Environment and the Evolutionary Trajectory of Horizontal Gene Transfer.

Rama P. Bhatia[1], Hande A. Kirit[1,2] and Jonathan P. Bollback[1*]

Affiliations:

[1] Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Liverpool, Merseyside, UK

[2] Laboratories of Molecular Anthropology and Microbiome Research, Stephenson Research and Technology Center, University of Oklahoma, USA

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 data supporting these predictions. Here, we address this gap by measuring the relative fitness of 44 genes horizontally transferred from *E. coli* to *S.* Typhimurium strain 4/74 in four infection relevant environments and estimated the distribution of fitness effects (DFE) in each environment. Roughly half of the genes, 40-50%, had significant fitness effects in at least one environment, with most genes showing significant gene by environment (G x E) interactions. We further identified 12 genes with dosage-dependent effects across the four environments, indicating that intolerance of the recipient strain to high protein dosage is a significant barrier to HGT. We also found longer genes had stronger fitness detriments than shorter ones, showing that gene length can significantly hinder HGT. In summary, a substantial fraction of transferred genes had a significant fitness cost on the recipient, with both gene characteristics and the environment providing 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. (Ochman et al., 2000). There is a growing need to explore factors that affect HGT, in order to improve our understanding of the evolution of bacterial genomes and thus our ability to design protective interventions.

Gene acquisition through HGT may exert a beneficial, neutral, or deleterious fitness effect on the recipient cell, with the nature of these effects determining the evolutionary fate of the newly transferred gene. While genes with a beneficial or neutral effect can persist in bacterial populations, genes with deleterious effects will most likely be purged (Popa and Dagan, 2011, Mira et al., 2001). The distribution of fitness effects (DFE) is an important predictor of HGT events. Therefore, understanding

the role of evolutionary barriers in defining the DFEs of horizontally transferred genes is critical if we wish to understand evolution by horizontal gene transfer. Therefore, understanding the role of factors affecting the probability or strength of deleterious fitness effects 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 (Ravenhall et al., 2015), 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 (Nakamura et al., 2004, Jain et al., 1999, Rivera et al., 1998), *(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 (Cohen et al., 2011, Wellner et al., 2007), *(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., Cas/CRISPR, H-NS) (Navarre, 2016, Labrie et al., 2010, Navarre et al., 2006, Lucchini et al., 2006), 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 (Papp et al., 2003).

Several studies have highlighted that the rate of HGT depends on the quality of the environment. For some examples, (i) The efficiency of transfer of a naturally occurring conjugative plasmid, encoding mercury resistance was studied in microcosms and *in situ*; the transfer frequencies were shown to be affected by temperature and nutrient concentrations (Bale et al., 1987), (ii) Conjugative transfer of IncP-1 plasmids was shown to be promoted at the air-liquid interface, whereas the transfer frequency was hindered on submerged *E. coli* K12 biofilms. Also, the efficiency of a conjugative transfer was dependent on the oxygen concentration in biofilms (Krol et al., 2011), (iii) Natural bacterial populations of the marine environment showed increased rates of prophage induction in response to environmental pollutants (Cochran et al., 1998).

Our current knowledge of horizontal gene transfer is thus mostly based on an understanding of genes/mutations that have survived the evolutionary sieve, but not of those that failed. We lack an in-depth understanding of evolutionary barriers that define the trajectory of HGT, and their interactions with the environment. In this study, we have tried to disentangle the influence of barriers and environments on the fitness effects of horizontally transferred genes. We do this through careful and precise measurement of the effects on growth rate of 44 *E. coli* orthologs horizontally transferred to *S.* typhimurium strain 4/74 in four different environments.

Materials and Methods

Bacterial Strains and Growth Conditions

Escherichia coli K-12 MG1655 and *Salmonella enterica* subsp. enterica serovar Typhimurium strain 4/74 (Hoiseth and Stocker, 1981) 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) (Xu and Hensel, 2010, Löber et al., 2006) were used for growing strains.

Experimental Approach

The experimental approach was to artificially transfer 44 orthologs from *E. coli* to *Salmonella* Typhimurium 4/74 using a low copy (3-4/cell) expression vector pZS* (Lutz and Bujard, 1997). The use of a low copy plasmid aided in tighter regulation of gene expression (Gruber et al., 2008). To control the expression of the transferred genes, the *tetR* repressor gene was introduced into the lambda phage attachment site on the *S*. Typhimurium chromosome using the modified pZS4Int plasmid (Acar Kirit et al., 2020, Lutz and Bujard, 1997) via recombineering (Murphy, 1998). Gene expression was induced using anhydrotetracycline (aTc) — a tetracycline derivative with no antibiotic activity — which interacts with TetR to induce gene expression (Ehrt et al., 2005).

Construction of pZS4-kan-tetR plasmid

The modified pZS4Int plasmid contains the *tetR* gene under the constitutive P_{N25} promoter, spectinomycin resistance gene, and origin of replication pSC101 (Acar Kirit et al., 2020, Lutz and Bujard, 1997). 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. Briefly, DNA fragments (0.03–0.2 pmols) were assembled into a vector using the NEBuilder HiFi DNA Assembly Master Mix at a vecor:insert ratio of 1:2. The assembly reaction mix was incubated in a thermocycler at 50°C for 4 hours. The ligation mix was cleaned using Zymo Research DNA Clean & Concentrator-5 PCR purification kit. The clean-up was then electroporated in Top10F(-) electrocompetent cells and plated on LB agar plates supplemented with 50µg/ml kanamycin (Kn^{50}). Transformants were screened, and correctly assembled plasmids were identified by PCR amplification and restriction enzyme digests. Stocks were prepared for colonies carrying the pZS4-kan-tetR plasmid in liquid LB supplemented with Kn^{50} and stored at -80°C with 15% glycerol.

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 (**Figure 1**) using λ red recombination plasmid pSIM5-tet (Koskiniemi et al., 2011, Datsenko and Wanner, 2000). For this, primers were designed to amplify the construct from the pZS4-kan-tetR plasmid, flanked by corresponding chromosomal position to be replaced with the construct. The construct size was verified by PCR on agarose gel and was subsequently gel extracted and purified for recombineering. For λ red recombination, a single colony of 4/74 strain carrying thermo-sensitive pSIM5-tet plasmid was inoculated in LB media supplemented with 25µg/ml tetracycline (Tc^{25}) and incubated for ~16 hours at 30°C and 220 rpm. A 1:100 dilution of the overnight culture in LB without NaCl, supplemented with Tc^{25} was incubated at 30°C to grow the cells to an optical density (OD_{600}) of

0.45 ±0.05. The cells were then incubated at 42°C in a water-bath for 15 mins with shaking to stimulate the expression of λ red genes. The culture was then transferred on ice and washed twice with cold sterile MilliQ water. Bacterial pellet was resuspended in 250 µl cold sterile 10% glycerol and 50 µl aliquots were prepared for storage at -80°C. The gel purified λ red fragment was then electroporated into competent 4/74 - pSIM5tet strain. Successful transformants were selected by plating on LB agar plates supplemented with Kn^{50} and screened for the correct insert using primers flanking the λ attachment site on 4/74 chromosome. After two rounds of streak purification on LB Kn^{50}, single colonies were tested on LB Tc^{25} for the loss of pSIM5-tet plasmid and the correct insert size was verified by PCR on agarose gel. A successful transformant was grown overnight in liquid LB and stored at -80°C with 15% glycerol. The integration site in the modified chromosome of 4/74 strain was Sanger sequenced to ensure no mutations were introduced during the recombineering process.

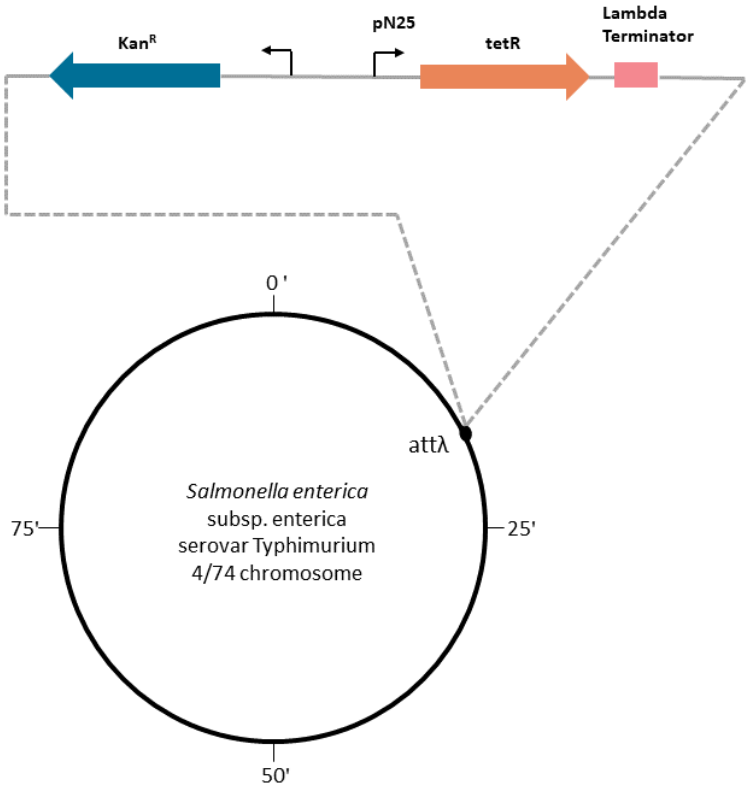


Figure 1 Schematic representation of *Salmonella enterica* serovar Typhimurium attλ::tetR-Kn^R 4/74 strain

Recipient strain for genes transferred from *Escherichia coli* K-12 MG1655.

Genes under study

The genes of interest to be artificially transferred were selected from a previous study (Acar Kirit et al., 2020). *Escherichia coli* str. K-12 substr. MG1655 (GenBank: U00096.2) was used as the donor. A random selection would be biased towards an overrepresentation of genes contained within large functional modules, among other biases. Therefore, an arbitrary selection was applied, although with some constraints, to ensure a good coverage of the donor's GC diversity, length, and approximately equal numbers of genes from the two distinct gene categories (informational genes and operational genes). To obtain a uniform distribution of protein-protein interactions (PPI), a database for *E.coli* genes whose interactions were experimentally validated by LCMS (liquid chromatography tandem mass spectrometry) and MALDI (matrix-assisted laser desorption/ionisation mass spectrometry) (Hu et al., 2009) was used to sample equal numbers of genes that had a representative range of PPI.

