BMC Microbiology

RESEARCH ARTICLE Open Access

Experimental determination of evolutionary barriers to horizontal gene transfer

Hande Acar Kirit[1,2], Mato Lagator[3] and Jonathan P. Bollback[1*]

Abstract

Background: Horizontal gene transfer, the acquisition of genes across species boundaries, is a major source of novel phenotypes that enables microbes to rapidly adapt to new environments. How the transferred gene alters the growth – fitness – of the new host affects the success of the horizontal gene transfer event and how rapidly the gene spreads in the population. Several selective barriers – factors that impact the fitness effect of the transferred gene – have been suggested to impede the likelihood of horizontal transmission, however experimental evidence is scarce. The objective of this study was to determine the fitness effects of orthologous genes transferred from *Salmonella enterica* serovar Typhimurium to *Escherichia coli* to identify the selective barriers using highly precise experimental measurements.

Results: We found that most gene transfers result in strong fitness costs. Previously identified evolutionary barriers — gene function and the number of protein-protein interactions — did not predict the fitness effects of transferred genes. In contrast, dosage sensitivity, gene length, and the intrinsic protein disorder significantly impact the likelihood of a successful horizontal transfer.

Conclusion: While computational approaches have been successful in describing long-term barriers to horizontal gene transfer, our experimental results identified previously underappreciated barriers that determine the fitness effects of newly transferred genes, and hence their short-term eco-evolutionary dynamics.

Keywords: Horizontal gene transfer, Evolutionary barriers, Gene length, Dosage sensitivity, Distribution of fitness effects

Background

Horizontal gene transfer (HGT) is the lateral transfer of genetic material between different individuals and species, and a major driver of evolution in all domains of microbial life [1–3]. HGT has contributed to a stunning array of phenotypic diversity that we observe in nature. As a source of phenotypic novelty, HGT differs from de novo mutations as new adaptive phenotypes can appear faster and sweep through the populations more rapidly, as demonstrated by the mannanase gene of the coffee berry borer beetle, *Hypothenemus hampei*, which was acquired from bacteria and enabled the beetle to digest the complex sugars in coffee beans, thus turning the beetle into an industrially relevant pest [4]. Similarly, the horizontal acquisition of vancomycin resistance by *Staphylococcus aureus* [5] and virulence factors by *Escherichia coli* O104:H4 in the 2011 European outbreak [6] are serious public health concerns. Despite the importance of HGT, we still have a limited understanding of the factors that determine the outcome of HGT events.

The success, or failure, of an HGT event is largely determined by the effect of the transferred gene on the fitness (differential growth compared to the wild-type not carrying a transferred gene) of the recipient species. Deleterious genes will be purged by selection, beneficial

* Correspondence: bollback@liverpool.ac.uk
[1]Institute of Integrative Biology, Functional and Comparative Genomics, University of Liverpool, Liverpool L69 7ZB, UK
Full list of author information is available at the end of the article

genes may be fixed, and the fate of neutral genes will be determined by stochastic forces, such as genetic drift [7, 8]. Thus, the fate of transferred genes depends on the distribution of their fitness effects (DFE). To date, we have little knowledge of the DFE for newly transferred genes, or of the factors that determine those fitness effects, especially following expression in the recipient cell.

Motivated by computational analyses, several non-mutually exclusive factors, or 'selective barriers', have been hypothesized to affect the DFE of horizontally transferred genes: the functional category of the transferred genes [9, 10], the number of protein-protein interactions (PPI) [10, 11], and the divergence between the donor and the recipient cell that is inferred as a difference in their GC content or the codon usage bias [12–14]. While computational approaches produced valuable insights into HGT, there are important barriers only assessable through experimental analyses. Notably, the suggestion that gene dosage, or dosage sensitivity, might play a role in HGT came from experimental insights and cannot be addressed with sequence data alone [15, 16].

Here we conduct a systematic experimental test of the barriers to HGT by transferring and expressing orthologs from *Salmonella enterica* serovar Typhimurium to *Escherichia coli*, and measuring their fitness effects. Specifically, we asked: What is the DFE of newly transferred genes? What selective barriers affect the likelihood of a newly transferred gene's spread in a population?

