Evolutionary barriers to horizontal gene transfer in macrophage associated *Salmonella*

Rama P. Bhatia [1], Hande A. Kirit [1,2,3], Cecil M. Lewis Jr [2,3], Krithivasan Sankaranarayanan [2,4], Jonathan P. Bollback [1*]

Affiliations:

[1] Institute of Infection, Veterinary, and Ecological Sciences, Department of Evolution, Ecology, and Behaviour, University of Liverpool, Liverpool, UK

[2] Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, Norman, USA

[3] Department of Anthropology, University of Oklahoma, Norman, OK

[4] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK

Corresponding author

Jonathan P. Bollback

Telephone number: +44 (0)151 795 4526

Email address: J.P.Bollback@liverpool.ac.uk

Abstract

Horizontal gene transfer (HGT) is a powerful evolutionary force facilitating bacterial adaptation and emergence of novel phenotypes. Several factors, including environmental ones, are predicted to restrict HGT, but we lack systematic and experimental data supporting these predictions. Here, we address this gap by measuring the relative fitness of genes horizontally transferred from *E. coli* to *S. enterica* in infection relevant environments. We estimated the distribution of fitness effects in each environment and identified dosage-dependent effects across different environments are a significant barrier to HGT. The majority of genes were found to be deleterious. We also found longer genes had stronger negative fitness consequences than shorter ones, showing that gene length can significantly hinder HGT. Furthermore, fitness effects of transferred genes were found to be environmentally dependent. In summary, a substantial fraction of transferred genes had a significant fitness cost on the recipient, with both gene characteristics and the environment acting as evolutionary barriers to HGT.

Introduction

Horizontal gene transfer (HGT) is the lateral transfer of genetic material between individuals of the same or different species rather than vertically from parent to offspring. HGT plays an important role in bacterial adaptation: e.g., acquisition of accessory genes can confer bacteria with novel phenotypes such as antimicrobial resistance, microbial toxins, and tolerance to heavy metals [1].

Gene acquisition through HGT may exert different fitness effects on the recipient cell, with the nature of these effects determining the evolutionary fate of the newly transferred gene. In the simplest terms, fitness can be defined as reproductive success and is key to understanding bacterial evolution and adaptation [2]. A fitness advantage or disadvantage can be considered at different levels of the biological hierarchy (e.g., genes, transposons, plasmids, cells) that are under selection [3,4,5,6]

Nonetheless, the sole purpose of this study was to explore which of the intrinsic properties or features of a horizontally transferred gene, upon its expression, could impact the fitness/reproductive ability of the host and not to understand fitness costs due to physical barriers to the acquisition and

integration (e.g., success of transmission or probability of homologous or illegitimate recombination) of the transferred genes. In this study, the transferred genes which on expression significantly increased the relative fitness of the host were said to have a beneficial effect; those that caused a significant decrease in the relative fitness were said to have a deleterious fitness effect and those that did not significantly differ from the 'wild type' were classed as neutral.

While genes with a beneficial or neutral effect can persist in bacterial populations, genes with deleterious effects will most likely be purged [7,8]. The relative frequency of these fitness effects, also known as the distribution of fitness effects (DFEs) [9,10] is thus an important predictor of the outcome of HGT events, as it provides insights into the evolutionary fate of a horizontally transferred gene determined by its fitness on the recipient cell. Therefore, understanding the role of factors in defining the DFEs of horizontally transferred genes is critical if we wish to understand bacterial evolution. Collectively, we will refer to these factors as evolutionary barriers to HGT.

Several potential evolutionary barriers to HGT have been predicted using computational approaches like phylogenetic and parametric/compositional analysis [11], including those due to *(i)* gene function: information processing, such as those involved in DNA replication, transcription and translation are less often transferred compared to operational genes [12,13,14], *(ii)* gene interactions: genes with multiple interacting partners (e.g., protein-protein, metabolic or regulatory) are less likely to experience a successful gene transfer event [15,16], *(iii)* deviation in GC content and codon usage: HGT frequencies can be affected by the donor-recipient similarity barrier. Genes acquired from donors of dissimilar GC content and codon usage are more likely to be targeted by anti-HGT systems (e.g., R-M systems, Cas/CRISPR, H-NS) [17,18,19,20,21], and *(iv)* gene dosage/protein dosage: an increase in protein levels due to extra copies of an ortholog can disrupt the stoichiometric balance in the cell [22].

In addition to the intrinsic characteristics of the genes themselves, the environment can substantially alter the fitness effects of horizontally acquired genes. For example, genes that confer antibiotic resistance are beneficial in the presences of the antibiotic but might be deleterious in its absence [23,24]. In fact, multiple studies have shown that the environment may affect the selective outcomes in evolving bacterial populations [25,26,27].

Here, we experimentally test the fitness effects of HGT events, by introducing genes *Escherichia coli* into *Salmonella enterica* serovar Typhimurium strain 4/74 (*S.* Typhimurium 4/74) in four different infection relevant environments. *S.* Typhimurium is an enteric pathogen infecting both humans and animals and can cause non-typhoidal Salmonella disease (NTS), including a virulent invasive form of the disease in sub-Saharan Africa [28,29,30]. The strain used in this study, *S.* Typhimurium 4/74, is a prototrophic bacteria that is highly virulent in cattle, pigs, mice, and chickens [31,32]. *Salmonella* is an intracellular pathogen, and to survive and proliferate within eukaryotic cells, it must adapt to a fluctuating intracellular environment. During the course of evolution, *Salmonella* has acquired virulence encoded *Salmonella* Pathogenicity Islands (SPI) through horizontal gene transfer. Among many other pathogenic properties, SPIs have allowed *Salmonella* to hoodwink host immune surveillance and survive in a replicative niche within a modified phagosome in the host macrophages, known as the *Salmonella*-containing vacuole (SCV) [28].

We tested the consequences on fitness of the 44 transferred genes in environments that mimic key features of the macrophage environment [32]; SPI-2 inducing PCN (phosphate/carbon/nitrogen) medium [33,34] as the base, and further modified it to either contain low magnesium, low oxygen (hypoxic), or an antibiotic challenge using a fluoroquinolone class of drugs – ciprofloxacin, a first-line antimicrobial for many Gram-negative pathogens, including *Salmonella* (see **Supplementary Table 1**).

In contrast to a retrospective computational approach, our experimental approach allows us to test which genes are likely to fail to survive the evolutionary sieve, and thus disentangle the influence of different evolutionary barriers and the role of the environment.

Results

The experimental approach was to artificially transfer 44 orthologs from *E. coli* to *Salmonella* Typhimurium 4/74 using a low copy (3-4/cell) inducible expression vector pZS* [35] in four infection relevant environments (see **Supplementary Table 1**) to study the DFE of horizontally acquired genes. Additionally, we wanted to investigate the extent of the effects of evolutionary barriers – functional

category, gene interactions, GC content, codon usage bias, gene length, protein dosage, and the environment on the persistence of horizontally transferred genes.

S. Typhimurium and *E. coli* were chosen primarily due to their genetic and ecological similarity, and history of HGT [36,37]. Furthermore, the genetic similarity between the donor and recipient species ensures the functioning of the transferred genes in the host background, as well as that the functional interacting partners exist for the majority of genes transferred; this is important as the study aims to understand the role of gene categories and gene interactions in the persistence of horizontally acquired genes. Additionally, the donor and recipient are sufficiently divergent so that the effect of several intrinsic factors of the transferred genes can be evaluated, such as their GC content and length.

DFEs and role of other barriers to HGT

Specifically, we looked at the effect of the functional category, gene interactions, GC content, codon usage bias, and gene length of the transferred orthologs on the fitness of *S*. Typhimurium 4/74 in four environments that mimic the signals encountered during an intracellular infection. We did this through careful and precise measurement of fitness effects from changes in gene frequencies using Next-Generation Sequencing (NGS). A graphical illustration of pooled growth setup has been provided (**Figure 1**).

Estimating the distribution of fitness effects (DFEs) of horizontally transferred genes

The quantitative effects of mutations can be summarised as a distribution—the distribution of fitness effects, or DFE—which models the frequency of mutations with different strength of effects [38].
Many experimental studies have been carried out to determine the DFE of new mutations. To mention a few examples, DFEs have been estimated for *E. coli* mutants generated by random transposon mutagenesis [39], spontaneous and induced mutations in diploid yeast cell lines [40], for random mutations in vesicular stomatitis virus [41], for amino acid mutations in humans and *Drosophila melanogaster* [10,42],

and for point mutations in genes [43]. Parametric distributions like unimodal log-normal and gamma distributions, have been widely used to model and infer properties of DFEs [44]. Although, there are differences in DFEs at the species and genomic level, they do exhibit some general properties, like beneficial/advantageous mutations are rare, strongly beneficial mutations are exponentially distributed and the DFE for deleterious/lethal mutations are complex and multi-modal [38].

In line with the previously observed DFEs [41,45,46], the fitness effects of our 44 genes in 4 different environments are best explained by a log-normal distribution (μ_{InSPI2} = -0.211 , σ_{InSPI2} = 0.243 ; $\mu_{InSPI2\ Hypoxic}$ = -0.174, $\sigma_{InSPI2\ Hypoxic}$ = 0.200 ; $\mu_{InSPI2\ Ciprofloxacin}$ = -0.262 , $\sigma_{InSPI2\ Ciprofloxacin}$ = 0.308 ; $\mu_{InSPI2\ Low\ Mg}$ = -0.148 , $\sigma_{InSPI2\ Low\ Mg}$ = 0.252) (**Figure 2**).

Categorising genes based on fitness effects

Based on the results from the one sample t-test, the genes were categorised into 5 groups in each environment with respect to their fitness effects (**Figure 3**). On average across the four environments ~70% of the genes were deleterious, although ~43% of these imposed a relatively minor fitness cost (mildly deleterious) (1 > mean fitness of gene > 0.8, P_{adj} < 0.05 with μ_0 < 1), whereas ~27% were highly deleterious (mean fitness of gene < 0.8, P_{adj} < 0.05 with μ_0 < 1). Approximately 22% of the genes were found to be neutral (P_{adj} > 0.05 with μ_0 > 1 and μ_0 < 1) on average across the four environments (**Figure 3**). FDR corrected p-values for fitness effects of *E.coli* orthologs deviating from neutrality are provided as **Supplementary Table 5**.

Functional category of genes

The ‘complexity hypothesis’ suggests that informational genes or genes involved in transcription, translation and replication are less likely to be horizontally transferred than the operational or housekeeping genes, those involved in amino acid synthesis, nucleotide biosynthesis, energy metabolism [12,13,14]. We separated our transferred genes based on their COG annotations as informational

(19 genes) and operational (20 genes) and tested whether or not the functional categories influenced the fitness effects of the transferred genes (excluding the genes that have not been classified (3 genes) and those that were classified as both informational and operational (2 genes)).

Contrary to the complexity hypothesis, we did not observe a significant difference in the mean fitness effects of genes between the two functional categories. This finding was consistent across all growth environments (InSPI2 : $Med_{Informational} = 0.886$, $Med_{Operational} = 0.913$, Wilcoxon rank sum test, W = 178, $p = 0.482$; InSPI2 Hypoxic : $Med_{Informational} = 0.860$, $Med_{Operational} = 0.959$, Wilcoxon rank sum test, W = 145, $p = 0.158$; InSPI2 Ciprofloxacin : $Med_{Informational} = 0.890$, $Med_{Operational} = 0.884$, Wilcoxon rank sum test, W = 192, $p = 0.642$; InSPI2 Low Mg : $Med_{Informational} = 0.969$, $Med_{Operational} = 0.966$, Wilcoxon rank sum test, W = 188, $p = 0.597$) (**Figure 4**). Although no significant difference was observed between the functional categories, genes with an average low fitness were mostly informational genes.

