0.6
envZ	Val241Gly	0.87±0.030*	2.6	2.7	2.6	0.8	1.0	1.5	1.6	0.5	1.0	0.8	0.7	0.6
soxR	Leu139*	0.72±0.023*	1.2	1.0	1.3	2.2	0.6	1.1	4.7	0.6	1.1	1.8	1.9	3.6
phoQ	Gly384Cys	0.94±0.032*	2.0	1.7	1.3	0.7	1.0	1.1	1.4	1.0	1.3	2.3	1.4	1.4
trkH	Thr350Lys	0.57±0.011*	0.5	0.8	0.6	0.3	0.8	0.6	0.9	0.5	0.5	0.6	3.4	2.1
gyrA	Ser83Leu	1.02±0.025	1.0	1.0	7.7	30.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
gyrA	Asp87Gly	1.04±0.006	1.8	1.3	7.7	30.1	0.9	1.0	1.0	0.8	1.0	1.0	1.4	1.9

TrkH antibiotic sensitivity data was partly based on results of a previous paper[7]. Relative fitness values are presented with the corresponding s.e. values.
*stands for cases of significance ($P<0.05$, $N=14$, *t*-test).

lines and cover a wide range of molecular functions. The selected mutations were inserted individually into wild-type *E. coli*. The mutations generally conferred mild, but significant declines in susceptibilities to several antibiotics (Table 3). For example, a mutation in PhoQ, a member of the two-component regulatory system involved in acid and low Mg^{2+} stress tolerance[35], increased resistance both to cell wall inhibitors and to the folic acid inhibitor trimethoprim. Beyond their beneficial effects, the