Results

We mimicked HGT by transferring genes from *S.* Typhimurium to an *E. coli* recipient to determine the DFE and test whether selective barriers — functional category, number of PPI, GC content, codon usage, and dosage sensitivity — affect the likelihood of transfer. We chose these species as they share the same environment – mammalian gut – which increases the potential of HGT between them. *S.* Typhimurium and *E. coli* are also close relatives [17]. HGT can result in a transfer of a new gene or an ortholog. By focusing on two closely related species, we ensured that all transferred *S.* Typhimurium genes had an existing ortholog in the *E.coli* genome. This also means that most of the native PPIs of the transferred orthologs are likely to be preserved. We transferred a total of 44 genes, placed them under the control of the same promoter, induced their expression, and measured their fitness effects with competition assays (Fig. 1). We specifically aimed to express the transferred genes in order to allow a more precise measurement of their effects on fitness. Furthermore, we expressed all the genes at the same expression level, to assess the intrinsic fitness costs associated with each gene, rather than the potential impact of their regulatory context. RNA-seq analysis showed that the transferred genes were expressed in the top 0.5% of all genes (see Methods section – RNA-seq: Data Processing). We also verified the expression of a subset of transferred genes at the level of translation (Sup. Figure 1). The recipient, *E. coli,* was chromosomally labeled at the p21 phage attachment site with two different fluorescent markers: CFP-labeled 'mutant' strain carried the plasmid containing one of the introduced genes (Sup. Figure 2); and YFP-labeled 'wild-type' strain carried the same plasmid only without the introduced gene. We confirmed that the chromosomal insertion of these two different fluorescent markers have similar effects on the fitness of the recipient (for details see Methods section – Accounting for the fitness effects of fluorescent markers). Two strains were mixed at equal frequencies and grown together with samples taken at regular intervals during the exponential phase ($t = 0$, 40, 80, and 120 min), and the change in frequencies of the two strains were measured by flow cytometry to determine fitness. Note that this means that the measure of fitness used in this study is relative to the wild-type. The fitness effects, or the selection coefficients (s), of transferred genes were estimated using the formula $ln\ (1 + s) = (ln\mathrm{R}_t - ln\mathrm{R}_0)/\mathrm{t}$, where R is the ratio of mutant to wild type, and t is the number of generations [18]. By performing 32 replicates for each transferred gene, we estimated the fitness effects of transferred genes with a previously not achievable degree of precision, $\Delta s \approx 0.005$ (Sup. Figure 3.). We opted to have such unprecedented precision in our estimates of Δs, at the cost of including more genes in the study, in order to obtain a more accurate distribution of fitness effects of horizontally transferred genes.

Distribution of fitness effects

Majority of *S.* Typhimurium genes transferred into *E. coli* had a negative effect on fitness with a median fitness effect $s = -0.020$ and a range of -0.606 to 0.009 (Fig. 2). Out of 44 transferred genes, 3 were beneficial, 5 were neutral (fitness not significantly different from zero), 25 were moderately deleterious, and 11 were highly deleterious ($s < -0.1$, Sup. Table 1). The shape of the DFE in our experiment was similar to DFEs observed in other biological systems, such as mutations in bacterial promoters [19, 20], viral sequences [21], transcription factors [22], and random transposon insertions [18]. The effect of the deleterious transferred genes is well described by a log-normal distribution ($\mu = -3.562$ and $\sigma = 1.693$), as previously observed for DMEs of mutations [23].

Evaluating potential barriers to HGT

The 44 genes chosen for experimental transfer differ with respect to several factors hypothesized to act as barriers to HGT. We asked whether these selective