Complexity of gene interactions

The complexity hypothesis originally proposed that the transferability of genes was dependent on their biological functions [13], and later suggested that the factor affecting HGT was instead the number of connections in gene networks [15]. The idea is that an increase in gene connectivity can reduce fitness of the bacterial host by creating a stoichiometric imbalance in the cell [47,48] or by improper interactions with their existing partners in the host [22].

To test whether high gene connectivity restricts HGT events, we investigated the relationship between fitness of the transferred orthologs and their corresponding interactions at three levels (protein-protein, metabolic, and regulatory) in *S.* Typhimurium 4/74. None of the gene interaction levels could significantly explain the fitness effects of the transferred orthologs in the four environments (Protein-protein interactions - InSPI2: $R^2 = 0.002$, $p = 0.75$, InSPI2 Ciprofloxacin: $R^2 = 0.008$, $p = 0.75$, InSPI2 Hypoxic: $R^2 = 0.004$, $p = 0.75$, InSPI2 Low Mg: $R^2 = 0.002$, $p = 0.75$; Regulatory interactions - InSPI2:

$R^2 = 0.010$, $p = 0.70$, InSPI2 Ciprofloxacin: $R^2 = 0.011$, $p = 0.70$, InSPI2 Hypoxic: $R^2 = 0.002$, $p = 0.73$, InSPI2 Low Mg: $R^2 = 0.009$, $p = 0.73$; Metabolic interactions - InSPI2: $R^2 = 0.111$, $p = 0.05$, InSPI2 Ciprofloxacin: $R^2 = 0.038$, $p = 0.20$, InSPI2 Hypoxic: $R^2 = 0.09$, $p = 0.05$, InSPI2 Low Mg: $R^2 = 0.097$, $p = 0.05$) (**Figure 5**).

There might be a potential trend for increasing number of metabolic interactions leading to reduced fitness of *S*. Typhimurium 4/74 (**Figure 5 c**). However, based on a power analysis [49], the likelihood of the regression model to detect a significant trend was found to be 62%, suggesting that an increased number of high complexity genes might be needed to identify the effect of increasing metabolic interactions on HGT.

GC content and codon usage

The base composition similarity between ancestral genes within a genome aid in distinguishing a horizontally transferred foreign gene exhibiting different characteristics, like GC content and codon usage patterns, in comparison to the ancestral genes. It has been proposed that the horizontally acquired foreign genes gradually undergo an amelioration process, whereby the foreign genes shift their base composition and codon usage patterns to match the resident genome, thus making an HGT event undetectable through parametric computational analysis [11,50]. Differences in the GC content of the foreign gene and recipient genome can impact the integration and persistence of the gene, and therefore can be a barrier to horizontal gene transfer. For example, the histone-like nucleoid structuring protein (H-NS) in *Salmonella* can silence the horizontally acquired foreign genes with a GC content lower than the recipient genome [20].

The acquired gene has a rich codon usage, if it has preferential usage for the abundant codons in the recipient genome, whereas if the gene has a bias for the rare codons, the gene is said to have a poor codon usage [51]. Codon usage bias can affect the translational efficiency and accuracy of the cell [52]. Incompatibility between the codon usage of the foreign gene and the recipient tRNA pool can lead to translational errors, for example, misfolded cytotoxic proteins. The misfolded proteins can cause

aggregation at the cell membranes, thus disrupting the cell integrity and reducing the fitness of the cell [53].

The effects of GC content and codon usage on fitness were examined individually, although these measures are often highly correlated [54]. We compared the absolute deviation in GC content between the transferred *E. coli* genes and their *S.* Typhimurium 4/74 orthologs with the observed fitness effects in four growth environments and did not find a significant relationship; InSPI2: R^2 = 0.004, p = 0.79, InSPI2 Ciprofloxacin: R^2 = 0.037, p = 0.79, InSPI2 Hypoxic: R^2 = 0.005, p = 0.79, InSPI2 Low Mg: R^2 = 0.00, p = 0.79 (**Supplementary Figure 3 a)**. Similarly, the absolute deviation in the Frequency of Optimal codon usage (FOP) between the transferred *E. coli* genes and their *S.* Typhimurium 4/74 copies could not explain the observed fitness effects in any environment; InSPI2: R^2 = 0.049, p = 0.29, InSPI2 Ciprofloxacin: R^2 = 0.014, p = 0.29, InSPI2 Hypoxic: R^2 = 0.056, p = 0.44, InSPI2 Low Mg: R^2 = 0.022, p = 0.44 (**Supplementary Figure 3 b**).

Gene length is a significant predictor of HGT

The length of the gene can impose fitness costs at the chromosome, transcript, and protein levels in the host cell. Although, extra DNA is unlikely to be costly at the chromosome/replication level, the energy expenditure inflates with transcription and even more with protein synthesis [55]. Ribosomal sequestration by lengthy genes will also limit the production of essential proteins, lowering host fitness and being selected against in a population [56]. We observed a statistically significant negative relationship between gene length and the fitness effects of the transferred genes in all growth environments; InSPI2: R^2 = 0.25, p = 6.17e-04, InSPI2 Ciprofloxacin: R^2 = 0.13, p = 0.016, InSPI2 Hypoxic: R^2 = 0.3, p = 1.29e-04, InSPI2 Low Mg: R^2 = 0.13, p = 0.019) (**Figure 6**).

Environment shapes the trajectory of HGT

The impact of the environment on the fitness effects of genes has been well documented in studies of antibiotic resistance genes and metabolic enzymes [24,57]. Thus, we posed the questions – Do environments affect the central tendency, shape, or spread of the DFEs? Do environments differentially affect the fitness effects of different genes?

The DFEs are fundamental in predicting the response of a population to selection, as the shape of the DFEs can provide an estimation of the effect of a new gene transfer event or mutation [38]. We observed that certain environments can significantly impact the median (Wilcoxon Rank Sum Test, **Supplementary Table 6 a**, **Supplementary Figure 4**) and shape or spread of DFEs (Kolmogorov-Smirnov Test, **Supplementary Table 6 b**).

Additionally, we observed a strong interaction between individual genes and the environment ($F_{126,516} = 13.32$, $p < 0.001$) suggesting that the environments interact with genes differentially (**Figure 7**). We observed 29 out of 43 genes showing significant gene by environment (G X E) interactions (see **Supplementary Table 7**). This indicates that the success of HGT events are largely dominated by specific G X E interactions.

In some cases, the specific function of the gene might be able to explain the observed fitness effect in a certain environment. For example, the *hfq* gene was beneficial for *S.* Typhimurium 4/74 in InSPI2 Ciprofloxacin and InSPI2 Low Mg environments (**Figure 7**). The *hfq* gene is an RNA chaperone facilitating transcriptional regulation in response to environment stress and changes in metabolite concentrations [58]. Additionally, *hfq* has been shown to regulate the formation of persister cells through the MqsR toxin in *E.coli* [59]. Similarly, *srmB* gene was beneficial for *S.* Typhimurium 4/74 in InSPI2 Ciprofloxacin but had a neutral effect on the host in InSPI2 Low Mg (**Figure 7**). The *srmB* gene is a DEAD-box RNA helicase facilitating ribosomal assembly [60]. Additionally, it has been shown that SrmB belongs to the AMR Gene Family - ABC-F ATP-binding cassette ribosomal protection proteins; that confer antibiotic resistance via ribosomal protection, unlike other ABC proteins which confer resistance through antibiotic efflux [61,62,63,64]. However, since ciprofloxacin is known to inhibit DNA replication, the *srmB* gene might have a potential indirect role in regulating fluoroquinolone resistance intracellularly.

Protein Dosage is a Barrier to HGT

To elucidate the role of protein dosage, the transferred *E. coli* orthologs were expressed in *S.* Typhimurium 4/74 using different inducer concentrations in each of the growth environments and the fitness for each ortholog gene was compared with the 'wild type'. We did this by using growth rates as a proxy for fitness. A graphical illustration of high throughput growth curve setup has been provided (**Figure 8**).

Protein dosage is a potential barrier to HGT, as changes in protein concentrations due to variation in gene copy numbers can cause an imbalance in the stoichiometry of the cell, thereby reducing bacterial fitness [65,66].

Due to the genetic similarity of *E. coli* and *Salmonella* [67], observed fitness effects due to increase in gene expression can be attributed to increased protein levels in the cell (i.e., a dosage effect). To this end, we applied Mandel's test on the fitness values obtained, and identified 12 genes showing a dosage dependent response (see **Supplementary Table 4**) (**Figure 9**).
The number of dosage dependent genes were found to be highest in the InSPI2 environment (11 out of 44 genes), followed by InSPI2 Ciprofloxacin (9 out of 44 genes) and InSPI2 Low Mg (6 out of 44 genes), and were lowest in the InSPI2 Hypoxic environment (2 out of 44 genes). In summary, 11 of the 44 transferred genes exhibited dosage dependence in a minimum of two growth environments (**Figure 9**). Classification of all genes based on results from Mandel's test across all environments are provided as **Supplementary Table 4**.

Discussion

Despite studies suggesting rampant uptake of foreign genetic material, certain selective forces can act as barriers to HGT [7,56,68]. This study set out to gain a quantitative understanding of the intrinsic properties of the transferred genes that prevent their successful movement from one species to another. A previous study by Acar Kirit et al. highlighted the importance of conducting experimental fitness measurements in natural environments [69], which are scarce due to experimental limitations and their

labour-intensive nature. Our study addresses this paucity, by conducting fitness assays in growth environments that capture the conditions encountered by *S.* Typhimurium 4/74 during an intracellular infection [32].

Fitness as a measure of reproductive potential is primarily evaluated using bacterial growth rates [70,71,72]. The density of the microbial culture is directly proportional to the optical density (OD), however, this is only true for a limited OD range, above which measuring culture density with high precision can be demanding [73]. Additionally, external factors affecting the morphology of bacterial cells, like filamentation, elongation or clumping of cells can interfere with OD measurements, thereby affecting growth rate measurements [74,75]. Although there were limitations to the use of OD for fitness estimations, the approach allowed us to test fitness effects across different levels of gene expression from no expression to saturation, and to understand the effect of protein dosage in a high throughput and cost effective manner.

Furthermore, we investigated the role of intrinsic properties of the transferred genes and the environment as potential barriers to HGT using NGS. To do this, we pooled the 44 mutants in equal ratios and grew them continuously for ~8 generations in the four infection relevant environments, and estimated fitness from changes in gene frequencies. The rationale behind pooling of mutants being the approach is time and cost effective, allows for a greater number of replicate experiments, and is less labour intensive. Additionally, studies have shown that copy numbers of a gene are exponential functions of growth rate [76,77,78].

Since the aim of the study was to understand what affects the persistence of a newly acquired gene in a population, we tested the immediate fitness effects after a new gene is expressed in the recipient. Although performing pooled growth experiments for a longer duration would help understand the dynamics between mutants due to constrained resources, it could also potentially lead to beneficial compensatory mutations that would balance out the effects of the newly transferred gene. The short duration of the experiment ensured that the energy source was not depleted thereby indicating that the mutants were not competing intensely against each other, and the fitness differences were mainly due to differences in the growth rates of mutants.