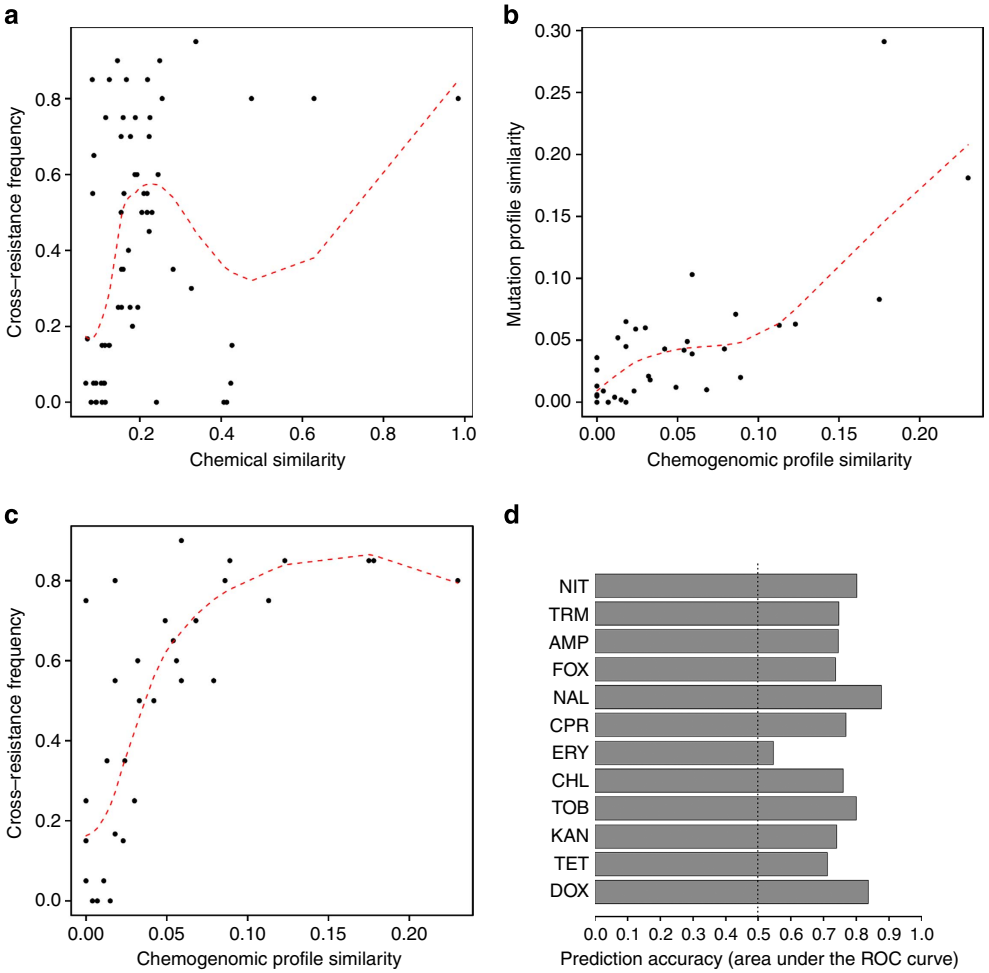


Figure 4 | Antibiotic properties and cross-resistance. (**a**) Weak association between chemical structural similarity between antibiotic pairs and cross-resistance frequency (Spearman's $\rho = 0.40$, $P < 10^{-3}$, $N = 66$), which disappears when aminoglycosides are excluded ($\rho = 0.21$, $P = 0.18$, $N = 45$). Structural similarity between antibiotics was estimated by the Tanimoto similarity of their molecular fingerprints. (**b**) Correlation between chemogenomic profile similarity and overlap in the set of accumulated mutations during laboratory evolution (Spearman's $\rho = 0.67$, $P < 10^{-5}$, $N = 36$). (**c**) Antibiotic pairs that frequently display cross-resistance interactions show relatively high overlap in their chemogenomic profiles (Spearman's $\rho = 0.77$, $P < 10^{-7}$, $N = 36$). Dashed red curves on scatterplots A–C indicate smooth curves fitted by Loess regression[56]. (**d**) Predicting antibiotic resistance phenotypes from genome sequences. Prediction performance for each antibiotic based on the set of accumulated mutations was measured by the area under the receiver operating characteristic (ROC) curve (AUC). This gives an overall measure of accuracy by taking into account both true positive and false positive rates across all possible cutoffs of the prediction score. Random prediction gives an AUC of 0.5. Variation in resistance among evolved strains can be predicted with 55–88% (76% average) accuracy, depending on the antibiotic studied. Special care was taken to avoid circularity in the predictions.

selected mutations frequently had significant fitness costs in antibiotic-free medium (ref. 17) and enhanced sensitivity to certain antimicrobial agents (Table 3). The cross-resistance patterns conferred by individual mutations and the corresponding laboratory-evolved lines showed 62% overlap (45% would be expected by chance, randomization test, $P = 0.002$, $N = 144$ and Supplementary Data 7).

Antibiotic features and cross-resistance patterns. By compiling a data set on the chemical and functional properties of antibiotics, we next examined the extent to which similarities in individual antibiotic properties shape the distribution of cross-resistance interactions in the network. One might argue that cross-resistance occurs mainly between antibiotics that target the same cellular subsystems. However, target mutations were present in less than half of the evolved lines and 88% of the cross-resistance interactions occurred between antibiotics with different cellular targets.

Relatedness of chemical structures (as captured by chemical fingerprint similarities as measured by the Tanimoto coefficient[36]) emerges as a weak predictor of antibiotic cross-resistance (Spearman's $\rho = 0.4$, $P < 10^{-3}$, $N = 66$, Fig. 4a). Furthermore, this marginal effect is entirely attributable to aminoglycosides, which have low chemical similarity with other antibiotics and rarely show cross-resistance interactions with them (Spearman's $\rho = 0.21$, $P = 0.17$, $N = 45$ when excluding aminoglycosides).

Last, the intrinsic resistome (that is, the set of genes that influence antibiotic sensitivity) provides an unbiased description of antibiotic action[37]. We, therefore, asked how the overlap in the intrinsic resistome shapes the distribution of cross-resistance interactions. Our molecular and phenotypic results were integrated with data from a previous chemogenomic screen[20]. That study exposed a nearly complete mutagenized *E. coli* library to each of 17 antibiotics and determined the fitness contribution of individual genes. Using this data set, we calculated the sets of

genes that influence susceptibility for each antibiotic used in our study (chemogenomic profile). Strikingly, antibiotic pairs that showed substantial overlap in their chemogenomic profiles also accumulated similar sets of mutations during the course of laboratory evolution (Spearman's $\rho = 0.67$, $P < 10^{-5}$, $N = 36$; Fig. 4b), and frequently displayed cross-resistance interactions (Spearman's $\rho = 0.78$, $P < 10^{-7}$, $N = 36$; Fig. 4c). Importantly, these results remained when excluding antibiotic pairs targeting the same subsystem (Spearman's $\rho = 0.59$, $P < 10^{-3}$, $N = 33$ and Spearman's $\rho = 0.73$, $P < 10^{-5}$, $N = 33$, respectively) or those involving aminoglycosides (Spearman's $\rho = 0.57$, $P < 0.005$, $N = 28$ and Spearman's $\rho = 0.75$, $P < 10^{-5}$, $N = 28$, respectively).