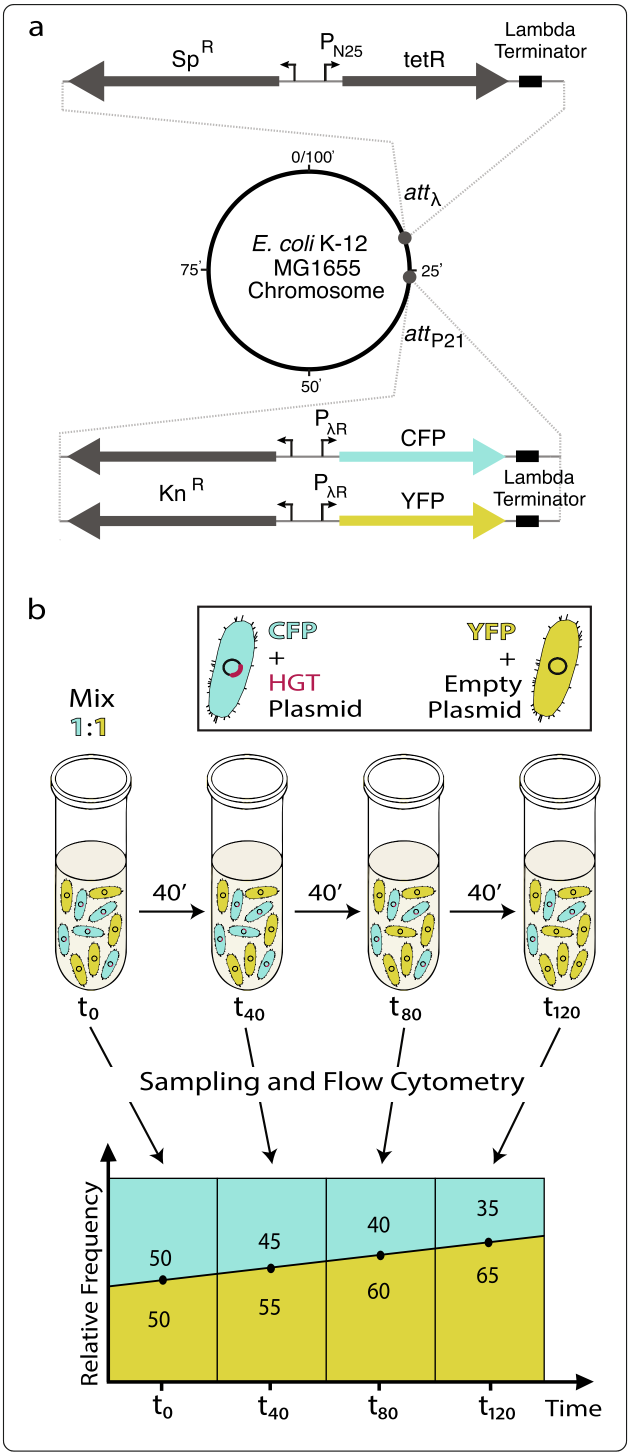

Fig. 1 Schematic representation of the experimental design. **a**. Chromosomal modifications in the recipient strain, *E. coli* MG1655 att-λ::(tetR-Sp^R) att-p21::(CFP/YFP-Kn^R), for the transferred genes. Att-λ and att-p21correspond to the attachment sites of the phage λ and p21, respectively. TetR is the repressor protein controlling the expression of the transferred genes. Sp^R and Kn^R are the resistance genes for spectinomycin and kanamycin, respectively. P$_{N25}$ and P$_{\lambda R}$ are the constitutive promoters. See Methods section for details. **b**. Depiction of the competition assay. Blue cells with CFP represent the 'mutant' strain that carries the pZS*-HGT plasmid containing the introduced gene, whereas yellow cells with YFP represent the 'wild type' strain that carries the empty pZS*-HGT plasmid without the introduced gene. The plot illustrates an example where the fitness effect of the gene is deleterious, resulting in a decrease in the frequency of blue cells over time. Numbers inside the segments represents the frequency of the type of the cell with same color

barriers predicted the fitness effects of the transferred genes. In particular, we looked at the effect of functional category, number of PPIs, gene length, and deviation in the GC content and codon usage between the transferred *S.* Typhimurium genes and their orthologs in *E. coli* (Sup. Table 2). We tested for these factors in a multiple linear regression model framework ($F_{5,37} = 2.24$, Sup. Table 3, for details see Methods section – Statistical analyses). Surprisingly, we found a significant effect only for the gene length.

Functional gene category

The 'functional category hypothesis' proposes that informational genes (those involved in DNA replication and repair, transcription, and translation) are less amenable to transfer than operational genes (those involved in processes like metabolism and biosynthesis) [9, 10, 24]. To test this hypothesis, we grouped genes by Gene Ontology and MultiFun annotations [25], classifying 24 genes as informational genes and 19 as operational (Sup. Table 2). No significant difference was observed between the two categories in either mean fitness effects (Fig. 3a, Wilcoxon rank sum test: median $\boldsymbol{s_{Info}} = -0.026$, median $\boldsymbol{s_{Oper}} = -0.010$, $W = 181$, $p = 0.130$; multiple regression model: $p = 0.354$) or in the variance in fitness effects (Levene's test: $\sigma^2_{Info} = 0.026$, $\sigma^2_{Oper} = 0.010$, $F_{1,41} = 1.164$, $p = 0.287$). Interestingly, 4 of the 5 most deleterious genes ($s < -0.25$, $\bar{s} = -0.422$) were informational. Moreover, dividing the genes further into functional categories using the COG database did not show any significant difference between these categories in their mean fitness effects ($F_{3,39} = 1.043$, $p = 0.384$, Additional Data File 1).

Protein-protein interactions

The 'complexity hypothesis' [11] proposes that newly acquired proteins may form spurious interactions with the proteins already present in the recipient cell because of the epistatic incompatibilities accrued during divergence.