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 (Lutz and Bujard, 1997) and cloned plasmids were electroporated in *E. coli* in a previous study (**Supplementary Figure 1**). 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 (Acar Kirit et al., 2020). A copy of the 45 strains (44 *E.coli* genes and a 'wild type') were obtained to use in the present study. Plasmids were isolated using the ZR Plasmid Miniprep-Classic kit (Zymo Research). The insert size was confirmed by PCR on 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 (Amp^{50}), single colonies were grown overnight in liquid LB supplemented with Amp^{50} and stored at -80°C with 15% glycerol.

Environments under study

S. Typhimurium is an enteric pathogen infecting both humans and animals and is one of the serovars causing non-typhoidal Salmonella disease (NTS). It is also a frequently associated serovar with invasive NTS disease in sub-Saharan Africa (Hurley et al., 2020, Okoro et al., 2012, Gordon, 2011). *S.* Typhimurium 4/74 is a prototrophic bacteria and is highly virulent in cattle, pigs, mice, and chickens (Canals et al., 2019, Kroger et al., 2013). During the course of evolution, *Salmonella* has acquired virulence encoded *Salmonella* Pathogenicity Islands (SPI) through horizontal gene transfer. This has allowed the organism 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) (Hurley et al., 2020). To proliferate within eukaryotic cells, bacteria must adapt to a fluctuating intracellular environment.

A previous study from our research group highlights the substantial role of the environment in altering the distribution of fitness effects (DFEs) of a horizontally transferred gene. The data suggests high variability, with some genes becoming deleterious in one environment, while becoming neutral or even beneficial in another, suggesting that fluctuating environments may increase the likelihood of HGT (Acar Kirit et al., Unpublished).

In order to better understand the evolutionary trajectory of an HGT event, we designed three infection relevant *in vitro* growth environments that reflect the intra-macrophage lifestyle of *S.* Typhimurium (Kroger et al., 2013). In addition, we also tested an antibiotic environment, where the InSPI2 medium was supplemented with a fluoroquinolone class of drugs – ciprofloxacin, a first-line antimicrobial for many Gram-negative pathogens, including *Salmonella*. Several studies have demonstrated uptake and accumulation of ciprofloxacin by macrophages and polymorphonuclear leucocytes (Michot et al., 2004). A detailed description of the growth environments is outlined in **Table 1.**

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.

Table 1 Infection relevant growth environments

Growth Environment	**Description**
InSPI2	PCN medium (pH 5.8) supplemented with Amp^{50}
InSPI2 $LowMg^{2+}$	PCN (InSPI2) medium with 10 mM MgSO4 supplemented with Amp^{50}
InSPI2 Hypoxic	PCN (InSPI2) medium supplemented with Amp^{50} and paraffin oil overlay
InSPI2 Ciprofloxacin	PCN (InSPI2) medium supplemented with Amp^{50} and 0.1 µg/ml CIP

Gene expression on the plasmids was induced by addition of the inducer aTc. 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 linear range of RFU curve was considered as the optimal inducer concentration.

In addition to the external environment, we also wanted to address the effect of the intrinsic environment of a cell on fitness, after an HGT event. Functional similarity of the horizontally transferred gene to the endogenous copy can cause an imbalance due to increased protein dosage, thereby leading to a negative fitness effect. Therefore, we induced the transferred genes on the plasmids using multiple inducer concentrations thereby leading to increased protein levels in the cell. By comparing the fitness of each ortholog gene with the wild type, we can elucidate when gene dosage acts as an evolutionary barrier to HGT. The inducer concentrations used were aTc 0 (control), aTc 2ng/ml (below optimal expression), aTc 4ng/ml (optimal expression), aTc 6ng/ml (above optimal expression) and aTc 8ng/ml (saturation).

Growth Rates as a proxy for fitness

A 1:20 dilution of individual overnight cultures of the 45 *S.* Typhimurium 4/74 attλ::tetR-Kn^R, pZS* carrying mutants were grown in InSPI2 medium supplemented with Amp^{50} to OD_{600} 0.2. The position of the strains were randomised on 96 well plates to avoid bias and stored in -80°C with 15% glycerol as starter plates. Three copies were prepared for each of these starter randomised 96 well plates to minimise the freeze thaw cycles for strains. For each environment, cultivations were carried out in Polystyrene greyish-white square 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. **Figure 2** gives 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 250 rpm, except for InSPI2 Hypoxic, where cultures were grown on static. 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). The conversion of G-values to OD_{600} values was determined by means of a calibration curve and fitting the values to $Gvalue = b * OD_{600} \ /a$ (parameters a and b are determined by applying non-linear regression). To perform the calibration curve, 25 ml of overnight culture of *S.* Typhimurium 4/74 in InSPI2 was used to prepare serial dilutions (across a range of OD_{600} values) on a greyish-white square 96-half-deepwell microplate (CR1496dg, Enzyscreen) and the corresponding images were captured by the Growth Profiler 960 platform (Enzyscreen).

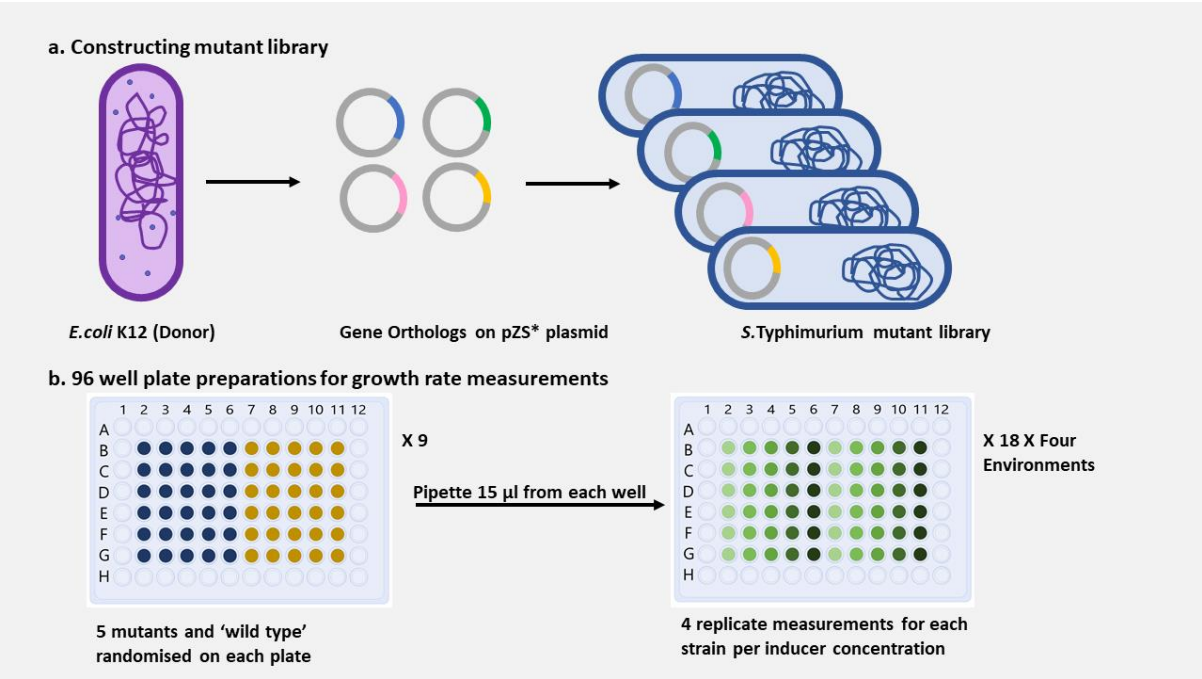


Figure 2 Graphical representation for preparation of mutant library and 96 well plates

a. shows the workflow for construction of *S.* Typhimurium strain 4/74 mutant library. **b.** left shows 96 well starter plate preparation, blue and yellow 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.

Statistical Analysis

Growth rates for all strains 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 study the effect of gene dosage on the DFEs of the transferred genes, 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 test assesses if the residual variances of the linear and quadratic model are significantly different using the Fischer–Snedecor F-test (Andrade and Gómez-Carracedo, 2013, Brüggemann et al., 2006).

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

Other factors affecting the DFEs of the transferred genes were studied using the fitness values for only the optimal inducer concentration (i.e., aTc 4ng/ml). To analyse if the fitness effects of informational genes was less than that of the operational genes, the 44 transferred *E.coli* orthologs were divided into the two groups using the Database for Clusters of Orthologous Groups (COGs) and a one-sided Wilcoxon Rank Sum test was applied on the mean fitness values of genes for 4 replicate measurements.

Additionally, we also looked at other intrinsic gene properties: the number of protein-protein (PPI), metabolic (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. Data used in the model is provided as **Supplementary Table 1**. The p-values obtained were corrected for multiple tests using the FDR method (Benjamini and Hochberg, 1995).

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

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 (Benjamini and Hochberg, 1995).

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 (Benjamini and Hochberg, 1995).

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

Results and Discussion

Distribution of Fitness Effects (DFEs)

Transferred genes may either exert a beneficial, neutral, or deleterious fitness effect on the recipient cell. We quantify this effect as relative fitness, or the effect on growth rate relative to a control strain of the same genetic background but lacking the transferred gene. 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 (Eyre-Walker and Keightley, 2007).