Mutational analysis captures antibiotic resistance profile. Our data indicate that the molecular mechanisms of antibiotic resistance evolve in a repeatable manner. This raises the question whether it is possible to predict antibiotic resistance phenotypes from the genome sequences of the laboratory-evolved lines. We employed a simple procedure that uses gene sets derived from our sequenced evolved lines to predict differences in resistance phenotypes among individual strains. Briefly, for each antibiotic, we compiled the complete list of genes that were mutated at least once in lines evolved under the given antibiotic selection pressure. This gene–antibiotic association set was compared with the set of genes mutated in each strain with known antibiotic resistance profile, resulting in a set of 12 scores measuring the likelihood of resistance of the evolved line against the complete panel of 12 antibiotics. The above procedure was repeated for each of our 61 sequenced evolved lines in turn. To quantify the agreement between this simple prediction score against experimentally determined resistance profiles (that is, increased resistance compared with wild-type), we used a combined measure of sensitivity (true positive rate) and specificity (true negative rate)[38]. In particular, we measured how accurately our prediction procedure separates resistance and sensitivity to a given antibiotic when averaged across all 61 evolved lines. The analyses demonstrated that variation in antibiotic resistance among evolved strains can be predicted with an average 76% accuracy, while only 50% would be expected by chance (Fig. 4d and Supplementary Fig. 3). For example, the method is able to discriminate doxycycline-resistant and sensitive strains with 84% accuracy. We emphasize that our attempt to predict resistance profiles is preliminary and future works should investigate whether incorporation of more antibiotics, a greater diversity of genomes or usage of more refined prediction algorithms could improve prediction success.

Discussion

By combining experimental evolution, genome sequencing and functional analyses, this work charted a map of cross-resistance interactions between antibiotics in *E. coli*, and explored, on a genome-wide scale, the mechanisms driving these evolutionary patterns. The following general conclusions can be drawn from our study.

First, our work indicates that the progressive accumulation of spontaneous mutations under antibiotic selection simultaneously changes the organism's sensitivity to many other antimicrobial agents (Fig. 1a). It also revealed differences in the efficacy by which different antibiotics can inhibit growth of resistant bacterial populations or select for the emergence of multidrug-resistant strains (Fig. 1c). Cross-resistance between two antibiotics was largely independent of whether they show synergistic effects in combination[12,39]. Thus, the networks based on evolutionary and physiological antibiotic interactions are generally governed by distinct mechanisms. As both synergism and cross-resistance interactions between antibiotic pairs can potentially influence long-term evolutionary pathways[4], combination of these two maps could be especially informative for future development of novel antimicrobial strategies.

Second, a strong signature of parallel evolution emerged across populations adapted to different antibiotics (Table 2), although the molecular mechanisms underlying antibiotic resistance and cross-resistance were diverse. Our work identified several genes where the observed mutations delivered resistance to multiple antimicrobial agents (Supplementary Table 3). In several instances (*phoQ*, *envZ*, *soxR* and *trkH*), the potential roles of these genes in multidrug resistance are yet to be investigated in the clinic. Unexpectedly, even a mutation in the molecular target of the antibiotic can alter sensitivity to multiple, unrelated antibiotics. Laboratory-evolved fluoroquinolone resistant lines frequently exhibited a specific mutation in the target topoisomerase gene (*gyrA*: A87G). This single mutation influenced sensitivity to several non-quinolone drugs (Table 3), probably through altering patterns of supercoiling and hence global expression of stress response pathways[40]. Strikingly, in several instances, individual mutations simultaneously enhanced sensitivity to other drugs (Table 3), indicating that negative trade-offs (collateral sensitivity interactions) are prevalent during antibiotic selection[6–8,41–43]. More generally, the presence of parallel mutations allowed us to predict the resistance profiles of evolved lines from their genome sequence based on catalogues of genes mutated under different antibiotic selection pressures.

Third, as high as 27% of the observed mutations generated proteins with compromised or no activities (Fig. 2b). While potential roles of loss-of-function mutations during antibiotic evolution have been suggested[22,44,45], our work provides the first estimate on the relative importance of this mutational class. Given their high rates and potential beneficial effects, loss-of-function mutations could play an important role during the early stage of resistance evolution (see also ref. 46).

Fourth, chemogenomic profile similarity between antibiotics emerges as the most significant determinant of cross-resistance (Fig. 4c). Thus, beyond their pivotal role in elucidating the mechanisms of drug actions[47], systematic chemogenomic studies could also be used in the future to infer general trends of resistance evolution.

Taken together, our analyses indicate that resistance evolution is governed by highly pleiotropic mutations in a relatively limited set of functional modules. The prevalence of mutations with pleiotropic effects indicates that the phenomenon of cross-protection may be more general and extend to other stressful conditions unrelated to antibiotic pressure[48]. Indeed, genes mutated in our study were enriched in the set of *E. coli* genes that influence sensitivity to toxic metal (for example, copper and nickel) and detergent exposure (Supplementary Table 4). Given the documented associations between levels of metal contamination and specific patterns of antibiotic tolerance in nature[49], future evolutionary studies should investigate how frequently metal and antibiotic resistance are co-selected in the laboratory. It will also be important to establish to what extent cross-resistance interactions remain conserved across (pathogenic) species or depend on the introduction of novel genes by horizontal transfer. As most laboratory-evolved lines displayed relatively low fitness in antibiotic-free medium, it will also be important to establish the extent to which adaptation through compensatory mutations can mitigate the costs of resistance.