Although computational approaches have the benefit of using a large sample size to infer the evolutionary outcomes of HGT, discrepancies in the method of analysis and their underlying assumptions can result in conflicting conclusions. Additionally, these approaches cannot provide information on the role of the environment at the time of the transfer event. This study provides a truly systematic quantitative understanding of the persistence of horizontally transferred genes, and the evolutionary barriers to the emergence and evolution of novel phenotypes. The current study did not observe gene interactions, GC content, and functional category of the genes as significant factors hindering the success of HGT. However, gene length was found to have a significant negative fitness effect on the persistence of horizontally acquired genes. This effect of gene length might be potentially explained by ribosomal sequestration [79]. An alternate explanation could be the intrinsically disordered protein regions lacking a tertiary structure, which have been found to be correlated with gene length [69]. These observations are in congruence with the study by Acar Kirit et al. [69], suggesting that these barriers are either only weak forces or are below the threshold of practical measurements, especially when the donor and recipient species have diverged recently.

There has been limited support from computational studies for effects of protein dosage as an evolutionary barrier to HGT, mainly because it is difficult to infer dosage/change in protein expression levels from genomic data. To this end, we have measured the fitness effects of the 44 mutants across different levels of gene expression in each of the four infection relevant environments. Our data provides evidence for protein dosage to be a significant barrier to HGT with 12 genes being identified as dosage dependent. Out of the 12 dosage dependent genes, 11 genes were found to be natively expressed at low levels in *S*. Typhimurium 4/74 [32]. Low levels of native expression potentially suggest that these genes might be maintained by *Salmonella* as single copy and any increase in their protein concentrations potentially imposes a significant fitness cost in the recipient. Interestingly, genes showing a dosage effect were found to be heavily influenced by the environment (**Figure 9**). Additionally, toxicity due to increased gene expression was also found to be a barrier to HGT (see **Supplementary Table 4**). A possible mechanism for toxicity effects could be attributed to the poorly

optimised gene interactions due to the lack of co-evolution between orthologs or other unknown promiscuous interactions.

Although the importance of the environment has been implicated in the gene acquisition process, we lack an understanding of how the environment determines if the acquired gene is likely to be maintained following HGT. Therefore, analysing the DFE of horizontally transferred genes across different environments is essential to understand if HGT is an opportunistic process or is primarily determined by the intrinsic properties of the newly acquired gene. Our data shows strong G X E interactions, with the fitness effects of the gene becoming unpredictable in each environment. This suggests that improving our knowledge of gene functions could potentially support our understanding of patterns of HGT.

An area of the study that could be expanded upon would be estimating fitness effects of the mutants in macrophages. Even so, the current data represents the closest distribution of fitness effects of horizontally transferred genes achievable using an *in vitro* model. Our data shows which of the environmental variables of a heterogeneous intracellular environment might be favouring HGT; highlighting that the environmental effects should be accounted for when making predictions about the outcome of HGT events. The approach outlined in this study can be further used to study the DFE of genes transferred between phylogenetically distant species to understand which properties of the gene favour or restrict the success of transfer. For example, on transferring orthologs from *E. coli* to *Lactobacillus* certain barriers like GC content or gene interactions might impose significant fitness costs on the recipient thereby hampering the persistence of the newly acquired gene. The fitness costs could arise due to disruption of functional interactions on the introduction of a highly divergent partner or that the divergent genes may be involved in novel negative interactions and that the recipient genome has not evolved to buffer these deleterious side effects.

In conclusion, our data provides evidence that the environment plays a crucial role in influencing the magnitude of gene length and protein dosage as significant evolutionary barriers to the success of gene transfer events.

Materials and Methods

Bacterial Strains and Growth Conditions

Escherichia coli K-12 MG1655 and *Salmonella enterica* subsp. enterica serovar Typhimurium strain 4/74 [80] were used as the donor and recipient strain respectively. LB Broth (Lennox) and SPI-2 inducing PCN (phosphate/carbon/nitrogen) medium (InSPI2 pH 5.8) [33,34] were used for growing strains.

Construction of pZS4-kan-tetR plasmid

Acar Kirit *et al*. (2020) modified the original pZS4Int plasmid [35] by removing the *lacI* gene. The modified pZS4Int plasmid contained the *tetR* gene under the constitutive P_{N25} promoter, spectinomycin resistance gene, and origin of replication pSC101 [69]. Since the 4/74 strain exhibited growth in spectinomycin concentrations ranging from 80-200µg/ml, the spectinomycin gene on the pZS4Int plasmid was replaced with a kanamycin gene using the NEBuilder HiFi DNA Assembly Reaction Protocol.

Chromosomal modification of *S*. Typhimurium strain 4/74

The construct (*tetR* gene under the constitutive P_{N25} promoter and kanamycin resistance gene) from the pZS4-kan-tetR plasmid was integrated into the lambda phage attachment site of 4/74 chromosome (see **Supplementary Figure 1**) using λ red recombination plasmid pSIM5-tet [81,82]. The integration site in the modified chromosome of 4/74 strain was Sanger sequenced to ensure no mutations were introduced during the recombineering process.

Genes under study

The genes of interest to be artificially transferred were selected from a previous study [69]. *Escherichia coli* strain K-12 substr. MG1655 (GenBank: U00096.2) was used as the donor. Since the study wanted

to understand the role of the barriers affecting HGT, genes that are known to be mobile (e.g., phage-related proteins, insertion sequences, transposable elements) were excluded when selecting genes for transfer. Only genes with validated interactions based on LCMS and MALDI after SPA tagging [83] were retained, ensuring that the genes were sampled to include a representative range of interactions. Additionally, since a random selection would be biased towards an overrepresentation of genes contained within large functional modules, sampling of the genes ensured that the genes belonged to different functional modules. Furthermore, sampled genes were not accounted for their origins, resulting in 36 genes being part of the core genome, 5 genes predicted to have been horizontally transferred, and 3 genes with unknown origin [69].

Construction of *S*. Typhimurium 4/74 mutant library

The selected 44 *E.coli* orthologs were cloned into a modified version of the pZS* class of plasmids under the control of the $p_{LtetO-1}$ promoter [35] and cloned plasmids were electroporated in *E. coli* in a previous study (see **Supplementary Figure 2**). Additionally, a random fragment of the *tetA* gene (721 bp, the mean length of all inserted genes) was cloned without a promoter into the pZS* plasmid and electroporated in *E.coli* to be used as the 'wild type' strain [69]. A copy of the 45 strains (44 *E.coli* genes and a 'wild type') were obtained to use in the present study. The insert size was confirmed by PCR on an agarose gel. The plasmids were then electroporated into *S*. Typhimurium 4/74 attλ::tetR-Kn^R. Successful transformants were selected and after two rounds of streak purification on LB agar plates supplemented with 50 µg/ml Ampicillin, single colonies were grown overnight in liquid LB supplemented with 50 µg/ml Ampicillin and stored at -80°C with 15% glycerol.

Antibiotic and inducer concentrations

To determine the concentration of ciprofloxacin, growth rate measurements were carried out using Tecan Infinite® 200 Pro Reader. For this, a 1:1000 dilution of an overnight culture of *S*. Typhimurium strain 4/74 attλ::tetR-Kn^R carrying the pZS*-tetA ('wild type' plasmid) was grown in a 96 well plate in

InSPI2 medium supplemented with different ciprofloxacin concentrations and OD_{600} was measured every 20 mins. The concentration of ciprofloxacin to be used in the study was aimed to be twice the doubling time compared to growth in InSPI2.

Gene expression on the plasmids was induced by addition of the inducer anhydrotetracycline (aTc) — a tetracycline derivative with no antibiotic activity — which interacts with TetR to induce gene expression [84]. To determine the inducer concentration, *S.* Typhimurium strain 4/74 attλ::tetR-Kn^R was electroporated with a pZS* plasmid carrying a *gfp* gene under the control of the p_{LtetO1} promoter. Relative Fluorescence Units were measured in InSPI2 with different inducer concentrations using Tecan Infinite® 200 Pro Reader (excitation wavelength, 485 nm; emission wavelength, 535 nm). An arbitrary aTc concentration within the midlinear range of RFU curve was considered as the optimal inducer concentration.

To understand the role of protein dosage as a potential barrier to HGT the transferred genes on the plasmids were induced using increasing inducer concentrations thereby leading to increased protein levels in the cell [69]. The inducer concentrations used were aTc 0ng/ml (control), aTc 2ng/ml (below optimal expression), aTc 4ng/ml (optimal expression), aTc 6ng/ml (above optimal expression) and aTc 8ng/ml (saturation).

Growth measurements using Growth Profiler 960

A 1:20 dilution of individual overnight cultures of the 44 *S.* Typhimurium 4/74 attλ::tetR-Kn^R, pZS* carrying mutants were grown individually in InSPI2 medium supplemented with 50 µg/ml Ampicillin to OD_{600} 0.2. The position of the strains were randomised on 96 well plates to minimise position effects. For each environment, cultivations were carried out in 96-half-deepwell microplates (CR1496dg, Enzyscreen) with gas-permeable sandwich covers and clamps, except InSPI2 Hypoxic, where cultures were overlaid with sterile paraffin oil and gas-permeable sandwich covers were used in conjunction with carbon-flush clamp units. A total volume of 250µl including a 15µl inoculum from the starter plates with randomised layout of genes was used throughout the experiment. The wells were

additionally supplemented with different aTc concentrations to induce gene expression (see **Antibiotic and Inducer Concentrations** above). **Figure 8** shows a graphical representation of the methodology. Each experiment was repeated to produce 4 biological replicates across two days. Plates were incubated for 18 hours at 37°C and shaken at 250 rpm, except for InSPI2 Hypoxic, where cultures were grown statically. Measurements were recorded every 20 mins using optimal image scanning by Growth Profiler 960 platform (Enzyscreen). The green values (G-values) obtained for each well were converted into their equivalent OD_{600} values using the image analysis software GP960Viewer (Enzyscreen).

Pooled Growth Assays

A 1:10 dilution of individual overnight cultures of the 44 *S.* Typhimurium 4/74 attλ::tetR-Kn^R, pZS* carrying mutants were grown in InSPI2 medium supplemented with 50 µg/ml Ampicillin to OD_{600} 0.2. The following steps were carried out on ice. After reaching the target OD_{600}, strains were mixed at equal volumes to make a total of 135mL. The pooled stock was thoroughly mixed and, 50 aliquots - each containing 1 ml pooled stock and 250µl 80% glycerol, were stored at -80°C. The remaining volume of the pooled stock was split to create two technical replicates for the starting frequency of the pooled growth experiments. To verify the cell concentration of the pooled stock, serial dilutions of a frozen stock were plated on LB agar plates supplemented with 50 µg/ml Ampicillin to obtain CFU counts. The pooled stock concentration was approximately 5 x 10^7 CFU/ml.

To perform the pooled growth experiments, frozen aliquots of the pooled stock were thawed on ice, and then gently mixed by pipetting. A total volume of 100ml including an inoculum of 10^7 cells from the pooled stock was used throughout the experiment. Gene expression was induced by the addition of 4ng/ml aTc, and 50 µg/ml Ampicillin was used for maintenance of pZS* plasmids carrying the transferred genes. For each of the four infection relevant environments (**Supplementary Table 1**), 100ml volume of cultures were grown in 200ml conical flasks, except InSPI2 Hypoxic, where cultures were filled up to the brim of 50ml conical tubes, capped with rubber septum, and further sealed using parafilm. The flasks were incubated at 37°C and 250 rpm and grown to OD_{600} 0.4. For the InSPI2

Hypoxic environment, cultures were grown on static at 37°C to OD_{600} 0.2. Four replicate experiments were performed for each of the environments under study. On reaching the target OD (i.e., endpoint of the experiment), plasmid DNA was extracted from the total culture volume using Invitrogen™ PureLink™ HiPure Plasmid Midiprep Kit (Catalog number: K210004).

Sequencing

DNA libraries for 16 biological replicates (4 replicate pooled experiments for each of the 4 environments) and 2 technical replicates (starting frequency of pooled growth experiments) were prepared and sequenced on Illumina NovaSeq 6000 platform (PE150bp) by our collaborators at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, USA, resulting in ~3.2 million read pairs per sample.