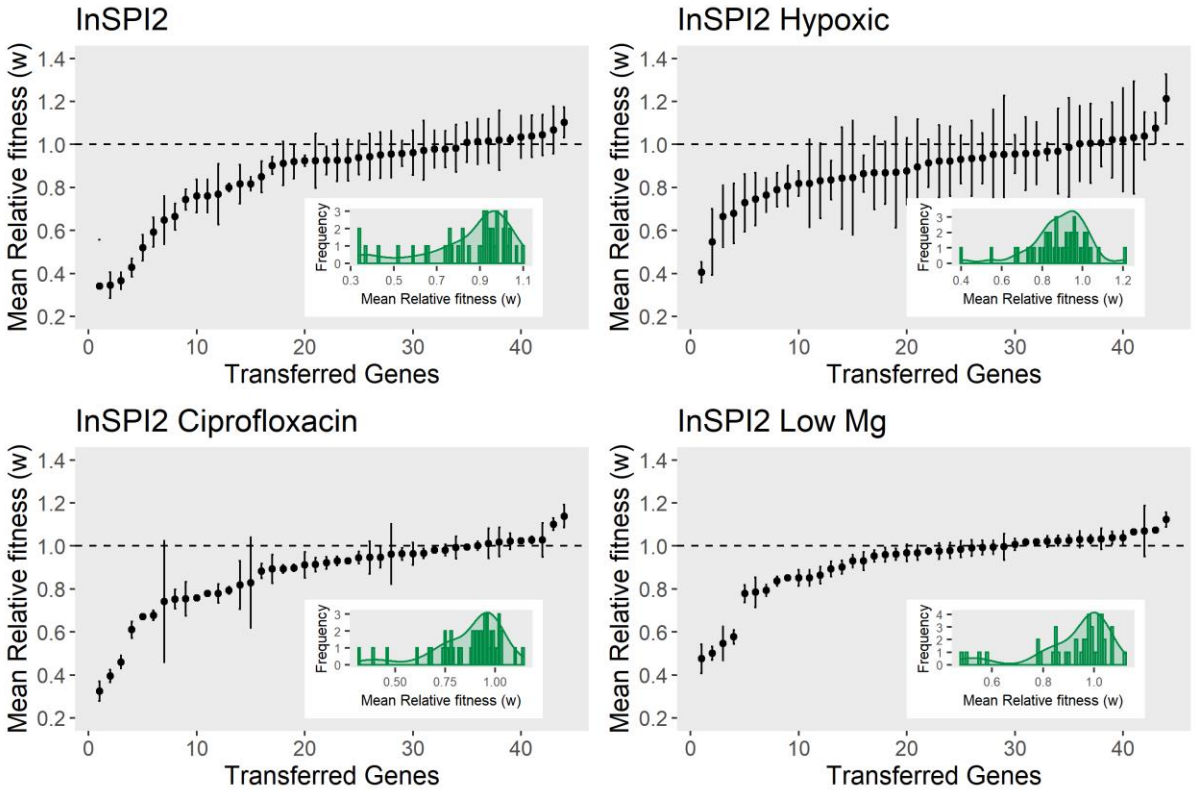


Figure 3 DFEs of transferred *E.coli* orthologs expressed in *S.* Typhimurium 4/74

The figure shows the distribution for the optimal inducer concentration (aTc 4ng/ml) for each of the growth environments. On the x-axis, transferred genes are sorted on 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.

Many studies have been carried out for the experimental determination of DFEs of new mutations. To mention a few examples, DFEs have been estimated for spontaneous and induced mutations in diploid yeast cell lines (Wloch et al., 2001), for random mutations in vesicular stomatitis virus (Sanjuan et al., 2004), for amino acid mutations in humans and *Drosophila melanogaster* (Keightley and Eyre-Walker, 2007, Eyre-Walker et al., 2006), for point mutations in genes (Lebeuf-Taylor et al., 2019). Parametric distributions like unimodal log-normal and gamma distributions, have been widely used to model and infer properties of DFEs (Kousathanas and Keightley, 2013). 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 (Eyre-Walker and Keightley, 2007).

In line with previous studies on experimental determination of DFEs, our study confirms a log-normal distribution for the DFEs of orthologs transferred from *E.coli* to *Salmonella* in each of the growth environments (μ_{InSPI2} = -0.189 , σ_{InSPI2} = 0.232 ; $\mu_{InSPI2\ Hypoxic}$ = -0.140, $\sigma_{InSPI2\ Hypoxic}$ = 0.158 ; $\mu_{InSPI2\ Ciprofloxacin}$ = -0.157 , $\sigma_{InSPI2\ Ciprofloxacin}$ = 0.193 ; $\mu_{InSPI2\ Low\ Mg}$ = -0.096 , $\sigma_{InSPI2\ Low\ Mg}$ = 0.161) (**Figure 3**).

Factors affecting fitness effects

Protein Dosage

Protein dosage is a potential barrier to HGT, as change in protein concentrations due to variation in gene copy numbers can cause an imbalance in the stoichiometry of the cell, thereby reducing bacterial fitness (Park and Zhang, 2012). Due to the genetic similarity of *E.coli* and *Salmonella* (McClelland et al., 2001), observed fitness effects due to increase in gene expression can be attributed to increased protein levels in the cell (i.e., a dosage effect). A previous study tested the hypothesis that host cells can be intolerant to increase in protein dosage due to functional similarity of the horizontally transferred gene to the endogenous copy (Sorek et al., 2007). The study shows 43 genes with > 98% identity to *E.coli* K12 could not be cloned from an *E.coli* HS strain into an *E.coli* host (DH10B), indicating intolerance to high gene dosage. They also identified a set of unclonable genes, found as a single-copy universally, implying that a high dosage of this set of genes can be detrimental universally.

To elucidate the role of protein dosage, the transferred *E.coli* orthologs were expressed in *S.* Typhimurium strain 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'. Based on the results from Mandel's test, we identified and categorised the genes into four dose response categories (**Figure 4**).

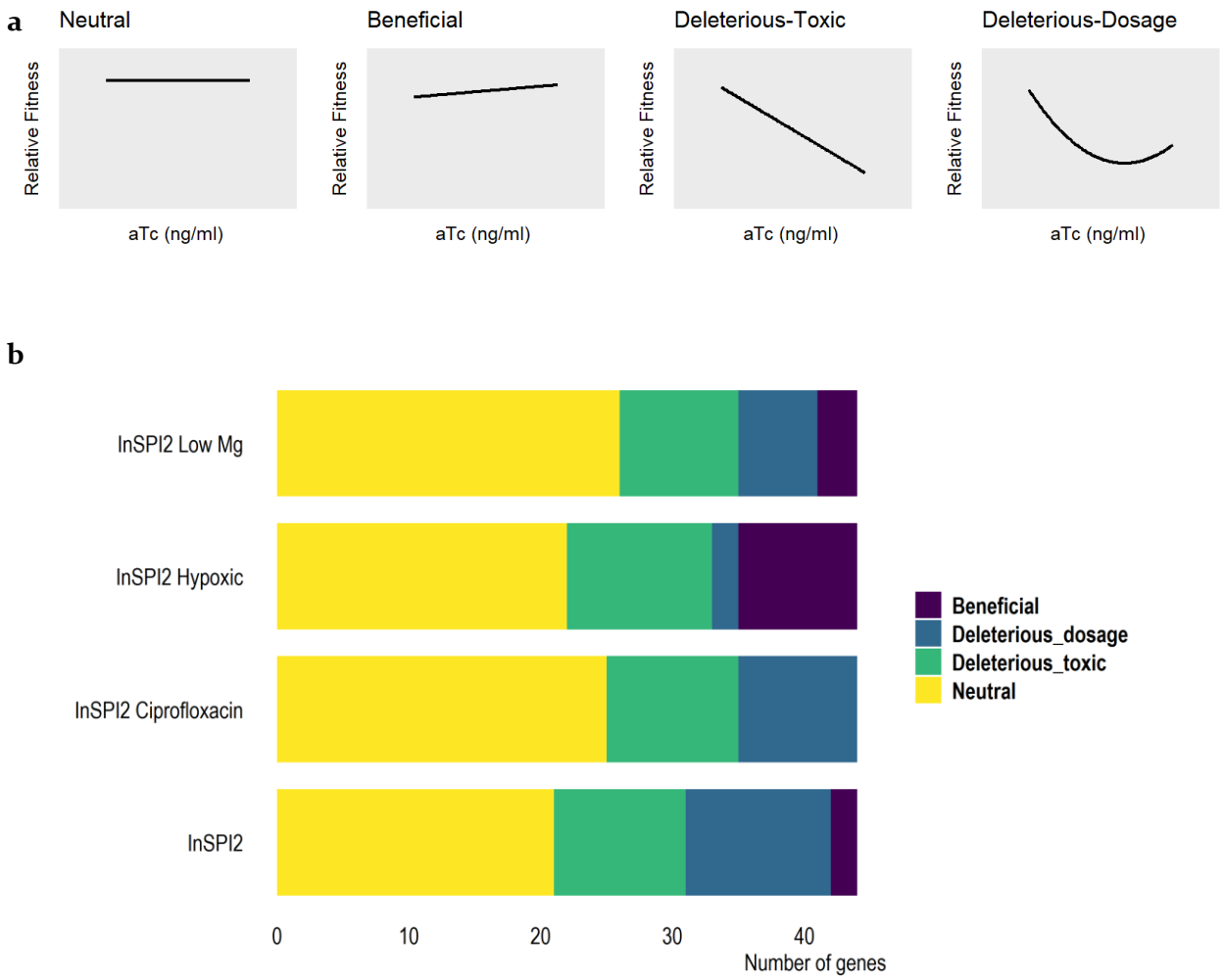


Figure 4 Categorizing genes based on dosage response

a. Graphical representation of the dosage trends observed for the horizontally transferred genes from *E.coli* K-12 to *S.* Typhimurium strain 4/74. **b.** Breakdown of 44 transferred genes based on the dosage response categories in each growth environment.

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) (**Figure 4 b**). In summary, 11 of the 44 transferred genes exhibited dosage dependence in a minimum of two growth environments (**Figure 5**). Classification of all genes across all environments are provided as **Supplementary Table 2**.