More generally, understanding the fitness consequences of genetic adaptations to different environments remains an important challenge for evolutionary biology[1]. Thanks to the recent availability of the necessary computational tools and

experimental techniques, it has become possible to predict certain aspects of genomic evolution[50]. Integrating experimental evolution, systems biology and genomics in a framework similar to that presented in this paper could result in the inference of general rules underlying the evolutionary trade-offs observed in nature.

Methods

Laboratory evolutionary experiment. Details of the laboratory evolution experiments have been described elsewhere[7]. Briefly, populations of *E. coli* K12 were grown in MS-minimal medium supplemented with glucose, casamino acids and 1 of 12 possible antibiotics. Parallel cultures were propagated in 96-well microtiter plates. Bacterial cells were transferred every 24 h by inoculating ~1% of the culture to 100 µl fresh medium. Starting with a subinhibitory (IC50) antibiotic concentration, antibiotic dosage was increased gradually (1.5 times the previous dosage) at every fourth transfer. We propagated 96 independent populations in the presence of each antibiotic up to ~336 generations. As expected, population sizes differed significantly across treatments and antibiotic dosages, reflecting independent evolutionary trajectories. For each antibiotic, the experiment halted at the last antibiotic dosage that permitted growth of at least 10 out of 96 parallel evolving populations (criteria was defined as the failure to obtain growth OD 600 < 0.05) or when the antibiotic concentration had reached its upper solubility limit (Supplementary Table 5). For each antibiotic, 10 populations with the highest final cell densities were selected for further analysis, resulting in 120 parallel evolved lines. We also established 10 parallel populations growing in an environment devoid of antibiotics for the same number of transfers, referred to as adapted control lines.

Measurement of antibiotic susceptibilities. Given a panel of resistant strains, our next goal was to detect changes in their sensitivity towards other antimicrobial agents. We developed a highly sensitive high-throughput screening and a robust statistical methodology[7]. Briefly, we tested the susceptibility of each evolved and control lines against the entire set of antibiotics by measuring their growth in liquid cultures at sublethal doses of antibiotics. Bacterial growth was monitored by measuring optical density (OD 600) of the liquid cultures at a single time point after 14 h of growth[7]. Prior experiments demonstrated that a single reading of optical density shows very strong linear correlation with the area under the growth curve[7].

To identify statistically significant cross-resistance interactions, we tested whether each evolved line showed a significant growth difference from the set of 10 control lines. To do this, for each evolved line, we calculated the median value of the four technical replicates and compared it with the distribution of the median growth values of the four technical replicates of the 10 control lines using a Z-test. This yielded a matrix of evolutionary interactions between evolved strains and antibiotics (Supplementary Data 1). Where multiple independent experimental runs were available, we used Fisher's method to aggregate *P*-values. All statistical analyses were carried out in Matlab. The results were confirmed by *E*-test assays, using standard protocols. Finally we calculated the the frequency of cross-resistance (FCR) for each antibiotic pair as follows: FCR = (NA → B + NB → A)/(NA + NB), where NA → B and NB → A are the number of populations adapted to antibiotic A showing enhanced resistance to B, and *vice versa*. NA and NB are the total number of populations adapted to antibiotic A and B, respectively.

Chemical and chemogenomic profile similarities. Chemical similarities of antibiotics were computed using an R implementation of the cheminformatics library CDK (Chemistry Development Kit)[51]. Chemical relatedness was captured by chemical fingerprint similarity as measured by the standard Tanimoto coefficient[52]. Chemogenomic similarity was calculated as pair-wise Jaccard similarity coefficient between sets of genes that influence antibiotic susceptibility based on a published chemogenomic screen[20]. This chemogenomic screen covered 9 of the 12 antibiotics employed in our study, and as it relied on a highly sensitive competition assay, it was particularly useful to identify genes whose inactivation increased antibiotic tolerance. MIC and dose response curve measurements were performed as described previously[7].