To briefly describe the library preparation protocol, DNA was first fragmented by sonication using Qsonica 800R, followed by construction for Illumina compatible sequencing libraries using the Kapa Hyperprep kit following manufacturer's protocols. Sequencing libraries were generated using PCR with Illumina-specific barcoded primers and dual index approach was used to allow multiplexing. The amplified library was then validated using 4150 TapeStation System (Agilent Technologies, Inc.), and size selected using Pippin Prep, BluePippin (Sage Science, Inc.). The libraries were pooled in equimolar concentrations and were sequenced.

Processing of Sequencing Data

Quality checks and demultiplexing of sequenced reads were performed by our collaborators. Sequenced reads were then processed using AdapterRemoval (v2) [85] to trim regions with low quality bases (q<30), merge overlapping read pairs, remove Illumina adapter sequences, and remove reads containing ambiguous bases (Ns).

A customised reference genome consisting of FASTA format sequences of the 44 transferred *E. coli* orthologs, 'wild type' tetA fragment, pZS* plasmid backbone, complete sequences of *S*. Typhimurium 4/74 chromosome (GenBank: CP002487.1); and its 3 native plasmids TY474p1 (GenBank: CP002488.1), TY474p2 (GenBank: CP002489.1), and TY474p3 (GenBank: CP002490.1) was built using Bowtie2 (Version 2.4.4) (Langmead et al., 2009). The sequenced reads were then mapped to the custom built reference genome using the defaults local alignment parameter in Bowtie2 (Version 2.4.4) (Langmead et al., 2009). The mapped SAM files were converted to BAM format and the corresponding BAM files were sorted using Samtools (Tools for Alignment in SAM format) (Version 1.13) (Li et al., 2009). Read depths at each nucleotide position were calculated using depth function of Samtools (Version 1.13) (Li et al., 2009). Nucleotide depths were converted to gene frequencies by calculating the median of depths obtained for every gene using the R software (Version 4.0.0). All steps of the sequencing data processing were performed on Ubuntu 16.04.5 LTS (GNU/Linux 4.15.0-52-generic x86_64). **Figure 1** provides a graphical illustration of the methodology.

Intrinsic properties of orthologs and statistical analysis

Effect of protein dosage was studied using growth rates of individual mutants relative to the 'wild type'. Growth rates for all mutants across all inducer concentrations were estimated from the log linear part of the growth curve through regression analysis. Relative fitness (w) for the transferred genes was calculated by normalising the growth rates of the mutants with the 'wild type' for the respective inducer concentrations.

To categorise genes based on dosage effect, Mandel's test for Linearity was applied to the relative fitness values for every gene to decide if the data points best fit a linear function than a quadratic function. The data used for Mandel's test is provided as **Supplementary Table 2**. The test assesses if the residual variances of the linear and quadratic model are significantly different using the Fisher–Snedecor F-test [86,87].

$$F_{test} = ((N-2){S_{y/x,lin}}^2 - (N-3){S_{y/x,non}}^2)/{S_{y/x,non}}^2 \quad (1)$$

In the above formula, N denotes sample size, S denotes the standard error of regression for the straight line (lin) and for quadratic (non) models.

Other selective factors affecting the DFEs of the transferred genes were studied using the NGS approach. Each replicate pooled growth assay was normalised by the read depth of the vector backbone to account for variation in sequencing library preparation. Fitness costs of selected genes were estimated for each replicate by using the following regression model.

$$\ln(1+s) = (\ln R_t - \ln R_0) \,/\, t$$

where R_0 and R_t are the ratios of the frequencies of mutant (depth of gene) to 'wild type' (depth of a non-functional *tetA* fragment) at start and end point of the experiment, respectively and *t* is the number of generations [39]. 43 genes were included in the analysis, as no reads were detected for one of the transferred genes (b1084).

To test whether the genes were neutral or not, a one-tailed one sample t-test was applied on 4 replicate measurements for each gene, with $\mu_0 < 1$ or $\mu_0 > 1$. The p-values obtained were corrected for multiple tests using the FDR method [88] and $\alpha = 0.05$ was used as the level of significance.

The transferred orthologs were divided into the two groups using the Database for Clusters of Orthologous Groups (COGs). A one-sided Wilcoxon Rank Sum test was applied on the mean fitness values of genes for 4 replicate measurements to analyse if the fitness effects of informational genes was smaller than those of the operational genes as predicted *a priori* [12,13,14].

Additionally, we also looked at other intrinsic gene properties: the number of protein-protein interactions (PPI), metabolic interactions (MI) and regulatory interactions (RI), gene length, GC content and codon usage. The number of protein-protein, metabolic and regulatory interactions for the transferred orthologs in *S.* Typhimurium strain 4/74 were identified using the SalmoNet database

(http://salmonet.org/). Gene length was calculated as the number of nucleotides in the coding sequence (CDS) (i.e., from the start to the stop codon) of the transferred *E.coli* genes. GC content was calculated as the absolute deviation between the transferred *E.coli* genes and their corresponding *S.* Typhimurium strain 4/74 orthologs. Codon usage was estimated as the absolute deviation of the frequency of optimal codon usage (FOP) in the transferred *E.coli* genes using the FOP in *S.* Typhimurium (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=602). Multiple Linear regression was used to study the relationship between the intrinsic gene properties and fitness effects, by using the following model.

$$Mean\ relative\ fitness\ (w) \sim PPI + MI + RI + Gene\ length\ (bp) + Deviation\ in\ GC\ content + Deviation\ in\ FOP$$

Data used in the model is provided as **Supplementary Table 3**. The p-values obtained were corrected for multiple tests using the FDR method [88].

To investigate the effect of the environments on the central tendency of the DFEs, we used a paired and two-sided Wilcoxon Signed Rank Test on each pairwise comparison of the four environments. To test if the shape and spread of the distributions were significantly different between environments, a two-sided Kolmogorov-Smirnov (K-S) test was used. For each of these tests, the fitness values for genes were provided as a mean of 4 replicate measurements. The p-values obtained were corrected for multiple tests using the FDR method [88].

The gene by environment (G x E) interactions were studied using a two-way analysis of variance (ANOVA) test for each pairwise comparison for the four environments, using the following model.

$$Relative\ fitness\ (w) \sim Environment * Gene$$

The p-values obtained were corrected for multiple tests using the FDR method [88].

All statistical analysis were carried out using the R software (Version 4.0.2) in RStudio (Version 1.1.456).

Acknowledgements

We thank Wai Yee Fong for sharing *S.* Typhimurium strain 4/74 and providing guidance on InSPI2 media preparation; Blanca Perez-Sepulveda, Rocío Canals and, Nicolas Wenner for their guidance on chromosomal modification of *Salmonella*; Simon (Xiaojun) Zhu for his guidance with using Growth Profiler 960; Hawra Al-Ghafli for helping with preparation of pooled stock of mutants; and to, Cecil M. Lewis, Jr., Krithivasan Sankaranarayanan and Hande A. Kirit at LMAMR, University of Oklahoma, USA for performing library preparation and sequencing on our samples. We thank Andrea J. Betancourt for helpful feedback on the manuscript. This work was supported by the European Research Council under the Horizon 2020 Framework Programme (FP/2007-2013) / ERC grant agreement no. [648440] to J.P.B. Library construction and sequencing was supported by OCAST (HF20-032-1) to CML and KS, and National Science Foundation grant (Subaward No. SDSMT-OK18-05) to KS.

Author Contributions

RPB, HAK and JPB designed the experiments together. RPB performed wet-lab experiments. KS and CML provided materials and support for genome sequencing. KS and HAK constructed genome libraries, sequencing and performed preliminary data quality analysis. RPB processed the sequenced reads with help from HAK. RPB performed statistical analysis. RPB wrote the initial draft of the manuscript and revised it together with HAK and JPB. All the authors have read and approved the manuscript.

References

1 Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405** (2000).

2 Orr, H. A. Fitness and its role in evolutionary genetics. *Nat Rev Genet* **10**, 531-539, doi:10.1038/nrg2603 (2009).

586 3 Rodriguez-Beltran, J., DelaFuente, J., Leon-Sampedro, R., MacLean, R. C. & San Millan, A.
587 Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat Rev Microbiol*
588 **19**, 347-359, doi:10.1038/s41579-020-00497-1 (2021).
589 4 Agren, J. A. Selfish genetic elements and the gene's-eye view of evolution. *Curr Zool* **62**, 659-
590 665, doi:10.1093/cz/zow102 (2016).
591 5 Draghi, J. A. & Turner, P. E. DNA secretion and gene-level selection in bacteria. *Microbiology*
592 *(Reading)* **152**, 2683-2688, doi:10.1099/mic.0.29013-0 (2006).
593 6 Eberhard, W. G. Evolution in bacterial plasmids and levels of selection. *Q Rev Biol* **65**, 3-22,
594 doi:10.1086/416582 (1990).
595 7 Popa, O. & Dagan, T. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin*
596 *Microbiol* **14**, 615-623, doi:10.1016/j.mib.2011.07.027 (2011).
597 8 Mira, A., Ochman, H. & Moran, N. A. Deletional bias and the evolution of bacterial genomes.
598 *Trends in Genetics* **17** (2001).
599 9 Keightley, P. D. & Eyre-Walker, A. What can we learn about the distribution of fitness effects
600 of new mutations from DNA sequence data? *Philos Trans R Soc Lond B Biol Sci* **365**, 1187-1193,
601 doi:10.1098/rstb.2009.0266 (2010).
602 10 Keightley, P. D. & Eyre-Walker, A. Joint inference of the distribution of fitness effects of
603 deleterious mutations and population demography based on nucleotide polymorphism
604 frequencies. *Genetics* **177**, 2251-2261, doi:10.1534/genetics.107.080663 (2007).
605 11 Ravenhall, M., Skunca, N., Lassalle, F. & Dessimoz, C. Inferring horizontal gene transfer. *PLoS*
606 *Comput Biol* **11**, e1004095, doi:10.1371/journal.pcbi.1004095 (2015).
607 12 Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. Biased biological functions of horizontally
608 transferred genes in prokaryotic genomes. *Nat Genet* **36**, 760-766, doi:10.1038/ng1381
609 (2004).

610 13 Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: The complexity
611 hypothesis. *Proceedings of the National Academy of Sciences* **96**, 3801,
612 doi:10.1073/pnas.96.7.3801 (1999).
613 14 Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct
614 gene classes. *Genetics* **95** (1998).
615 15 Cohen, O., Gophna, U. & Pupko, T. The complexity hypothesis revisited: connectivity rather
616 than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol* **28**, 1481-1489,
617 doi:10.1093/molbev/msq333 (2011).
618 16 Wellner, A., Lurie, M. N. & Gophna, U. Complexity, connectivity, and duplicability as barriers
619 to lateral gene transfer. *Genome Biol* **8**, R156, doi:10.1186/gb-2007-8-8-r156 (2007).
620 17 Navarre, W. W. The Impact of Gene Silencing on Horizontal Gene Transfer and Bacterial
621 Evolution. *Adv Microb Physiol* **69**, 157-186, doi:10.1016/bs.ampbs.2016.07.004 (2016).
622 18 Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat Rev*
623 *Microbiol* **8**, 317-327, doi:10.1038/nrmicro2315 (2010).
624 19 Navarre, W. W. *et al.* Selective silencing of foreign DNA with low GC content by the H-NS
625 protein in Salmonella. *Science* **313**, 236-238, doi:10.1126/science.1128794 (2006).
626 20 Lucchini, S. *et al.* H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS*
627 *Pathog* **2**, e81, doi:10.1371/journal.ppat.0020081 (2006).
628 21 Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Curr Opin Microbiol*
629 **8**, 466-472, doi:10.1016/j.mib.2005.06.003 (2005).
630 22 Papp, B., Pal, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast.
631 *Nature* **424**, 194-197, doi:10.1038/nature01771 (2003).
632 23 Roux, D. *et al.* Fitness cost of antibiotic susceptibility during bacterial infection. *Sci Transl Med*
633 **7**, 297ra114, doi:10.1126/scitranslmed.aab1621 (2015).
634 24 Melnyk, A. H., Wong, A. & Kassen, R. The fitness costs of antibiotic resistance mutations. *Evol*
635 *Appl* **8**, 273-283, doi:10.1111/eva.12196 (2015).