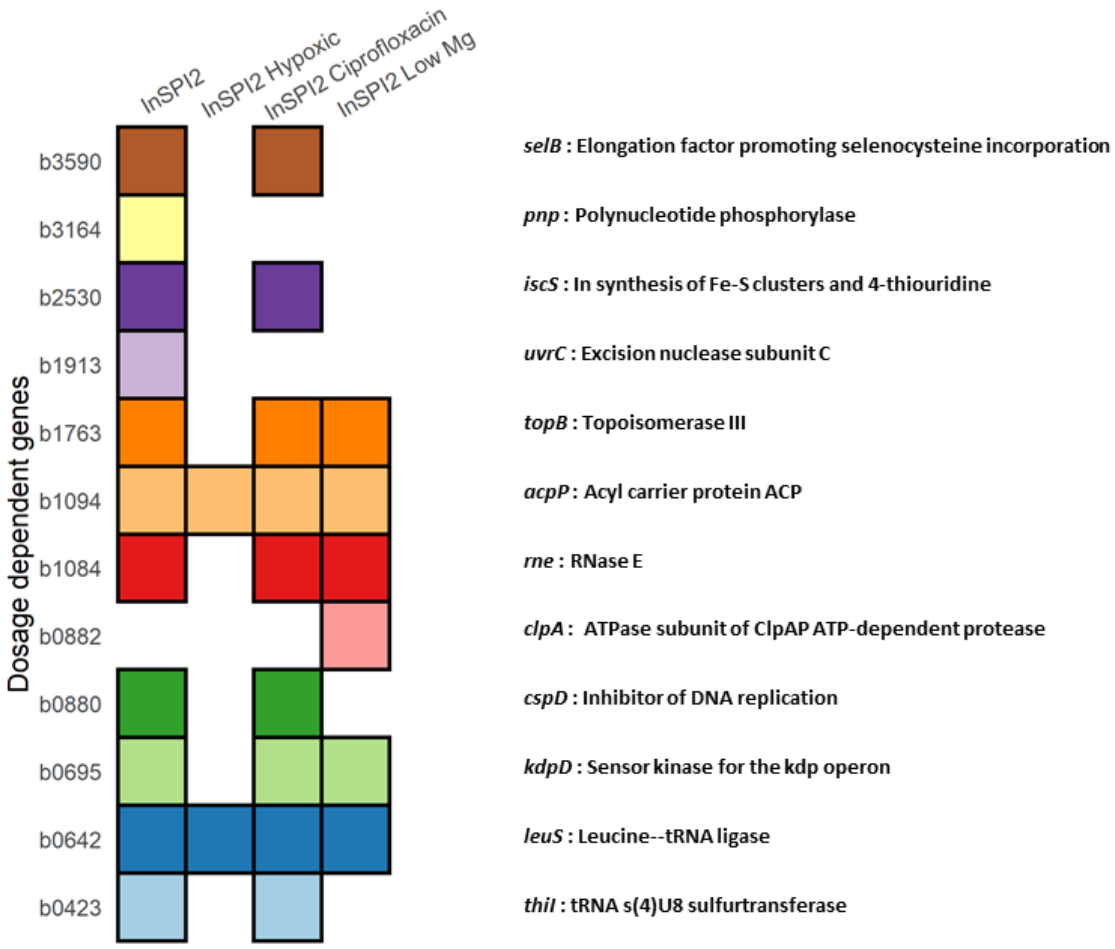


Figure 5 Dosage dependent genes

The coloured representation shows transferred *E.coli* orthologs, each assigned a colour (on y-axis) that exhibit a dosage effect across different environments (on x- axis) when expressed in *S.* Typhimurium 4/74. White spaces are indicative of the absence of dosage effects for the corresponding gene and environment combination. Gene names and functions are mentioned on the right.

Other factors affecting the DFEs of the transferred genes were studied using the fitness values for only the optimal inducer concentration (aTc 4ng/ml). They are detailed as follows.

Functional Categories

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 (Nakamura et al., 2004, Jain et al., 1999, Rivera et al., 1998). We separated our 44 genes informational and operational categories based on their COG annotations (**Table 2**) and tested whether or not the functional categories influenced the fitness effects of the transferred genes (excluding the ‘null‘ and ‘informational and operational’ categories).

Table 2 Genes grouped into functional categories

Functional Category	COG Categories	Number of genes
Informational	INFORMATION STORAGE AND PROCESSING	19
Operational	CELLULAR PROCESSES AND SIGNALING METABOLISM	20
Null	POORLY CHARACTERIZED	3
Informational and Operational	INFORMATION STORAGE AND PROCESSING CELLULAR PROCESSES AND SIGNALING METABOLISM	2

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 : $Med_{Informational} = 0.924$, $Med_{Operational} = 0.934$, Wilcoxon rank sum test, $W = 160$, $p = 0.205$; InSPI2 Hypoxic : $Med_{Informational} = 0.863$, $Med_{Operational} = 0.917$, Wilcoxon rank sum test, W

= 138, p = 0.074; InSPI2 Ciprofloxacin : Med $_{Informational}$ = 0.912, Med $_{Operational}$ = 0.937, Wilcoxon rank sum test, W = 167, p = 0.265; InSPI2 Low Mg : Med $_{Informational}$ = 0.968, Med $_{Operational}$ = 0.977, Wilcoxon rank sum test, W = 174, p = 0.333) (**Figure 6**). Although, we did not see a significant difference in fitness between the two categories, we observed that 6 out of 11 genes from the 'Deleterious-dosage' category were informational genes.

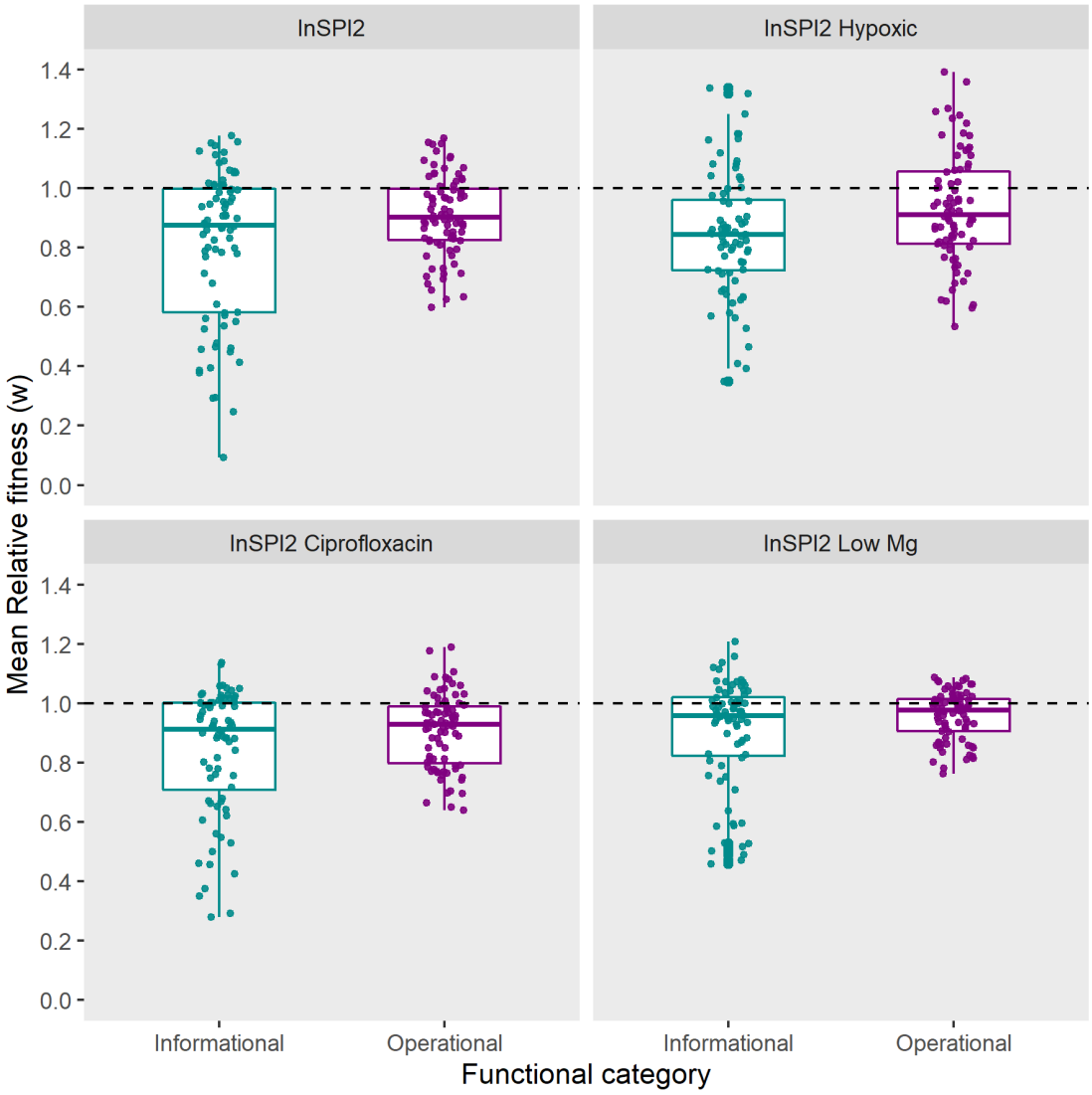


Figure 6 Fitness effects by functional category

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.

Complexity of Gene Interactions

The underlying rationale behind the complexity hypothesis is that the informational genes form part of large apparatuses, with complex gene product interaction assemblies, which inhibits their successful transfer, as opposed to the operational genes, that have fewer gene product interactions (Jain et al., 1999). The complexity hypothesis was revisited, and it was asserted that the number of protein-protein interactions, and not the gene function, acts as a barrier to HGT (Cohen et al., 2011). Increase in connectivity can result in multiple scenarios. Firstly, the newly acquired gene with multiple interactions, might fail to interact with its existing partners in the host cell, thereby making its acquisition effectively neutral. Secondly, such genes may properly interact with their existing partners, but may reduce host fitness by disrupting the cellular stoichiometry (Omer et al., 2010, Wellner and Gophna, 2008). A third possible scenario is, where genes interact improperly with the existing partners in the host and have a deleterious effect (Papp et al., 2003). The sampling of 44 *E.coli* orthologs ensured that the genes had a representative range (1 - 60) of PPI (Hu et al., 2009). To test if multiple interactions pose a barrier to HGT, we investigated the relationship between fitness of the transferred orthologs and their corresponding interactions at three levels (protein-protein, metabolic, and regulatory) in the recipient strain. Our data does not show support for gene interactions as evolutionary barriers to HGT (**Figure 7**).