Whole-genome sequencing. The ancestral and 63 selected evolved strains were subjected to next-generation sequencing to identify mutations. Genomic DNA (gDNA) was extracted from selected *E. coli* strains (SIGMA GenElute Bacterial Genomic DNA kit) and the subsequent library preparation was performed using the 5500 SOLiD Fragment Library Core Kit (Life Technologies; LT). Briefly, 3 µg of purified bacterial gDNA was fragmented by Covaris S2 System to 100–250 bp. The fragmented DNA was end-repaired and ligated to P1 and P2 adaptors, which provide the primary sequences for both amplification and sequencing of the sample library fragments; the P2 adaptor contains a 10-bp barcode sequence that provided the basis for multiplex sequencing (5500 SOLiD Fragment Library Barcode Adaptors; LT). The templates were size-selected using Agencourt AMPure XP system (Beckman Coulter), nick-translated using Platinum PCR Amplification Mix and the template library was quantified by qPCR using SOLiD Library TaqMan Quantitation Kit (LT). The templates were clonally amplified by emulsion PCR (ePCR) with P1 primer covalently attached to the bead surface. Emulsions were broken with butanol and ePCR beads enriched for template-positive beads by hybridization with P2-coated capture beads. Template-enriched beads were extended at the 3′ end in the presence of terminal transferase and 3' bead linker. Beads with clonally amplified DNA were then deposited onto a SOLiD Flowchip and the slide was loaded into a SOLiD 5500xl System (LT) and the 50-base sequences were obtained according to the manufacturer's protocol.

Bioinformatic analysis of genome sequences. The obtained sequences from each strain were first trimmed in order to filter out low-quality reads that were shorter than 50 bp. The remaining high quality sequences from each strain were then aligned to the *E. coli* K-12 substr. MG1655 chromosome (GenBank Accession No. NC000913; Version NC_000913.2 GI:49175990) in colour space using Genomics Workbench 6.5 (CLC Bio). Within a single read, the maximum gap and mismatch count was set to two and the similarity fraction was set to 0.8. Two mappings were performed for each strain which differed in setting the length fraction to 0.5 for relaxed or 0.6 for stringent analysis. Minimum coverage of ≥51-fold and ≥44-fold was accomplished for each strain when using relaxed or stringent parameters, respectively. A minimum of six reads were required to call a point mutation or short indel (<15 bp) upon relaxed analysis; in contrast, 20 reads were required to call a structural variation (SV; for example, inversion, duplication, replacement, translocation) upon stringent analysis.

For quality-based variant detection we used an approach based on the Neighbourhood Quality Standard algorithm that is implemented in Genomics Workbench. Relaxed alignment was used to identify point mutations or short indels; the minimum variant frequency was set to 50%. Variants identified in the ancestral genome were excluded from further analyses. All remaining potential variants were manually checked with a visual output in order to exclude false variant calls due to insufficient mapping accuracy.

The soft-clipped, unaligned ends of the sequence reads were used to map SVs and long indels. For this, stringent alignment was used and the resulting self-mapped, cross-mapped, multiple, close and paired breakpoints (for details see http://www.clcsupport.com/clcgenomicsworkbench/current/) were identified and manually checked; indels and SVs identified in the ancestral genome were again excluded. All identified breakpoints were validated by re-mapping: consensus sequence resulting from large indel or SV was extracted, re-mapping was performed using stringent setup and the breakpoint considered valid if perfectly matching sequence tags overlapped the breakpoint.

Validation of whole-genome sequencing data. Several structural variants were randomly chosen and validated by either PCR followed by Sanger sequencing (for example, point mutations, deletions and inversions) or by quantitative PCR (for example, duplications). For this latter, DNA levels were determined using StepOne Plus Real-Time PCR system (LT). Reactions were performed by using Power SybrGreen Master Mix (LT); the primer sequences are available on request. All of the measurements were performed in duplicates; the ratio of each amplicon relative to the normalizing control was calculated using the $2^{-\Delta\Delta CT}$ method.

Allele replacements. Allele replacements were constructed by a suicide plasmid-based method. Standard steps and plasmids (pST76-A, pSTKST) of the procedure were described previously[53]. In brief, an ~800-bp long targeting DNA fragment carrying the desired point mutation in the middle was synthesized by PCR, then cloned into a thermosensitive suicide plasmid. The plasmid construct was then transformed into the cell, where it was able to integrate into the chromosome by way of a single crossover between the mutant allele and the corresponding chromosomal region. The desired cointegrates were selected by the antibiotic resistence carried on the plasmid at a nonpermissive temperature for plasmid replication. Next, the pSTKST helper plasmid was transformed, then induced within the cells, resulting in the expression of the I-SceI meganuclease enzyme, which cleaves the chromosome at the 18-bp recognition site present on the integrated plasmid. The resulting chromosomal gap is repaired by way of RecA-mediated intramolecular recombination between the homologous segments in the vicinity of the broken ends. The recombinational repair results in either a reversion to the wild-type chromosome, or in a markerless allele replacement, which can be distinguished by sequencing of the given chromosomal region. For all primers, see Supplementary Table 6.