636 25 Kishony, R. & Leibler, S. Environmental stresses can alleviate the average deleterious effect of
637 mutations. *J Biol* **2**, 14, doi:10.1186/1475-4924-2-14 (2003).
638 26 Remold, S. K. & Lenski, R. E. Contribution of individual random mutations to genotype-by-
639 environment interactions in Escherichia coli. *Proc Natl Acad Sci U S A* **98**, 11388-11393,
640 doi:10.1073/pnas.201140198 (2001).
641 27 Cochran, P. K., Kellogg, C. A. & Paul, J. H. Prophage induction of indigenous marine lysogenic
642 bacteria by environmental pollutants. *Marine Ecology Progress Series* **164**, 125-133 (1998).
643 28 Hurley, D. *et al.* Atypical Salmonella enterica Serovars in Murine and Human Macrophage
644 Infection Models. *Infect Immun* **88**, doi:10.1128/IAI.00353-19 (2020).
645 29 Okoro, C. K. *et al.* Intracontinental spread of human invasive Salmonella Typhimurium
646 pathovariants in sub-Saharan Africa. *Nat Genet* **44**, 1215-1221, doi:10.1038/ng.2423 (2012).
647 30 Gordon, M. A. Invasive nontyphoidal Salmonella disease: epidemiology, pathogenesis and
648 diagnosis. *Curr Opin Infect Dis* **24**, 484-489, doi:10.1097/QCO.0b013e32834a9980 (2011).
649 31 Canals, R. *et al.* Adding function to the genome of African Salmonella Typhimurium ST313
650 strain D23580. *PLoS Biol* **17**, e3000059, doi:10.1371/journal.pbio.3000059 (2019).
651 32 Kroger, C. *et al.* An infection-relevant transcriptomic compendium for Salmonella enterica
652 Serovar Typhimurium. *Cell Host Microbe* **14**, 683-695, doi:10.1016/j.chom.2013.11.010
653 (2013).
654 33 Xu, X. & Hensel, M. Systematic analysis of the SsrAB virulon of Salmonella enterica. *Infect*
655 *Immun* **78**, 49-58, doi:10.1128/IAI.00931-09 (2010).
656 34 Löber, S., Jackel, D., Kaiser, N. & Hensel, M. Regulation of Salmonella pathogenicity island 2
657 genes by independent environmental signals. *Int J Med Microbiol* **296**, 435-447,
658 doi:10.1016/j.ijmm.2006.05.001 (2006).
659 35 Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia
660 coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research* **25**
661 (1997).

662 36 Mugnai, R., Sattamini, A., Albuquerque dos Santos, J. A. & Regua-Mangia, A. H. A Survey of
663 Escherichia coli and Salmonella in the Hyporheic Zone of a Subtropical Stream: Their
664 Bacteriological, Physicochemical and Environmental Relationships. *PLoS One* **10**, e0129382,
665 doi:10.1371/journal.pone.0129382 (2015).
666 37 Winfield, M. D. & Groisman, E. A. Role of nonhost environments in the lifestyles of Salmonella
667 and Escherichia coli. *Appl Environ Microbiol* **69**, 3687-3694, doi:10.1128/AEM.69.7.3687-
668 3694.2003 (2003).
669 38 Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat Rev*
670 *Genet* **8**, 610-618, doi:10.1038/nrg2146 (2007).
671 39 Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. Distribution of fitness effects
672 caused by random insertion mutations in Escherichia coli. *Genetica* (1998).
673 40 Wloch, D. M., Szafraniec, K., Borts, R. H. & Korona, R. Direct estimate of the mutation rate and
674 the distribution of fitness effects in the yeast Saccharomyces cerevisiae. *Genetics* **159**, 441-
675 452, doi:10.1093/genetics/159.2.441 (2001).
676 41 Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-
677 nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* **101**, 8396-8401,
678 doi:10.1073/pnas.0400146101 (2004).
679 42 Eyre-Walker, A., Woolfit, M. & Phelps, T. The distribution of fitness effects of new deleterious
680 amino acid mutations in humans. *Genetics* **173**, 891-900, doi:10.1534/genetics.106.057570
681 (2006).
682 43 Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A. & Kassen, R. The distribution of fitness
683 effects among synonymous mutations in a gene under directional selection. *Elife* **8**,
684 doi:10.7554/eLife.45952 (2019).
685 44 Kousathanas, A. & Keightley, P. D. A comparison of models to infer the distribution of fitness
686 effects of new mutations. *Genetics* **193**, 1197-1208, doi:10.1534/genetics.112.148023 (2013).

687 45 Kinney, J. B., Murugan, A., Callan, C. G., Jr. & Cox, E. C. Using deep sequencing to characterize
688 the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci U S A*
689 **107**, 9158-9163, doi:10.1073/pnas.1004290107 (2010).
690 46 Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. Distribution of fitness effects
691 caused by random insertion mutations in Escherichia coli. *Genetica* **102-103**, 349-358 (1998).
692 47 Omer, S., Kovacs, A., Mazor, Y. & Gophna, U. Integration of a foreign gene into a native
693 complex does not impair fitness in an experimental model of lateral gene transfer. *Mol Biol*
694 *Evol* **27**, 2441-2445, doi:10.1093/molbev/msq145 (2010).
695 48 Wellner, A. & Gophna, U. Neutrality of foreign complex subunits in an experimental model of
696 lateral gene transfer. *Mol Biol Evol* **25**, 1835-1840, doi:10.1093/molbev/msn131 (2008).
697 49 Cohen, J. Statistical Power Analysis. *Current Directions in Psychological Science* **1**, 98-101
698 (1992).
699 50 Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and
700 exchange. *J Mol Evol* **44**, 383-397, doi:10.1007/pl00006158 (1997).
701 51 Medrano-Soto, A., Moreno-Hagelsieb, G., Vinuesa, P., Christen, J. A. & Collado-Vides, J.
702 Successful lateral transfer requires codon usage compatibility between foreign genes and
703 recipient genomes. *Mol Biol Evol* **21**, 1884-1894, doi:10.1093/molbev/msh202 (2004).
704 52 Qian, W., Yang, J. R., Pearson, N. M., Maclean, C. & Zhang, J. Balanced codon usage optimizes
705 eukaryotic translational efficiency. *PLoS Genet* **8**, e1002603,
706 doi:10.1371/journal.pgen.1002603 (2012).
707 53 Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant
708 constraint on coding-sequence evolution. *Cell* **134**, 341-352, doi:10.1016/j.cell.2008.05.042
709 (2008).
710 54 Sueoka, N. Correlation between Base Composition of Deoxyribonucleic Acid and Amino Acid
711 Composition of Protein. *Proc Natl Acad Sci U S A* **47**, 1141-1149, doi:10.1073/pnas.47.8.1141
712 (1961).

713 55 Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A* **112**,
714 15690-15695, doi:10.1073/pnas.1514974112 (2015).
715 56 Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol* **28**, 489-495,
716 doi:10.1016/j.tree.2013.04.002 (2013).
717 57 Zhong, Q., Kobe, B. & Kappler, U. Molybdenum Enzymes and How They Support Virulence in
718 Pathogenic Bacteria. *Frontiers in Microbiology* **11**, doi:10.3389/fmicb.2020.615860 (2020).
719 58 Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E. & Gross, C. A. Hfq modulates the sigmaE-
720 mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in
721 Escherichia coli. *J Bacteriol* **189**, 1963-1973, doi:10.1128/JB.01243-06 (2007).
722 59 Kim, Y. & Wood, T. K. Toxins Hha and CspD and small RNA regulator Hfq are involved in
723 persister cell formation through MqsR in Escherichia coli. *Biochem Biophys Res Commun* **391**,
724 209-213, doi:10.1016/j.bbrc.2009.11.033 (2010).
725 60 Rabuck-Gibbons, J. N. *et al.* SrmB Rescues Trapped Ribosome Assembly Intermediates. *J Mol*
726 *Biol* **432**, 978-990, doi:10.1016/j.jmb.2019.12.013 (2020).
727 61 Yang, Z. *et al.* Alteration in microbial community and antibiotic resistance genes mediated by
728 microplastics during wastewater ultraviolet disinfection. *Sci Total Environ* **825**, 153918,
729 doi:10.1016/j.scitotenv.2022.153918 (2022).
730 62 Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive
731 antibiotic resistance database. *Nucleic Acids Res* **48**, D517-D525, doi:10.1093/nar/gkz935
732 (2020).
733 63 Murina, V. *et al.* ABCF ATPases Involved in Protein Synthesis, Ribosome Assembly and
734 Antibiotic Resistance: Structural and Functional Diversification across the Tree of Life. *J Mol*
735 *Biol* **431**, 3568-3590, doi:10.1016/j.jmb.2018.12.013 (2019).
736 64 Sharkey, L. K., Edwards, T. A. & O'Neill, A. J. ABC-F Proteins Mediate Antibiotic Resistance
737 through Ribosomal Protection. *mBio* **7**, e01975, doi:10.1128/mBio.01975-15 (2016).

738 65 Park, C. & Zhang, J. High expression hampers horizontal gene transfer. *Genome Biol Evol* **4**,
739 523-532, doi:10.1093/gbe/evs030 (2012).
740 66 Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene
741 transfer. *Science* **318**, 1449-1452, doi:10.1126/science.1147112 (2007).
742 67 McClelland, M. *et al.* Complete genome sequence of Salmonella enterica serovar
743 Typhimurium LT2. *Nature* **413**, 852-856, doi:10.1038/35101614 (2001).
744 68 Gonzalez-Candelas, F. & Francino, M. 47-73 (2012).
745 69 Acar Kirit, H., Lagator, M. & Bollback, J. P. Experimental determination of evolutionary barriers
746 to horizontal gene transfer. *BMC Microbiol* **20**, 326, doi:10.1186/s12866-020-01983-5 (2020).
747 70 Lenski, R. E. *et al.* Evolution of competitive fitness in experimental populations of E. coli: what
748 makes one genotype a better competitor than another? *Antonie Van Leeuwenhoek* **73**, 35-47,
749 doi:10.1023/a:1000675521611 (1998).
750 71 Lenski, R. E., Simpson, S. C. & Nguyen, T. T. Genetic analysis of a plasmid-encoded, host
751 genotype-specific enhancement of bacterial fitness. *J Bacteriol* **176**, 3140-3147,
752 doi:10.1128/jb.176.11.3140-3147.1994 (1994).
753 72 Bennett, A. F., Dao, K. M. & Lenski, R. E. Rapid evolution in response to high-temperature
754 selection. *Nature* **346**, 79-81, doi:10.1038/346079a0 (1990).
755 73 Mira, P., Yeh, P. & Hall, B. G. Estimating Microbial Population Data from Optical Density.
756 *bioRxiv*, 2021.2012.2022.473948, doi:10.1101/2021.12.22.473948 (2022).
757 74 Sridhar, S. *et al.* Inhibitory Concentrations of Ciprofloxacin Induce an Adaptive Response
758 Promoting the Intracellular Survival of Salmonella enterica Serovar Typhimurium. *mBio* **12**,
759 e0109321, doi:10.1128/mBio.01093-21 (2021).
760 75 Vadia, S. & Levin, P. A. Growth rate and cell size: a re-examination of the growth law. *Curr*
761 *Opin Microbiol* **24**, 96-103, doi:10.1016/j.mib.2015.01.011 (2015).