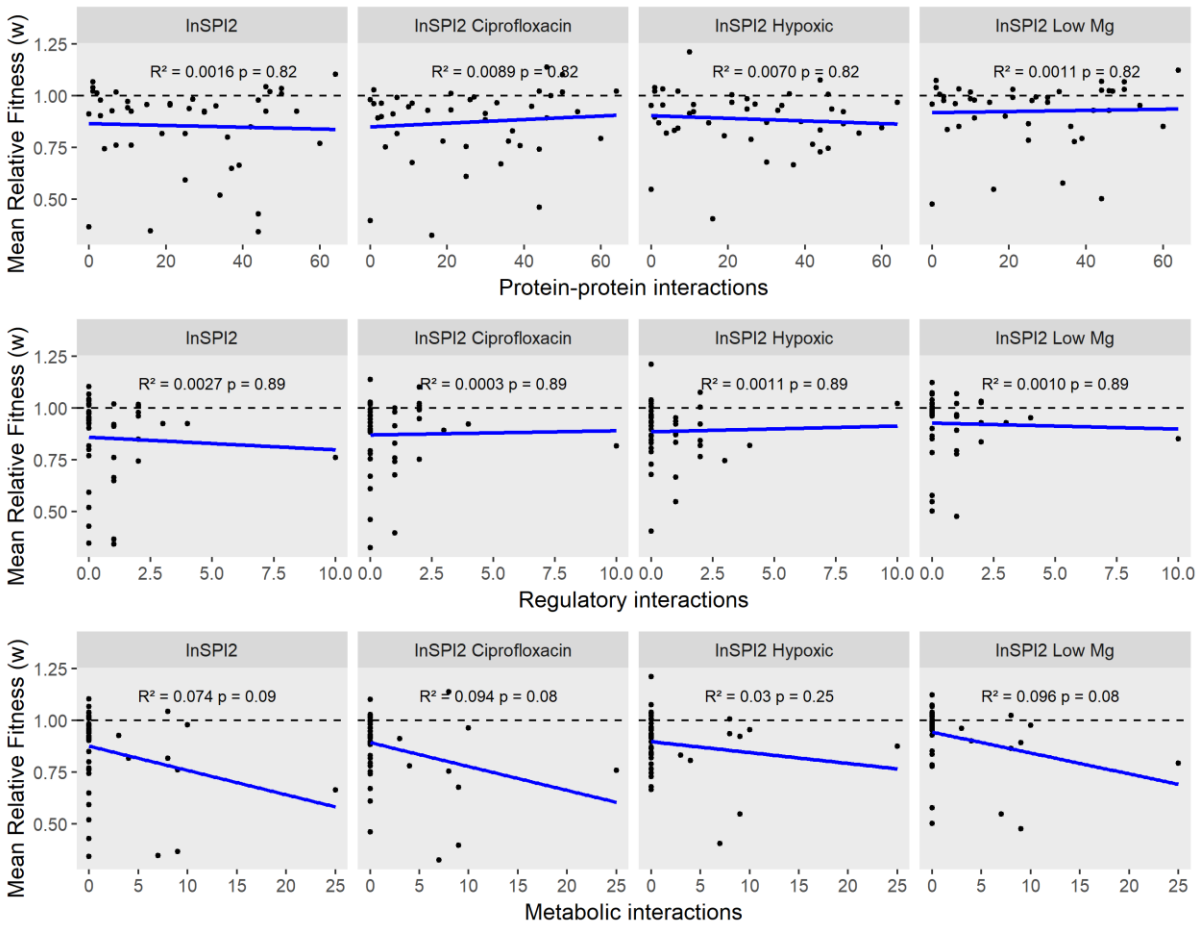


Figure 7 Relationship between relative fitness and number of interactions

Relative fitness of transferred *E.coli* orthologs in four growth environments plotted against the number of protein-protein interactions (top panel), regulatory interactions (middle panel) and metabolic interactions (bottom panel) in *S.* Typhimurium 4/74. The blue line is the regression between the two variables. Black dashed line shows a fitness of 1. FDR corrected p-values and R^2 values are shown on the plot.

There is computational evidence suggesting that barriers to HGT due to multiple interacting partners is not only limited to informational genes, and may be observed in all functional categories (Cohen et al., 2011). Horizontal gene acquisition within the *E. coli* metabolic network indicated that the frequency of gene transfers was higher among enzymes involved in nutrient uptake and metabolism compared to those involved in central reactions like biomass production. Evolution of metabolic networks is largely due to adaptation to new environments (Pal et al., 2005). There might be a potential trend for increasing number of metabolic interactions leading to reduced fitness of *S.* Typhimurium strain 4/74 (**Figure 7**

bottom panel). However, based on power analysis (Cohen, 1992), the likelihood of the regression model to detect a significant trend was found to be 45%, suggesting that an increased number of high complexity genes might be needed to identify the effect of increasing metabolic interactions on HGT.

Gene Length

The length of the gene can infer 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 (Lynch and Marinov, 2015). Ribosomal sequestration by lengthy genes will limit the production of essential proteins, lowering host fitness and being selected against in a population (Baltrus, 2013). We observed a statistically significant negative relationship between gene length and the fitness effects of the transferred genes in all growth environments (**Figure 8**).

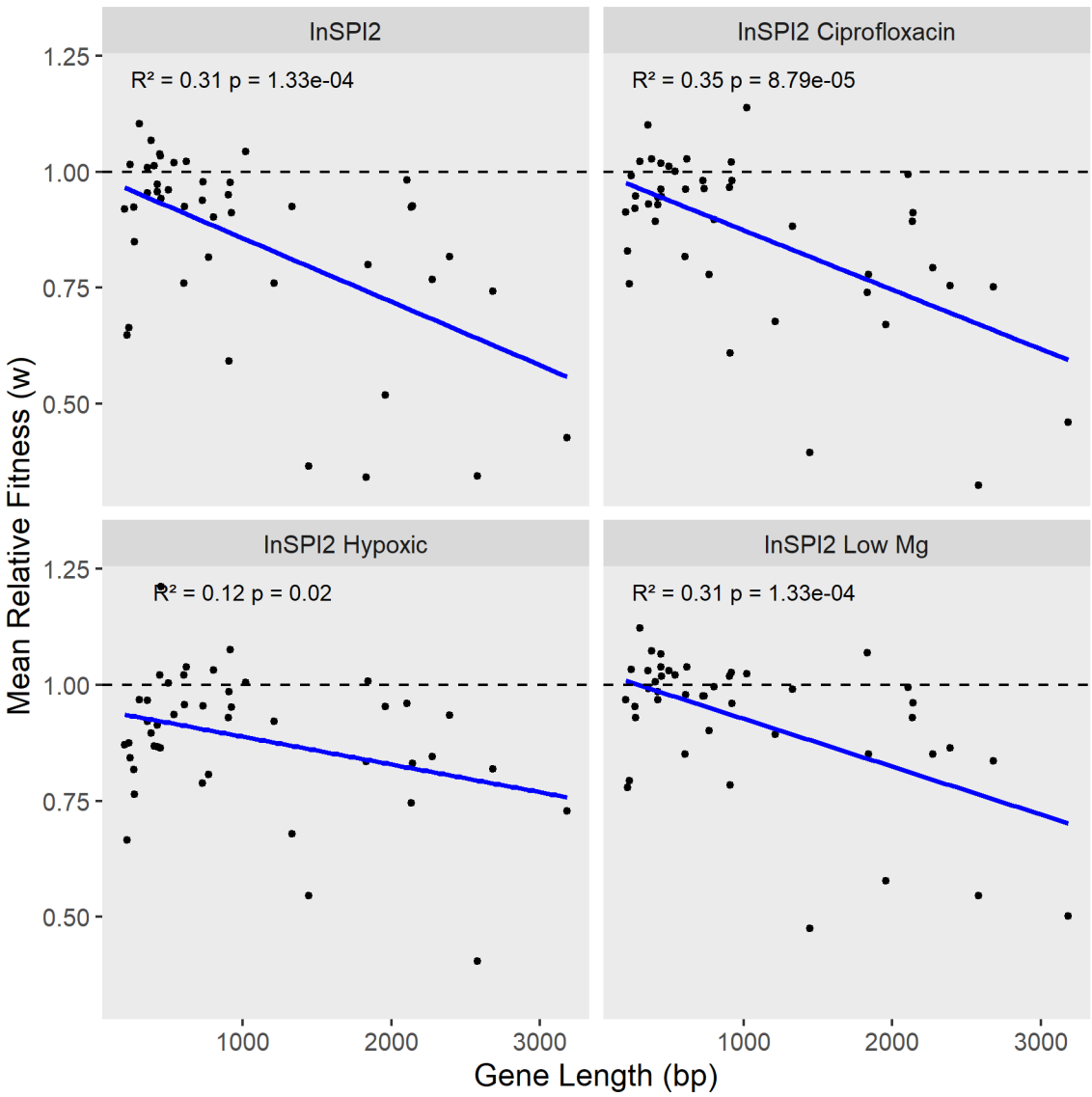


Figure 8 Relationship between relative fitness and gene length

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

GC Content and Codon Usage

The gene sequences within a species have a similar base composition, irrespective of the wide variation in the total GC content between bacterial species. 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 (Ravenhall et al., 2015, Lawrence and Ochman, 1997). 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 (Lucchini et al., 2006).

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 (Medrano-Soto et al., 2004). Codon usage bias can affect the translational efficiency and accuracy of the cell (Qian et al., 2012). 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 (Drummond and Wilke, 2008). The translation efficiency hypothesis asserts that there is an abundance of tRNA's for the preferred codons as opposed to the non-preferred codons in a cell, enabling the translation of the preferred codons at a faster rate. Therefore, a foreign gene with a codon usage bias for the unpreferred codons will affect the fitness of the cell, and thus act as a barrier to horizontal gene transfer. Experimental evidence supporting the translation efficiency hypothesis is available. For instance, a significant reduction in the levels of ADH (Alcohol Dehydrogenase) protein was observed, when unpreferred codons were introduced into the *Adh* gene in *Drosophila melanogaster* (Carlini and Stephan, 2003).

The effects of GC Content and Codon Usage on fitness were examined individually. We compared the absolute deviation in GC content between the transferred *E. coli* genes and their *S.* Typhimurium orthologs with the observed fitness effects in four growth environments and did not find a significant relationship between the two variables (**Figure 9** top panel**)**. Similarly, the absolute deviation in FOP

between the transferred *E. coli* genes and their *S.* Typhimurium copies could not explain the observed fitness effects in any environment (**Figure 9** bottom panel).