As other methods failed, the oligonucleotide-mediated λ Red recombination was used to generate the *gyrA* variant S83 → L and D87 → G in *E. coli* BW25113. The applied wild-type strain contained the pBADαβγ λ Red expression plasmid for inducible λ Red recombinase production. Oligonucleotides for allelic replacement were designed according to standard guidelines[54]. Briefly, oligos applied for allelic replacement have complementary sequences to the replicating lagging strand and have minimized secondary structure (less than -12 kcal mol^{-1}). Additionally, each oligo contained two subsequent phosphorothioate linkages at both 5′ and 3′ termini for endogenous nuclease evasion. Oligos were ordered with standard purification and desalting from Integrated DNA Technologies (IDT). To perform allelic replacement, cells were grown in 10 ml Luria Bertani (LB) broth,

supplemented with 100 μg ml^{-1} ampicillin, from overnight starter culture at 37 °C, 250 r.p.m. to OD_{550} 0.5–0.7. Expression of λ Red proteins were induced by the addition of L-Arabinose at 0.2% concentration for 30 min. For recombination, cells were pelleted (3,800 r.p.m. for 7 min) and washed twice in ice-cold dH_2O, resuspended in 160 μl dH_2O. 40 μl cell suspension was electroporated with oligo GyrAS83L or GyrAD87G at 2.5 μM final concentration. Electroporated cells were allowed to recover in 10 ml LB at 37 °C overnight. Cells were plated on LB agar plates supplemented with 100 ng ml^{-1} ciprofloxacin. Clones with desired mutation were identified by sequencing target site in *gyrA* using GyrA2F and GyrA2R primers.

Mutation rate measurements. Mutation rates of two laboratory-evolved lines (AMP6, CPR6) were measured by using rifampicin (Rif^S to Rif^R) forward fluctuation test. The rifampicin minimum inhibitory concentration (MIC) for the two evolved lines does not differ from that of the control line. Overnight cultures (grown in LB broth, on 30 °C) were diluted to 10^4 cells per ml and six parallel cultures per each line were started in glass tubes. After 24 h incubation at 30 °C, appropriate dilutions were plated to LB agar plates for CFU determination, and to LB agar plates containing 100 μg ml^{-1} rifampicin for detection of rifampicin resistant mutants. Colonies were counted after 24 and 48 h, respectively. Mutation rates were calculated by using the MSS maximum-likelihood method[55].

Predicting antibiotic resistance phenotypes from genomic data. To predict antibiotic resistance phenotypes from genome sequences of the evolved lines, we employed a procedure that uses gene sets derived from our sequenced evolved lines to predict differences in resistance phenotypes among individual genomes. First, for each antibiotic, we compiled the list of genes that were mutated in at least one of our lines evolved under the given antibiotic selection pressure (for example, genes mutated in ampicillin-evolved lines for ampicillin). To avoid circularity in the predictions, these gene–antibiotic association lists were defined by leaving out the genome (G_x) for which resistance prediction was attempted (that is, yielding slightly different association lists for each G_x). Next, for each antibiotic, we counted the number of protein-coding genes that are both mutated in G_x and present in the gene–antibiotic association list of the given antibiotic. This procedure results in a set of 12 scores measuring the likelihood of resistance of evolved line G_x against our panel of 12 antibiotics. Finally, the above procedure was repeated for each of our 61 sequenced evolved lines in turn. To quantify the agreement between this simple prediction score against experimentally determined resistance profiles (that is, increased resistance compared to wild-type), we used a combined measure of sensitivity (true positive rate) and specificity (true negative rate)[38]. In particular, we measured how accurately our prediction procedure separates resistance and sensitivity to a given antibiotic when averaged across all 61 evolved lines. We note that not all gene–antibiotic association lists were equally informative in the prediction process as mutations occurring in aminoglycoside-evolved lines were especially relevant to discriminate between the presence and absence of resistance to a number of antibiotics (Supplementary Table 7). This is unsurprising given the distinct mutational profiles and resistance mechanisms of aminoglycoside-adapted lines.