762 76 Roller, B. R., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA operon copy number to
763 investigate bacterial reproductive strategies. *Nat Microbiol* **1**, 16160,
764 doi:10.1038/nmicrobiol.2016.160 (2016).
765 77 Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in
766 bacteria. *Cell* **139**, 1366-1375, doi:10.1016/j.cell.2009.12.001 (2009).
767 78 Bremer, H. & Dennis, P. P. Modulation of Chemical Composition and Other Parameters of the
768 Cell at Different Exponential Growth Rates. *EcoSal Plus* **3**, doi:10.1128/ecosal.5.2.3 (2008).
769 79 Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein
770 translation. *Cell* **153**, 1589-1601, doi:10.1016/j.cell.2013.05.049 (2013).
771 80 Hoiseth, S. K. & Stocker, B. A. Aromatic-dependent Salmonella typhimurium are non-virulent
772 and effective as live vaccines. *Nature* **291**, 238-239, doi:10.1038/291238a0 (1981).
773 81 Koskiniemi, S., Pranting, M., Gullberg, E., Nasvall, J. & Andersson, D. I. Activation of cryptic
774 aminoglycoside resistance in Salmonella enterica. *Mol Microbiol* **80**, 1464-1478,
775 doi:10.1111/j.1365-2958.2011.07657.x (2011).
776 82 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia
777 coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640-6645,
778 doi:10.1073/pnas.120163297 (2000).
779 83 Hu, P. *et al.* Global functional atlas of Escherichia coli encompassing previously
780 uncharacterized proteins. *PLoS Biol* **7**, e96, doi:10.1371/journal.pbio.1000096 (2009).
781 84 Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet
782 repressor. *Nucleic Acids Res* **33**, e21, doi:10.1093/nar/gni013 (2005).
783 85 Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming,
784 identification, and read merging. *BMC Res Notes* **9**, 88, doi:10.1186/s13104-016-1900-2
785 (2016).

86 Andrade, J. M. & Gómez-Carracedo, M. P. Notes on the use of Mandel's test to check for nonlinearity in laboratory calibrations. *Analytical Methods* **5**, 1145-1149, doi:10.1039/C2AY26400E (2013).

87 Brüggemann, L., Quapp, W. & Wennrich, R. Test for non-linearity concerning linear calibrated chemical measurements. *Accreditation and Quality Assurance* **11**, 625-631, doi:10.1007/s00769-006-0205-x (2006).

88 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300, doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x (1995).

Figures

Figure 1

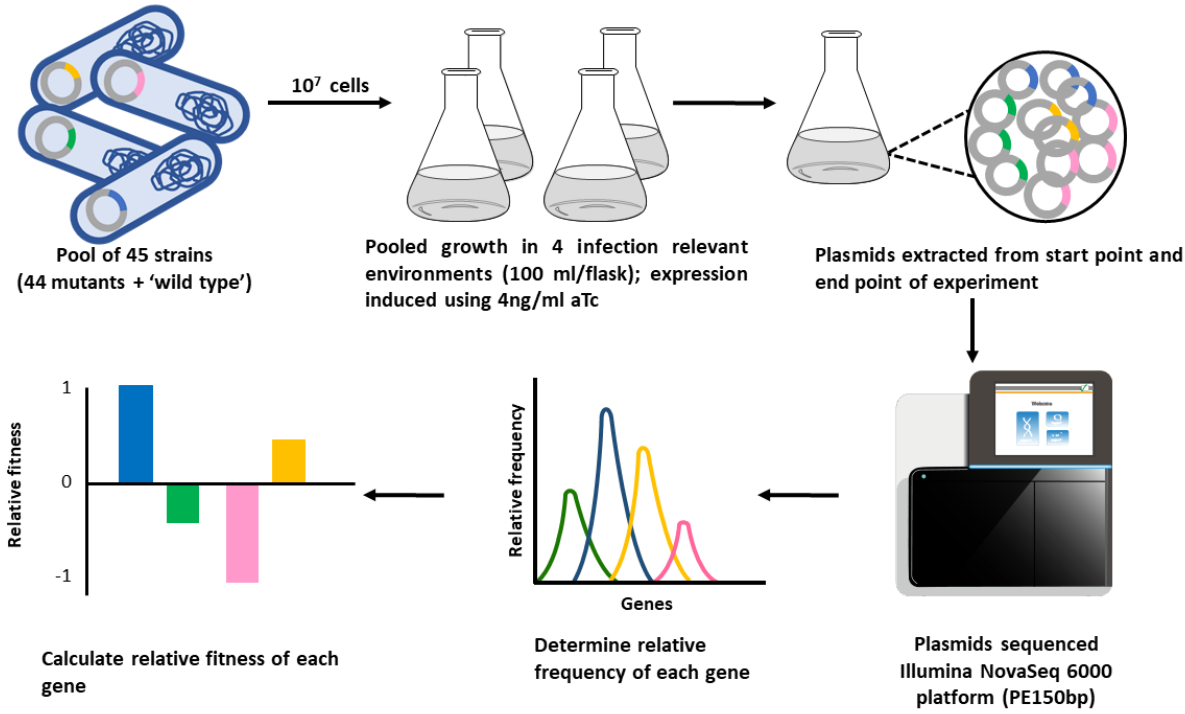

Figure 1 Graphical representation for pooled growth experiments

The flask-2 icon in the figure by DCBLS is licensed under CC-BY 4.0 Unported / Object properties modified from original. The sequence_histogram icon by Marcel Tisch is licensed under CC0 / Object properties modified from original. The genomesequencer9 icon by DCBLS is licensed under CC-BY 4.0 Unported .

Figure 2

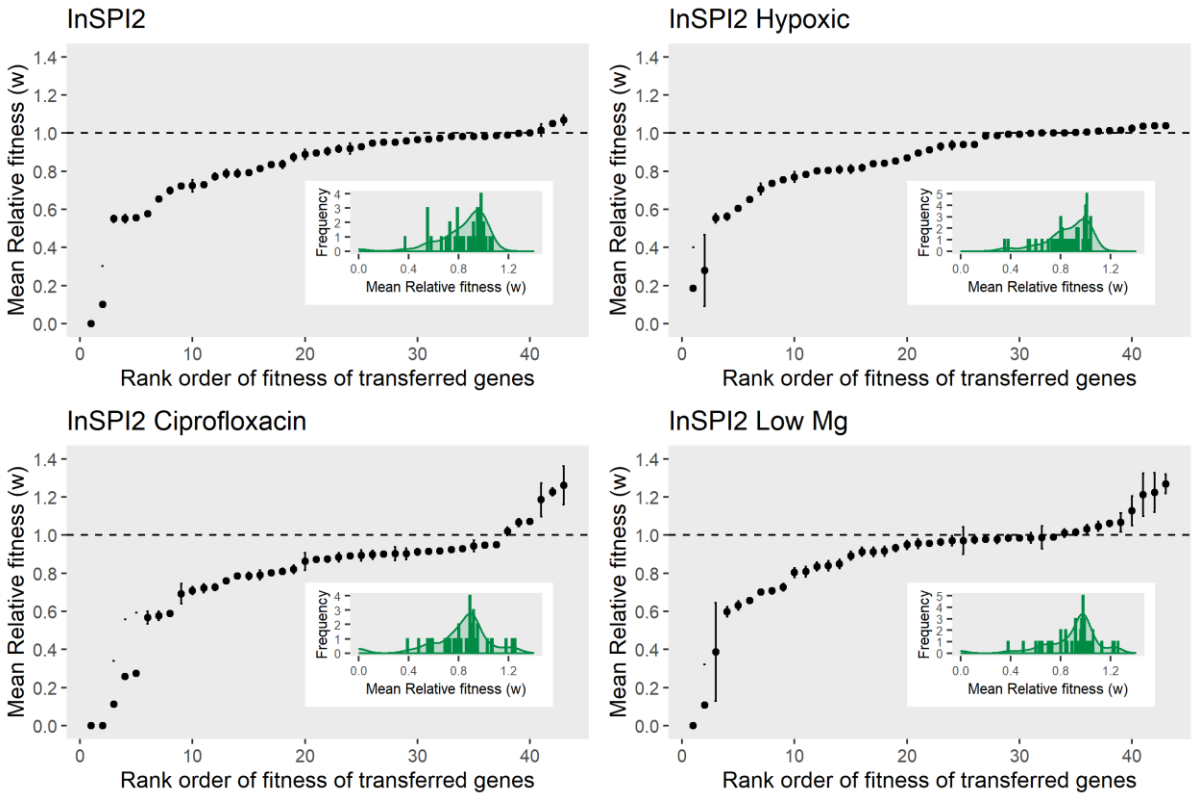

Figure 2 DFEs of transferred *E.coli* orthologs expressed in *S.* Typhimurium 4/74

The figure shows the distribution of fitness effects obtained from pooled growth assays in each infection relevant environment. On the x-axis, transferred genes are ordered by their relative fitness values in the respective environments. Error bars represent standard deviation of the mean for 4 replicate measurements for each gene. Embedded plots are histogram and density representations for the same data. Black dashed line shows a fitness of 1.

Figure 3

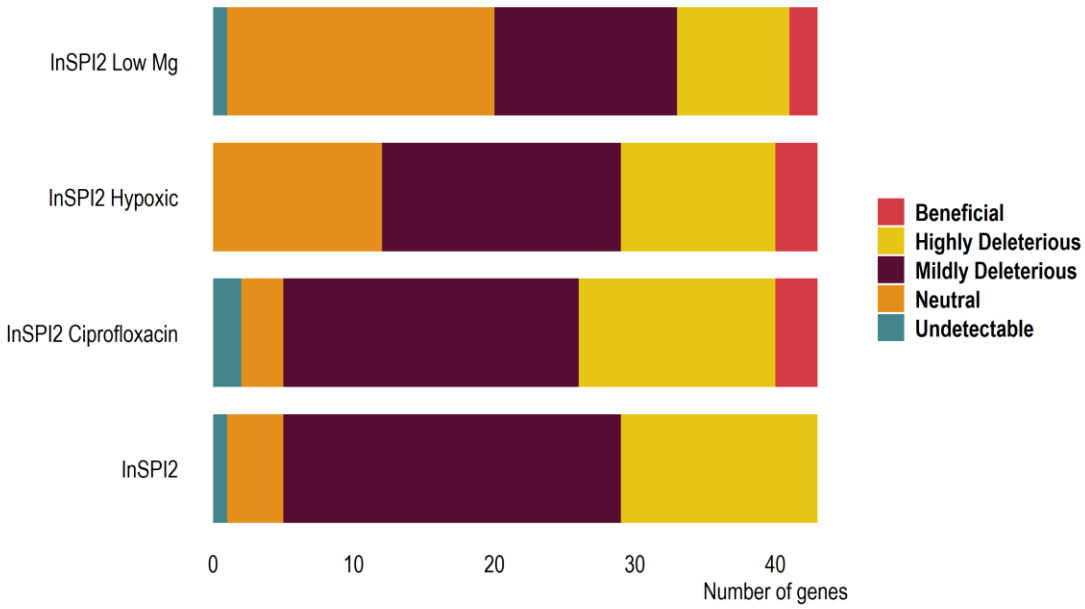

Figure 3 Grouping of genes based on fitness effects

Breakdown of the transferred orthologs based on results from one sample t-test in each growth environment. Undetectable genes: mean fitness of gene = 0, $P_{adj} < 0.05$ with $\mu_0 < 1$; Neutral genes: $P_{adj} > 0.05$ with $\mu_0 > 1$ and $\mu_0 < 1$; Mildly Deleterious genes: 1 > mean fitness of gene > 0.8, $P_{adj} < 0.05$ with $\mu_0 < 1$; Highly Deleterious genes: mean fitness of gene < 0.8, $P_{adj} < 0.05$ with $\mu_0 < 1$.; Beneficial genes: $P_{adj} < 0.05$, with $\mu_0 > 1$.