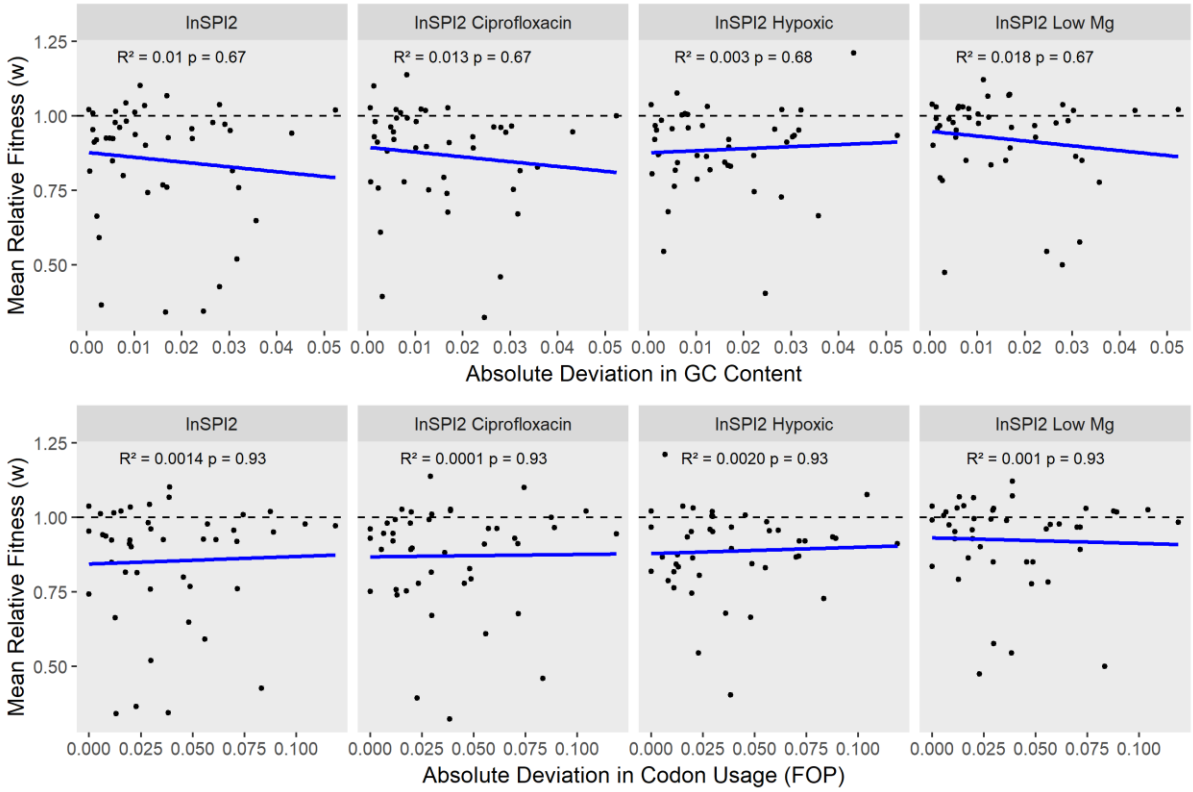


Figure 9 Relationship of GC content and Codon Usage with fitness

Relative fitness of transferred *E.coli* orthologs in *S.* Typhimurium 4/74 plotted against absolute deviation in GC Content (top panel) and absolute deviation in FOP (bottom panel) in four growth environments. The blue line is the regression between the two variables. Black dashed line shows a fitness of 1. FDR corrected p-values and R^2 values are shown on the plot.

Effect of the Environment on DFEs

Substantial dependency of fitness effects of genes on environmental niches has been well documented in studies of antibiotic resistance genes and metabolic enzymes (Zhong et al., 2020, Melnyk et al., 2015). There are three possible outcomes for how the environment shapes the DFEs of newly transferred genes. Under the first possibility, the environment may not have any effect on the new gene. Since, the mutants and the 'wild type' were compared under the same environmental conditions, this would result in identical DFEs for each environment. A second possibility is where, the environment affects the fitness of all mutants equally, resulting in an overall shift of the DFEs while maintaining the rank order

of the mutants. Lastly, the environment might affect the fitness of each gene differentially, indicative of strong gene by environment interactions. Under this outcome, the fitness effect of a gene becomes unpredictable with a shift in environment.

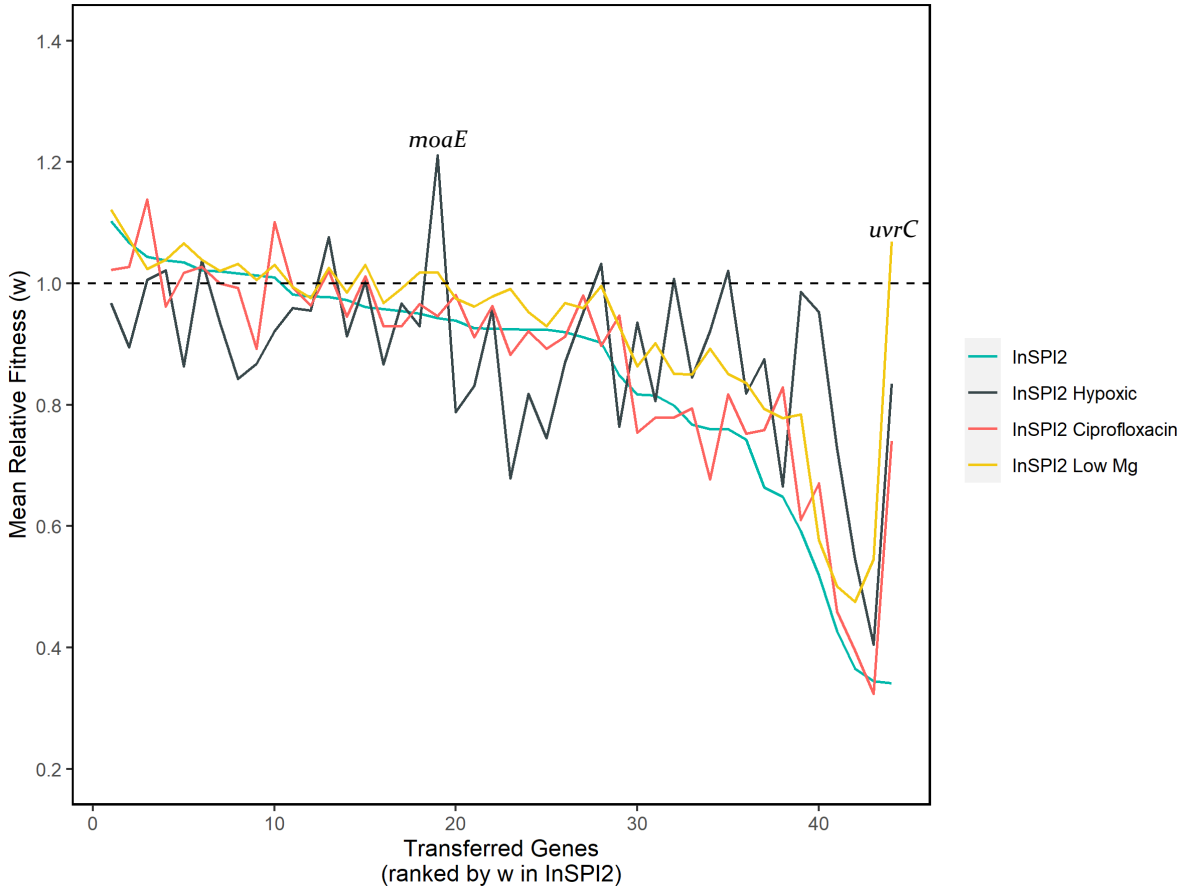


Figure 10 Relative fitness of transferred genes in four environments

The above figure shows the variation in fitness effects of individual genes between environments, two 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.

Based on the possible outcomes for G x E interactions outlined above, our data (**Figure 10**) can be explained by the third scenario, indicating each gene being affected by the environment differently, likely based on their functions. Hence, we investigated if the environment affected the central tendency, shape, or spread of the DFEs.

Central Tendency

The median fitness effects of DFEs were compared using the Wilcoxon Signed Rank test for each pairwise comparison of the four environments. The median fitness effects for InSPI2 Low Mg environment were found to be significantly different from InSPI2 (P_{adj} < 0.001, Z-statistic = -6.540) and InSPI2 Ciprofloxacin (P_{adj} < 0.001, Z-statistic = -4.380) environments. The DFEs for InSPI2, InSPI2 Hypoxic and InSPI2 Ciprofloxacin were not found to be significantly different from each other. This is surprising, as a much larger variation in fitness effects of a gene between these three environments is evident from **Figure 10**.

Shape and Spread

The Kolmogorov-Smirnov (K-S) test is sensitive to differences in shape of the sample distributions and was therefore used to detect differences in the shape and spread of DFEs between environments. We observed that none of the distributions were significantly different from each other in terms of shape and spread. This potentially indicates that shape and spread are intrinsic properties for DFEs of horizontal gene transfer events between closely related species.

Gene by Environment Interactions

Since some environments altered the central tendency of DFEs significantly in the study, we posed the question – Do environments differentially affect the fitness effects of different genes? We observed a strong interaction between individual genes and the environment ($F_{129,528}$ = 3.039, p < 0.001). All pairwise comparisons of gene by environment interactions between the four environments were found to be significant. At the single gene level, these gene by environment interactions show a completely unpredictable distribution of fitness effects (**Figure 10**), indicating the substantial role of the environment in shaping the fitness effects of transferred genes. Interestingly, this is also true for environments with median fitness effects not significantly different from each other.

In this study, we investigated if any, evolutionary barriers exist, and what the role of environment is on the distribution of fitness effects of genes transferred from *E.coli* to *Salmonella*. The study excludes the dominance of insertional effects on DFEs of HGT events as the genes were transferred on a low copy

plasmid and their expression was induced by the repressor construct inserted into a defined position in the *S.* Typhimurium chromosome, with no detectable fitness cost. Also, there was no detectable cost to carrying the low copy plasmid used for gene transfers (Acar Kirit et al., 2020). A previous study by Acar Kirit et al., 2020 looking at the DFEs obtained by transferring the same set of 44 orthologs from *Salmonella enterica* serovar Typhimurium strain LT2 to *E. coli* in minimal media using gene frequencies as a proxy for fitness, showed 37 out of 44 genes to impose a significant fitness cost on the host. The study also highlighted the importance of conducting experimental fitness measurements in natural environments, 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 simulate signals encountered by *S.* Typhimurium during an intracellular infection (Kroger et al., 2013). We hypothesize that the magnitude of fitness effects observed in the study might be close to what we would observe intracellularly.

Our results show that on average, 17 genes were deleterious, 3 genes were beneficial, and 24 genes were neutral across the four environments studied. In each environment, approximately 50 - 60 % of the genes were found to have fitness effects not significantly different from 1 (**Figure 4**), suggesting that a large fraction of transfers may be neutral, i.e., its persistence/fixation in a population depends on random forces such as genetic drift (Eyre-Walker and Keightley, 2007) or genetic draft (Gillespie, 2000).