References

1. Cooper, V. S. & Lenski, R. E. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407,** 736–739 (2000).
2. MacLean, R. C., Bell, G. & Rainey, P. B. The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. *Proc. Natl Acad. Sci. USA* **101,** 8072–8077 (2004).
3. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19,** 207–233 (1988).
4. Palmer, A. C. & Kishony, R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat. Rev. Genet.* **14,** 243–248 (2013).
5. Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44,** 101–105 (2012).
6. Szybalski, W. & Bryson, V. Genetic studies on microbial cross resistance to toxic agents. I. Cross resistance of *Escherichia coli* to fifteen antibiotics. *J. Bacteriol.* **64,** 489–499 (1952).
7. Lazar, V. *et al.* Bacterial evolution of antibiotic hypersensitivity. *Mol. Syst. Biol.* **9,** 700 (2013).
8. Imamovic, L. & Sommer, M. O. Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. *Sci. Transl. Med.* **5,** 204ra132 (2013).
9. Taber, H. W., Mueller, J. P., Miller, P. F. & Arrow, A. S. Bacterial uptake of aminoglycoside antibiotics. *Microbiol. Rev.* **51,** 439–457 (1987).
10. Chait, R., Craney, A. & Kishony, R. Antibiotic interactions that select against resistance. *Nature* **446,** 668–671 (2007).
11. Bliss, C. I. The toxicity of poisons applied jointly. *Ann. Appl. Biol.* **26,** 585–615 (1939).
12. Yeh, P., Tschumi, A. I. & Kishony, R. Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* **38,** 489–494 (2006).
13. Oliver, A., Canton, R., Campo, P., Baquero, F. & Blazquez, J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288,** 1251–1254 (2000).
14. Wiegand, I. *et al.* Mutator genes giving rise to decreased antibiotic susceptibility in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **52,** 3810–3813 (2008).
15. Denamur, E. *et al.* Intermediate mutation frequencies favor evolution of multidrug resistance in *Escherichia coli*. *Genetics* **171,** 825–827 (2005).
16. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57,** 3348–3357 (2013).
17. Andersson, D. I. The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr. Opin. Microbiol.* **9,** 461–465 (2006).
18. Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335,** 457–461 (2012).
19. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109,** E2774–E2783 (2012).
20. Girgis, H. S., Hottes, A. K. & Tavazoie, S. Genetic architecture of intrinsic antibiotic susceptibility. *PLoS ONE* **4,** e5629 (2009).
21. McCalla, D. R., Kaiser, C. & Green, M. H. Genetics of nitrofurazone resistance in *Escherichia coli*. *J. Bacteriol.* **133,** 10–16 (1978).
22. Okusu, H., Ma, D. & Nikaido, H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J. Bacteriol.* **178,** 306–308 (1996).
23. Alekshun, M. N. & Levy, S. B. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol.* **7,** 410–413 (1999).
24. Piddock, L. J. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clin. Microbiol. Rev.* **19,** 382–402 (2006).
25. Alekshun, M. N. & Levy, S. B. Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128,** 1037–1050 (2007).
26. Stephan, J., Mailaender, C., Etienne, G., Daffe, M. & Niederweis, M. Multidrug resistance of a porin deletion mutant of *Mycobacterium smegmatis*. *Antimicrob. Agents Chemother.* **48,** 4163–4170 (2004).
27. Proctor, R. A. *et al.* Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat. Rev. Microbiol.* **4,** 295–305 (2006).
28. Cui, L. *et al.* An RpoB mutation confers dual heteroresistance to daptomycin and vancomycin in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **54,** 5222–5233 (2010).
29. Hansen, S., Lewis, K. & Vulic, M. Role of global regulators and nucleotide metabolism in antibiotic tolerance in *Escherichia coli*. *Antimicrob. Agents Chemother.* **52,** 2718–2726 (2008).
30. Nguyen, D. *et al.* Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science* **334,** 982–986 (2011).
31. Gilbert, P., Collier, P. J. & Brown, M. R. Influence of growth rate on susceptibility to antimicrobial agents: biofilms, cell cycle, dormancy, and stringent response. *Antimicrob. Agents Chemother.* **34,** 1865–1868 (1990).
32. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.* **8,** 423–435 (2010).
33. Tkachenko, A. G., Akhova, A. V., Shumkov, M. S. & Nesterova, L. Y. Polyamines reduce oxidative stress in *Escherichia coli* cells exposed to bactericidal antibiotics. *Res. Microbiol.* **163,** 83–91 (2012).
34. Wang, X. *et al.* Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* **1,** 147 (2010).
35. Zwir, I. *et al.* Dissecting the PhoP regulatory network of *Escherichia coli* and *Salmonella enterica*. *Proc. Natl Acad. Sci. USA* **102,** 2862–2867 (2005).
36. Leach, A. R. & Gillet, V. J. *An introduction to chemoinformatics* (Springer, 2007).
37. Jansen, G. *et al.* Chemogenomic profiling predicts antifungal synergies. *Mol. Syst. Biol.* **5,** 338 (2009).
38. Hanley, J. A. & Mcneil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143,** 29–36 (1982).
39. Klein, M. & Schorr, S. E. The role of bacterial resistance in antibiotic synergism and antagonism. *J. Bacteriol.* **65,** 454–465 (1953).
40. Webber, M. A. *et al.* Clinically relevant mutant DNA gyrase alters supercoiling, changes the transcriptome, and confers multidrug resistance. *MBio* **4,** e00273–00213 (2013).
41. Perichon, B. & Courvalin, P. Synergism between beta-lactams and glycopeptides against VanA-type methicillin-resistant *Staphylococcus aureus* and heterologous expression of the vanA operon. *Antimicrob. Agents Chemother.* **50,** 3622–3630 (2006).
42. Macvanin, M. & Hughes, D. Hyper-susceptibility of a fusidic acid-resistant mutant of Salmonella to different classes of antibiotics. *FEMS Microbiol. Lett.* **247,** 215–220 (2005).
43. Goulart, C. P. *et al.* Designing antibiotic cycling strategies by determining and understanding local adaptive landscapes. *PLoS ONE* **8,** e56040 (2013).