Figure 4

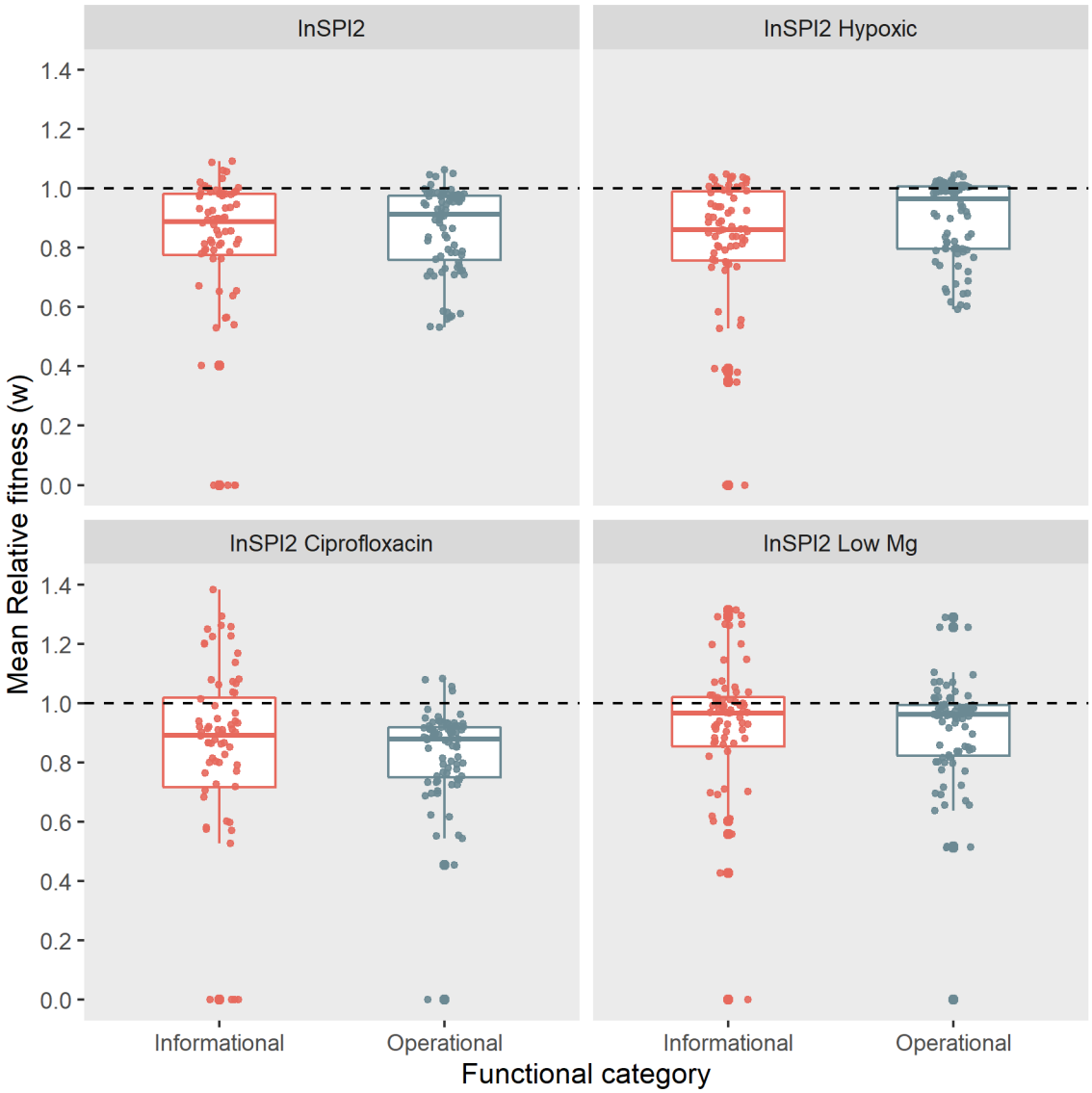

Figure 4 Fitness effects by functional categories

Boxplot with jitter points showing the fitness effects of transferred genes (4 replicate measurements) in four growth environments, with genes divided into two functional categories. Black dashed line shows a fitness of 1.

Figure 5

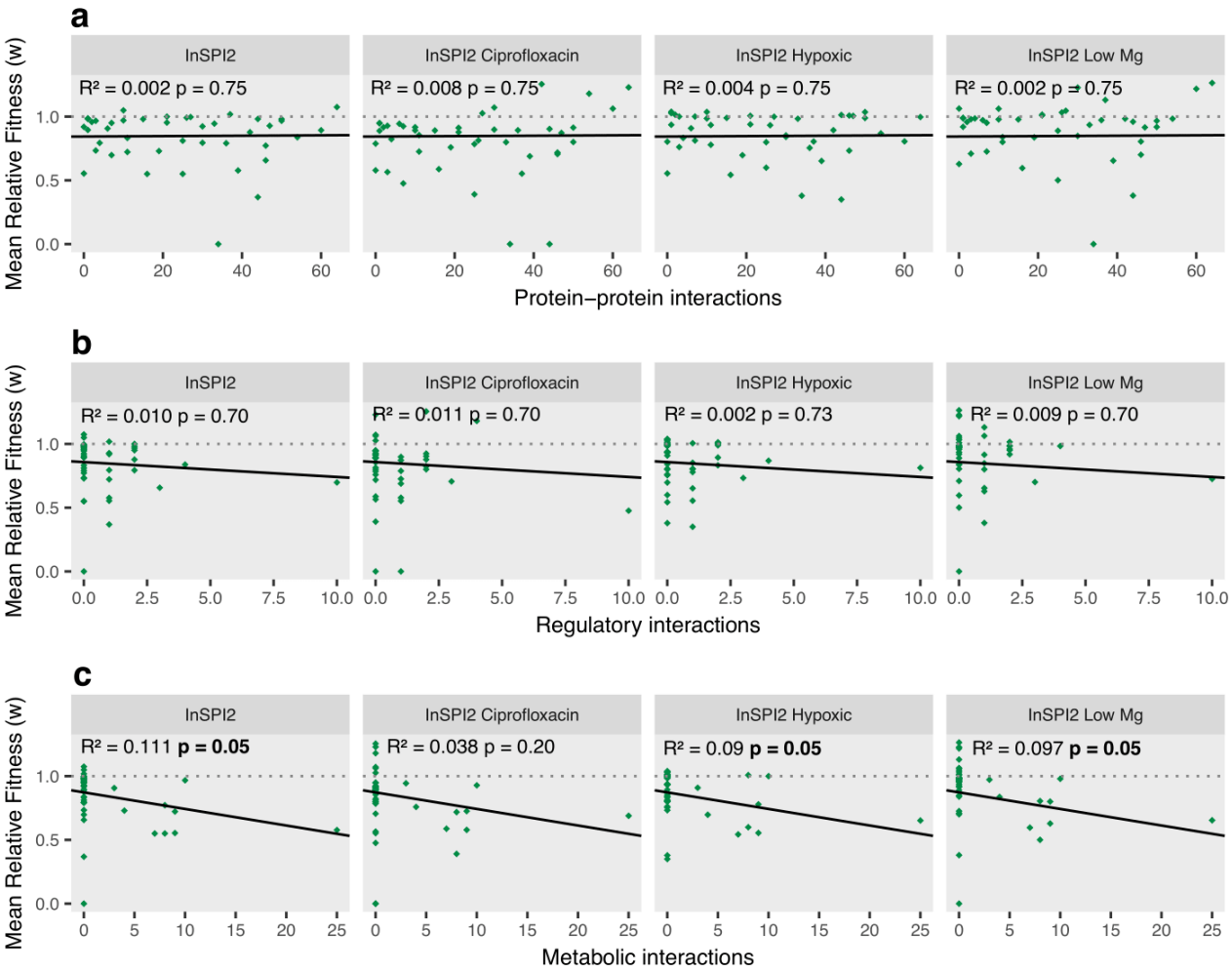

Figure 5 Relationship between relative fitness and number of interactions

Relative fitness of transferred *E.coli* orthologs in four growth environments plotted against the number of protein-protein interactions (**a**), regulatory interactions (**b**), and metabolic interactions (**c**) in *S.* Typhimurium 4/74. The black line is the regression between the two variables. Grey dotted line shows a fitness of 1. FDR corrected p-values and R^2 values from the linear regression analysis are shown on the plot.

Figure 6

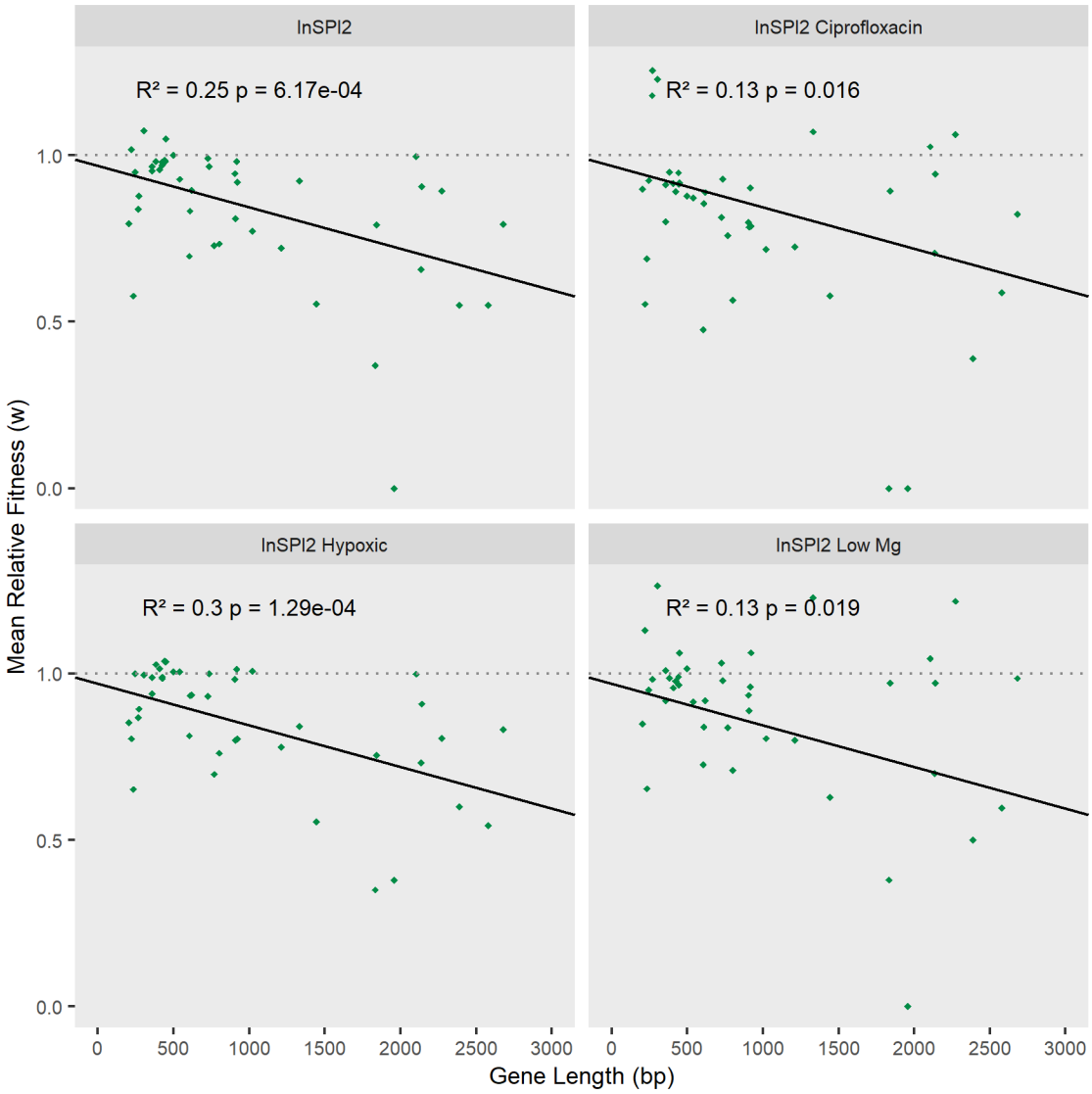

Figure 6 Relationship between relative fitness and gene length

Relative fitness of transferred *E.coli* orthologs in *S.* Typhimurium 4/74 plotted against the gene length in four growth environments. The black line is the regression between the two variables. Grey dotted line shows a fitness of 1. FDR corrected p-values and R^2 values are shown on the plot.