It is very rare for horizontally acquired genes imposing a neutral effect to fix in a population (Kimura, 1983). Although, horizontally acquired neutral genes may not extend an immediate selective advantage to the population, but they may provide a genetically diverse pool as raw material for selection to act upon. Further, adaptation due to selection pressures imposed by environmental stressors can promote a neutral gene to be important for evolving novel beneficial functions (Knoppel et al., 2014).

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. We have identified protein dosage and toxicity as barriers to HGT. We

identified 12 dosage dependent genes that, when transferred into *S.* Typhimurium 4/74, dramatically reduced its fitness. Of these, 11 genes consistently exhibited dosage dependency regardless of the environment, suggesting an increased intolerance of *S.* Typhimurium to a high dosage of these genes in an intracellular environment. InSPI2 Hypoxic had the least number of genes exhibiting dosage dependency amongst all environments, suggesting that hypoxia/low oxygen availability may be an important component of the intracellular environment offering better opportunities for *S.* Typhimurium to undergo adaptive evolution. In addition, the length of the transferred gene was identified as significant barrier to HGT.

Other hypothesized barriers to HGT - functional categories, gene interactions, GC content and codon usage bias could not significantly explain the fitness effects of transferred genes. This is likely due to the fact that *E. coli* and *Salmonella* are genetically similar and orthologs have minimal differences in codon usage and GC content, which may prevent us from detecting the role of these factors as prominent barriers to HGT.

Our study also reveals the influential role of the environment in the HGT trajectory. For genes with substantial variability in fitness effects between environments (**Figure 10**), the host cells become more tolerant to having these genes under specific conditions, i.e., some genes can have a strong fitness cost in one environment, while becoming neutral or even beneficial in another. Such genes might therefore be positively selected by fluctuating environments. In other words, the likelihood of a successful HGT event might increase in a heterogeneous or fluctuating environment. The strong G x E interactions can be better understood by analysing the gene regulatory networks in the host. For example, the *moaE* gene was shown to increase *S.* Typhimurium 4/74 fitness in the InSPI2 Hypoxic environment (**Figure 10**). The *moaE* gene is a part of the *moaABCDE* operon that is regulated by the FNR transcriptional master switch, which senses oxygen availability. The *moaABCDE* operon is involved in the Molybdenum cofactor (Moco) biosynthesis. This cofactor serves as the catalytic centre of Molybdoenzymes that are essential for anaerobic respiration (Zupok et al., 2019). Similarly, the *uvrC* gene increased the fitness of 4/74 in all environments except in InSPI2 (**Figure 10**). This could be attributed to its role in SOS response and DNA repair.

Conclusion

Although comparative genomics can provide valuable insight into the extent of occurrences of gene transfer events, it does not necessarily provide information on the evolutionary fate of horizontally acquired genes, including those quickly purged by selection. Hence, we have adopted the approach of systematic experimental evolution to understand the factors affecting HGT. We show that the environment and protein dosage of the transferred genes can have some level of interaction to collectively determine the outcome of an HGT event. Also, the length of the gene transferred is an important predictor for the outcome of HGT.

In addition, we also show that the fitness effects of horizontally transferred genes are highly dependent on the environment. Some of these genes may contribute to *Salmonella* virulence indirectly, for example by facilitating processes that aid bacteria survival. Given a small sample size in the experimental approach, our study not only provides a systematic experimental understanding of trends in HGT, but also highlights the importance of investigating the role played by genes indirectly involved in pathogenicity to better understand invasiveness, resistance, and environment adaptation of *Salmonella*.

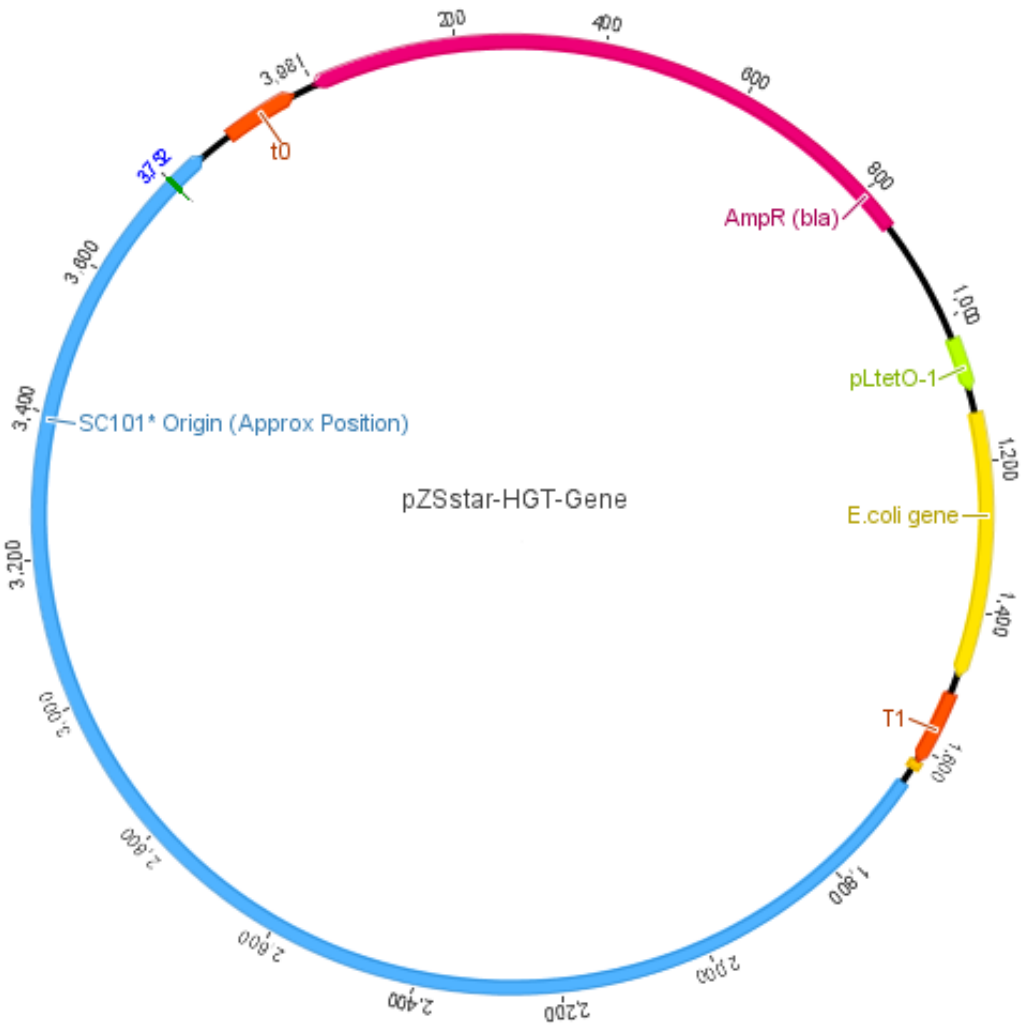
Acknowledgements

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*; and to, Simon (Xiaojun) Zhu for his guidance with using Growth Profiler 960. 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.

Author Contributions

RPB, HAK and JPB designed the experiments together. RPB performed wet-lab experiments. RPB performed statistical analysis. RPB wrote the initial draft of the manuscript and revised it together with HAK and JPB.

Supplementary Figures



Supplementary Figure 1 Expression plasmid used to construct the mutant library

The pZS* plasmid consisting of an SC101 origin of replication (blue), terminators (t0 and T1)(orange), Ampicillin resistance gene (pink), transferred *E.coli* gene (yellow) under the control of the inducible p_{LtetO1} promoter (green).

References

ACAR KIRIT, H., LAGATOR, M. & BOLLBACK, J. P. 2020. Experimental determination of evolutionary barriers to horizontal gene transfer. *BMC Microbiol,* 20**,** 326.

ACAR KIRIT, H., LAGATOR, M., POECHER, K. & BOLLBACK, J. P. Unpublished. The role of environment on bacterial gene acquisition

ANDRADE, J. M. & GÓMEZ-CARRACEDO, M. P. 2013. Notes on the use of Mandel's test to check for nonlinearity in laboratory calibrations. *Analytical Methods,* 5**,** 1145-1149.

BALE, M. J., FRY, J. C. & DAY, M. J. 1987. Plasmid transfer between strains of Pseudomonas aeruginosa on membrane filters attached to river stones. *J Gen Microbiol,* 133**,** 3099-107.

BALTRUS, D. A. 2013. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol,* 28**,** 489-95.

BENJAMINI, Y. & HOCHBERG, Y. 1995. 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.

BRÜGGEMANN, L., QUAPP, W. & WENNRICH, R. 2006. Test for non-linearity concerning linear calibrated chemical measurements. *Accreditation and Quality Assurance,* 11**,** 625-631.

CANALS, R., HAMMARLOF, D. L., KROGER, C., OWEN, S. V., FONG, W. Y., LACHARME-LORA, L., ZHU, X., WENNER, N., CARDEN, S. E., HONEYCUTT, J., MONACK, D. M., KINGSLEY, R. A., BROWNRIDGE, P., CHAUDHURI, R. R., ROWE, W. P. M., PREDEUS, A. V., HOKAMP, K., GORDON, M. A. & HINTON, J. C. D. 2019. Adding function to the genome of African Salmonella Typhimurium ST313 strain D23580. *PLoS Biol,* 17**,** e3000059.

CARLINI, D. B. & STEPHAN, W. 2003. In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein. *Genetics,* 163**,** 239-43.

COCHRAN, P. K., KELLOGG, C. A. & PAUL, J. H. 1998. Prophage induction of indigenous marine lysogenic bacteria by environmental pollutants. *Marine Ecology Progress Series,* 164**,** 125-133.

COHEN, J. 1992. Statistical Power Analysis. *Current Directions in Psychological Science,* 1**,** 98-101.

COHEN, O., GOPHNA, U. & PUPKO, T. 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol,* 28**,** 1481-9.

DATSENKO, K. A. & WANNER, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences,* 97**,** 6640-6645.

DRUMMOND, D. A. & WILKE, C. O. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell,* 134**,** 341-52.

EHRT, S., GUO, X. V., HICKEY, C. M., RYOU, M., MONTELEONE, M., RILEY, L. W. & SCHNAPPINGER, D. 2005. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res,* 33**,** e21.