44. Jeong, J. Y. *et al.* Sequential inactivation of rdxA (HP0954) and frxA (HP0642) nitroreductase genes causes moderate and high-level metronidazole resistance in *Helicobacter pylori*. *J. Bacteriol.* **182,** 5082–5090 (2000).
45. McKeegan, K. S., Borges-Walmsley, M. I. & Walmsley, A. R. Microbial and viral drug resistance mechanisms. *Trends Microbiol.* **10,** S8–S14 (2002).
46. Hottes, A. K. *et al.* Bacterial adaptation through loss of function. *PLoS Genet.* **9,** e1003617 (2013).
47. Kapitzky, L. *et al.* Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action. *Mol. Syst. Biol.* **6,** 451 (2010).
48. Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J. & Tagkopoulos, I. Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in *Escherichia coli*. *Mol. Syst. Biol.* **9,** 643 (2013).
49. Baker-Austin, C., Wright, M. S., Stepanauskas, R. & McArthur, J. V. Co-selection of antibiotic and metal resistance. *Trends Microbiol.* **14,** 176–182 (2006).
50. Papp, B., Notebaart, R. A. & Pal, C. Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* **12,** 591–602 (2011).
51. Steinbeck, C. *et al.* The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* **43,** 493–500 (2003).
52. Walsh, C. T. *Antibiotics: Actions, Origins, Resistance* (ASM press, 2003).
53. Feher, T. *et al.* Scarless engineering of the *Escherichia coli* genome. *Methods Mol. Biol.* **416,** 251–259 (2008).
54. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
55. Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25,** 1564–1565 (2009).
56. Cleveland, W. S., Grosse, E. & Shyu, W. M. *Statistical Models in S.* (eds Chambers, J. M. & Hastie, T. J.) Ch 8 (Wadsworth & Brooks/Cole, 1992).

Acknowledgements

This work was supported by grants from the European Research Council (202591) and the EMBO Young Investigator Programme (to C.P.), the Wellcome Trust, and the 'Lendület Program' of the Hungarian Academy of Sciences (to C.P. and B.P.). B.C. was supported by the Hungarian Research Fund (OTKA PD 109572). We also acknowledge the use of the NGS platform established by the ERC Advanced Grant 'SymBiotics' of Eva Kondorosi.

Author contributions

B.P. and C.P. conceived and supervised the project; V.L., I.N. and R.S. designed the experiments; V.L., I.N., R.S., B.C., Á.N., B.H., A.V., M.H., B.B. and O.M. performed the experiments; V.L., I.N., Á.G., R.B.-F., G.F., B.S., B.K., B.P. and C.P. developed data analysis procedures and interpreted the data; C.P. and P.B. wrote the manuscript with contributions from all other authors.

Additional information

Accession codes: The raw sequences and assemblies have been deposited in NCBI Bioproject database under the accession code PRJNA248327 (accession SRP042209). The individual accession numbers for the 64 deposited samples are as follows: SRR1297006, SRR1297043, SRR1297049, SRR1297054 to SRR1297056, SRR1297060 to SRR1297064, SRR1297067, SRR1297069, SRR1297073, SRR1297077, SRR1297079, SRR1297081, SRR1297096, SRR1297101, SRR1297103 to SRR1297107, SRR1297109, SRR1297112, SRR1297114, SRR1297117, SRR1297123 to SRR1297126, SRR1297129, SRR1297132, SRR1297133, SRR1297135, SRR1297137, SRR1297139, SRR1297142, SRR1297144, SRR1297148, SRR1297150, SRR1297153, SRR1297155, SRR1297157, SRR1297159, SRR1297160, SRR1297163, SRR1297168 to SRR1297173 and SRR1297175 to SRR1297184.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Lázár, V. *et al.* Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nat. Commun.* 5:4352 doi: 10.1038/ncomms5352 (2014).