Figure 7

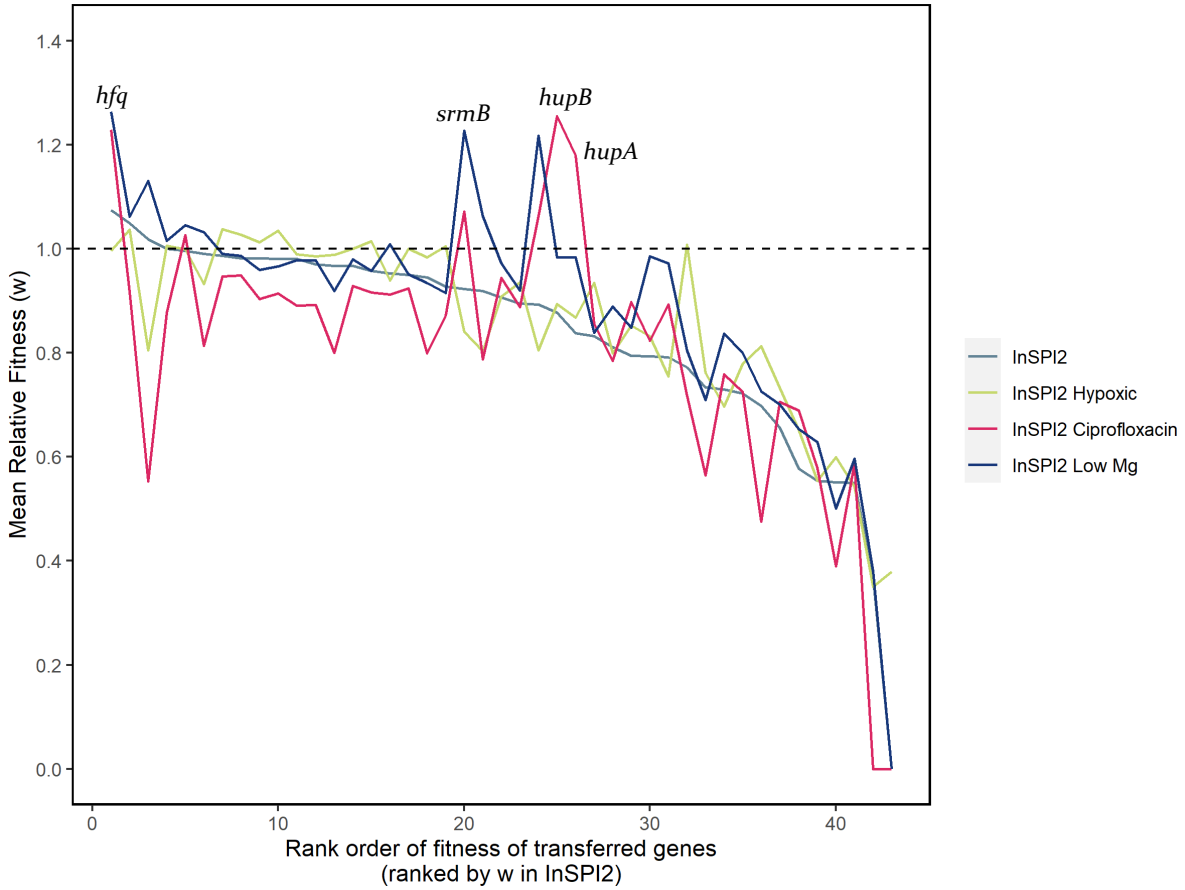

Figure 7 Relative fitness of transferred genes in four environments

The above figure shows the variation in fitness effects of individual genes between environments, few examples have been labelled. Genes are ranked according to the relative fitness measurements in the control environment (InSPI2) only. Black dashed line shows a fitness of 1.

Figure 8

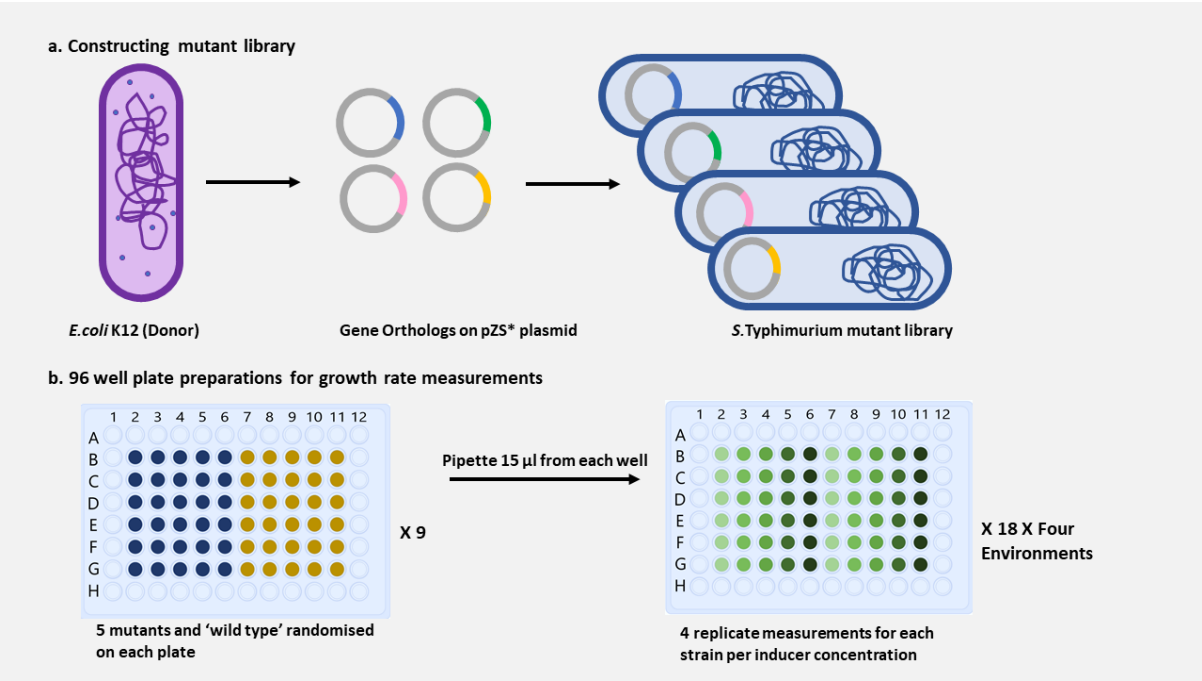

Figure 8 Graphical representation for preparation of mutant library and 96 well plates

a. shows the workflow for construction of *S.* Typhimurium strain 4/74 mutant library. **b.** left shows 96 well starter plate preparation, yellow and blue colours showing the two replicates for 5 mutants and a 'wild type'; and right shows 96 well plate preparation for growth measurements, shades of green from left to right showing increasing inducer concentrations.

Figure 9

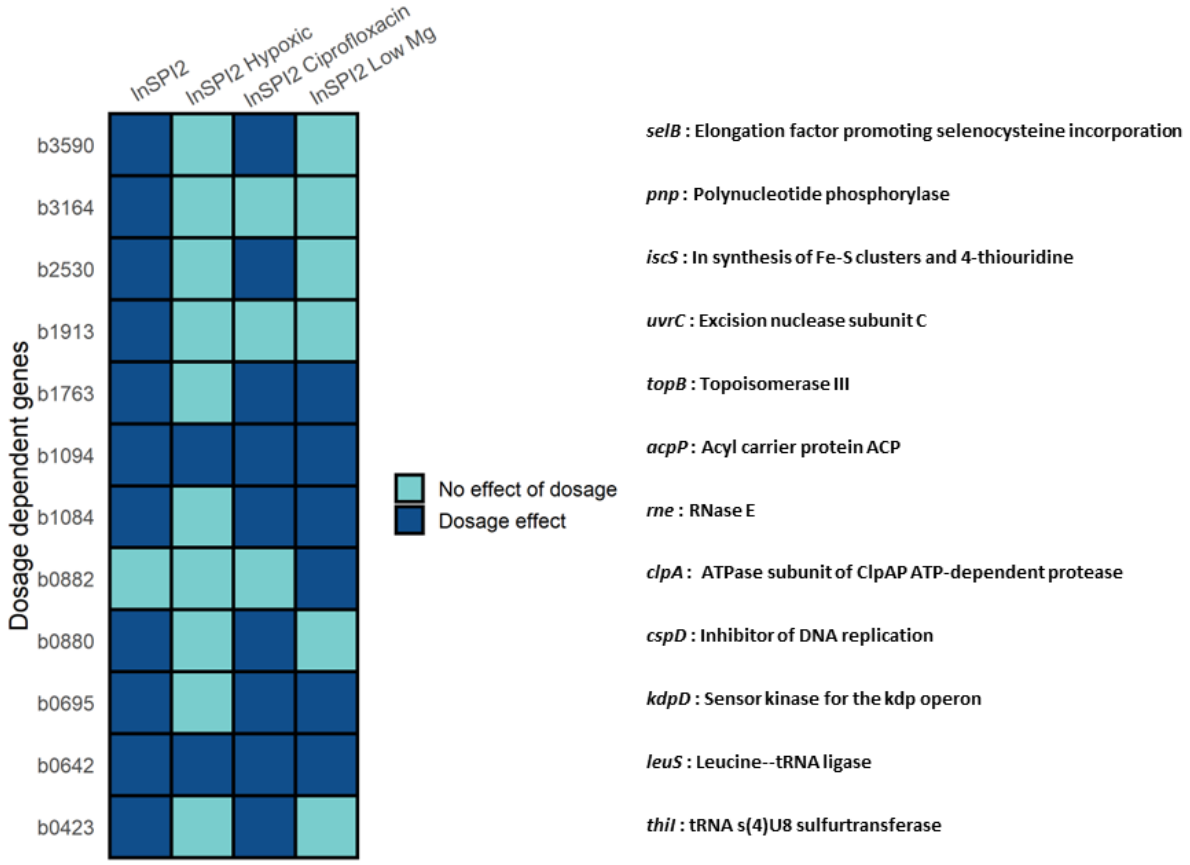

Figure 9 Dosage dependent genes

The figure shows *E.coli* orthologs (on y-axis) that exhibit a dosage effect across different environments (on x- axis) when expressed in *S.* Typhimurium 4/74. Gene names and functions are mentioned on the right.