EYRE-WALKER, A. & KEIGHTLEY, P. D. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet,* 8**,** 610-8.

EYRE-WALKER, A., WOOLFIT, M. & PHELPS, T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics,* 173**,** 891-900.

GILLESPIE, J. H. 2000. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics,* 155**,** 909-19.

GORDON, M. A. 2011. Invasive nontyphoidal Salmonella disease: epidemiology, pathogenesis and diagnosis. *Curr Opin Infect Dis,* 24**,** 484-9.

GRUBER, D. F., PIERIBONE, V. A., PORTON, B. & KAO, H. T. 2008. Strict regulation of gene expression from a high-copy plasmid utilizing a dual vector system. *Protein Expr Purif,* 60**,** 53-7.

HOISETH, S. K. & STOCKER, B. A. 1981. Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature,* 291**,** 238-9.

HU, P., JANGA, S. C., BABU, M., DIAZ-MEJIA, J. J., BUTLAND, G., YANG, W., POGOUTSE, O., GUO, X., PHANSE, S., WONG, P., CHANDRAN, S., CHRISTOPOULOS, C., NAZARIANS-ARMAVIL, A., NASSERI, N. K., MUSSO, G., ALI, M., NAZEMOF, N., EROUKOVA, V., GOLSHANI, A., PACCANARO, A., GREENBLATT, J. F., MORENO-HAGELSIEB, G. & EMILI, A. 2009. Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. *PLoS Biol,* 7**,** e96.

HURLEY, D., HOFFMANN, M., MURUVANDA, T., ALLARD, M. W., BROWN, E. W., MARTINS, M. & FANNING, S. 2020. Atypical Salmonella enterica Serovars in Murine and Human Macrophage Infection Models. *Infect Immun,* 88.

JAIN, R., RIVERA, M. C. & LAKE, J. A. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences,* 96**,** 3801.

KEIGHTLEY, P. D. & EYRE-WALKER, A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics,* 177**,** 2251-61.

KIMURA, M. 1983. *The neutral theory of molecular evolution*, Cambridge University Press.

KNOPPEL, A., LIND, P. A., LUSTIG, U., NASVALL, J. & ANDERSSON, D. I. 2014. Minor fitness costs in an experimental model of horizontal gene transfer in bacteria. *Mol Biol Evol,* 31**,** 1220-7.

KOSKINIEMI, S., PRANTING, M., GULLBERG, E., NASVALL, J. & ANDERSSON, D. I. 2011. Activation of cryptic aminoglycoside resistance in Salmonella enterica. *Mol Microbiol,* 80**,** 1464-78.

KOUSATHANAS, A. & KEIGHTLEY, P. D. 2013. A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics,* 193**,** 1197-208.

KROGER, C., COLGAN, A., SRIKUMAR, S., HANDLER, K., SIVASANKARAN, S. K., HAMMARLOF, D. L., CANALS, R., GRISSOM, J. E., CONWAY, T., HOKAMP, K. & HINTON, J. C. 2013. An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe,* 14**,** 683-95.

KROL, J. E., NGUYEN, H. D., ROGERS, L. M., BEYENAL, H., KRONE, S. M. & TOP, E. M. 2011. Increased transfer of a multidrug resistance plasmid in Escherichia coli biofilms at the air-liquid interface. *Appl Environ Microbiol,* 77**,** 5079-88.

LABRIE, S. J., SAMSON, J. E. & MOINEAU, S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol,* 8**,** 317-27.

LAWRENCE, J. G. & OCHMAN, H. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol,* 44**,** 383-97.

LEBEUF-TAYLOR, E., MCCLOSKEY, N., BAILEY, S. F., HINZ, A. & KASSEN, R. 2019. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *Elife,* 8.

LUCCHINI, S., ROWLEY, G., GOLDBERG, M. D., HURD, D., HARRISON, M. & HINTON, J. C. 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog,* 2**,** e81.

LUTZ, R. & BUJARD, H. 1997. Independent and tight regulation of transcriptional

units in Escherichia coli via the LacR/O, the TetR/O

and AraC/I1-I2 regulatory elements. *Nucleic Acids Research,* 25.

LYNCH, M. & MARINOV, G. K. 2015. The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A,* 112**,** 15690-5.

LÖBER, S., JACKEL, D., KAISER, N. & HENSEL, M. 2006. Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. *Int J Med Microbiol,* 296**,** 435-47.

MCCLELLAND, M., SANDERSON, K. E., SPIETH, J., CLIFTON, S. W., LATREILLE, P., COURTNEY, L., PORWOLLIK, S., ALI, J., DANTE, M., DU, F., HOU, S., LAYMAN, D., LEONARD, S., NGUYEN, C., SCOTT, K., HOLMES, A., GREWAL, N., MULVANEY, E., RYAN, E., SUN, H., FLOREA, L., MILLER, W., STONEKING, T., NHAN, M., WATERSTON, R. & WILSON, R. K. 2001. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature,* 413**,** 852-6.

MEDRANO-SOTO, A., MORENO-HAGELSIEB, G., VINUESA, P., CHRISTEN, J. A. & COLLADO-VIDES, J. 2004. Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol,* 21**,** 1884-94.

MELNYK, A. H., WONG, A. & KASSEN, R. 2015. The fitness costs of antibiotic resistance mutations. *Evol Appl,* 8**,** 273-83.

MICHOT, J. M., VAN BAMBEKE, F., MINGEOT-LECLERCQ, M. P. & TULKENS, P. M. 2004. Active efflux of ciprofloxacin from J774 macrophages through an MRP-like transporter. *Antimicrob Agents Chemother,* 48**,** 2673-82.

MIRA, A., OCHMAN, H. & MORAN, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends in Genetics,* 17.

MURPHY, K. C. 1998. Use of Bacteriophage Lambda Recombination Functions To Promote

Gene Replacement in Escherichia coli. *Journal of Bacteriology,* 180.

NAKAMURA, Y., ITOH, T., MATSUDA, H. & GOJOBORI, T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet,* 36**,** 760-6.

NAVARRE, W. W. 2016. The Impact of Gene Silencing on Horizontal Gene Transfer and Bacterial Evolution. *Adv Microb Physiol,* 69**,** 157-186.

NAVARRE, W. W., PORWOLLIK, S., WANG, Y., MCCLELLAND, M., ROSEN, H., LIBBY, S. J. & FANG, F. C. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science,* 313**,** 236-8.

OCHMAN, H., LAWRENCE, J. G. & GROISMAN, E. A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature,* 405.

OKORO, C. K., KINGSLEY, R. A., CONNOR, T. R., HARRIS, S. R., PARRY, C. M., AL-MASHHADANI, M. N., KARIUKI, S., MSEFULA, C. L., GORDON, M. A., DE PINNA, E., WAIN, J., HEYDERMAN, R. S., OBARO, S., ALONSO, P. L., MANDOMANDO, I., MACLENNAN, C. A., TAPIA, M. D.,

LEVINE, M. M., TENNANT, S. M., PARKHILL, J. & DOUGAN, G. 2012. Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet,* 44**,** 1215-21.

OMER, S., KOVACS, A., MAZOR, Y. & GOPHNA, U. 2010. Integration of a foreign gene into a native complex does not impair fitness in an experimental model of lateral gene transfer. *Mol Biol Evol,* 27**,** 2441-5.

PAL, C., PAPP, B. & LERCHER, M. J. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet,* 37**,** 1372-5.

PAPP, B., PAL, C. & HURST, L. D. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature,* 424**,** 194-7.

PARK, C. & ZHANG, J. 2012. High expression hampers horizontal gene transfer. *Genome Biol Evol,* 4**,** 523-32.

POPA, O. & DAGAN, T. 2011. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol,* 14**,** 615-23.

QIAN, W., YANG, J. R., PEARSON, N. M., MACLEAN, C. & ZHANG, J. 2012. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet,* 8**,** e1002603.

RAVENHALL, M., SKUNCA, N., LASSALLE, F. & DESSIMOZ, C. 2015. Inferring horizontal gene transfer. *PLoS Comput Biol,* 11**,** e1004095.

RIVERA, M. C., JAIN, R., MOORE, J. E. & LAKE, J. A. 1998. Genomic evidence for two functionally distinct gene classes. *Genetics,* 95.

SANJUAN, R., MOYA, A. & ELENA, S. F. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A,* 101**,** 8396-401.

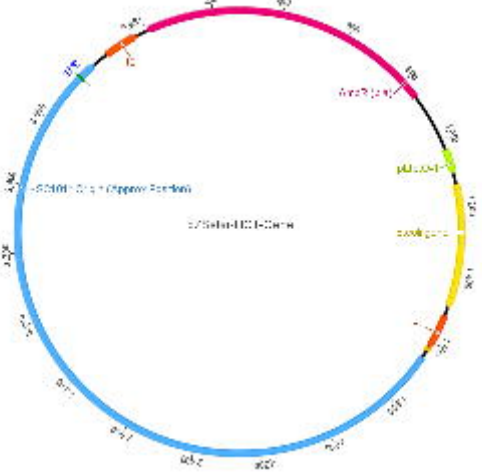
SOREK, R., ZHU, Y., CREEVEY, C. J., FRANCINO, M. P., BORK, P. & RUBIN, E. M. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science,* 318**,** 1449-52.

WELLNER, A. & GOPHNA, U. 2008. Neutrality of foreign complex subunits in an experimental model of lateral gene transfer. *Mol Biol Evol,* 25**,** 1835-40.

WELLNER, A., LURIE, M. N. & GOPHNA, U. 2007. Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol,* 8**,** R156.

WLOCH, D. M., SZAFRANIEC, K., BORTS, R. H. & KORONA, R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast Saccharomyces cerevisiae. *Genetics,* 159**,** 441-52.

XU, X. & HENSEL, M. 2010. Systematic analysis of the SsrAB virulon of Salmonella enterica. *Infect Immun,* 78**,** 49-58.

ZHONG, Q., KOBE, B. & KAPPLER, U. 2020. Molybdenum Enzymes and How They Support Virulence in Pathogenic Bacteria. *Frontiers in Microbiology*, 11.

ZUPOK, A., GORKA, M., SIEMIATKOWSKA, B., SKIRYCZ, A. & LEIMKUHLER, S. 2019. Iron-Dependent Regulation of Molybdenum Cofactor Biosynthesis Genes in Escherichia coli. *J Bacteriol*, 201.