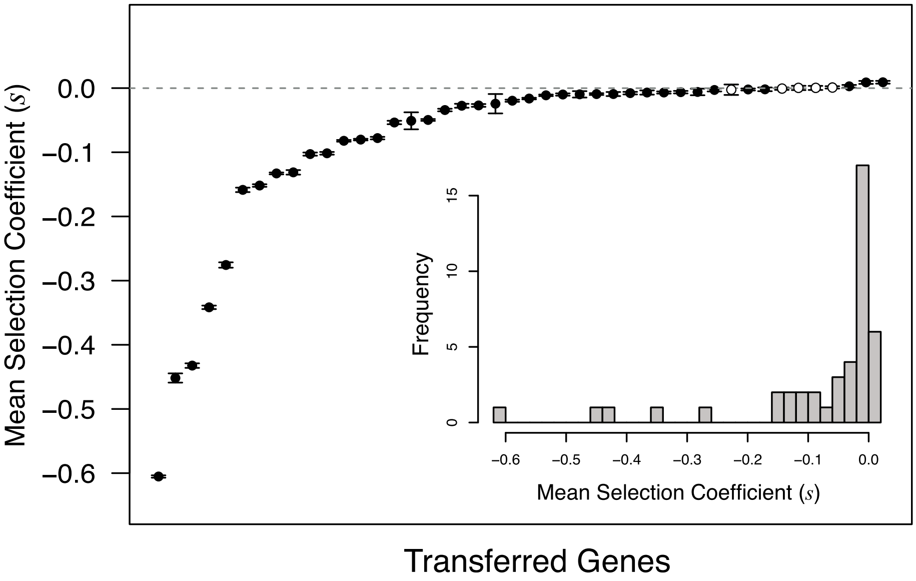

Fig. 2 DFE of 44 newly transferred *S. Typhimurium* orthologs expressed in *E. coli*. On the x-axis, genes are sorted by their fitness effects (*s*). Error bars indicate the 95% CI of the selection coefficients from 32 replicate measurements for each gene. Empty circles represent the 5 genes with effects not significantly different from zero. Embedded plot gives the histogram representation of the same data

In other words, an increase in the number of PPIs may decrease the likelihood of successful HGT. We asked whether the number of PPIs affected the fitness effects of transferred genes. We selected 44 genes that represent a range of PPI levels from 1 to 40, covering 83% of the range of *E. coli* genes given in the database of Hu et al. [26] (see Methods section — Selection of genes). We determined whether potential interacting partners are expressed in the recipient cell under our experimental conditions and adjusted the predicted PPI levels to exclude unexpressed genes (see Methods section — RNA-seq). In contrast to the predictions of the complexity hypothesis, we found that the number of PPI were uncorrelated with observed fitness effects (Fig. 3b, multiple regression model, $p = 0.245$).

GC content and codon usage

Differences in GC content [27] and codon usage [28] between a donor and a recipient have been proposed as barriers to horizontal acquisition of genes. Horizontally transferred genes with low GC content may avoid deleterious consequences through H-NS mediated gene silencing in gram-negative bacteria [27]. On the other hand, a fitness cost can arise because of non-optimal codon usage of transferred genes, leading to high rates of translation error, creating toxic side products in the recipient cell [13], or leading to ribosomal sequestration that slows down the overall growth of the recipient cell [12, 29–31].

To test the effects of GC content and codon usage, F_{OP} [32], we calculated the absolute deviations between our 44 *S.* Typhimurium genes and their *E. coli* orthologs. We examined the effects of the absolute deviation in GC content and codon usage bias separately, while also accounting for the potential correlation between them (simple linear regression, $F_{1,42} = 0.008$, $p = 0.93$). In terms of the absolute deviation in GC content, the 44 gene pairs cover a range of 0.1 to 5.2%, which spans the range of 94% of all the ortholog pairs between *S.* Typhimurium and *E. coli*. We found that the absolute deviation in GC content between the transferred *S.* Typhimurium genes and their *E. coli* orthologs did not correlate with the observed fitness effects (Fig. 3c, multiple regression model, $p = 0.325$). Considering higher average GC content of *S.* Typhimurium, we also looked at the effect of the actual GC content (rather than the deviation), which has been previously shown to affect the likelihood of HGT [33]. We did not observe a correlation between the GC content and fitness effects of genes (simple linear regression, $F_{1,42} = 0.429$, $p = 0.52$). With respect to absolute deviation in codon usage bias, the 44 gene pairs cover a range of 0 to 12%, spanning the range of 99% of all the ortholog pairs between the donor and recipient. The factor of absolute deviation in codon usage bias was not a significant predictor of the observed fitness effects either (Fig. 3d, multiple regression model: $p = 0.173$).

Gene length and intrinsic protein disorder

We identified a significant negative relationship between gene length and the fitness effects of transferred genes (Fig. 3e, multiple regression model: $p = 0.016$, simple linear regression: $R^2 = 0.15$, $p = 0.011$). The observed effect

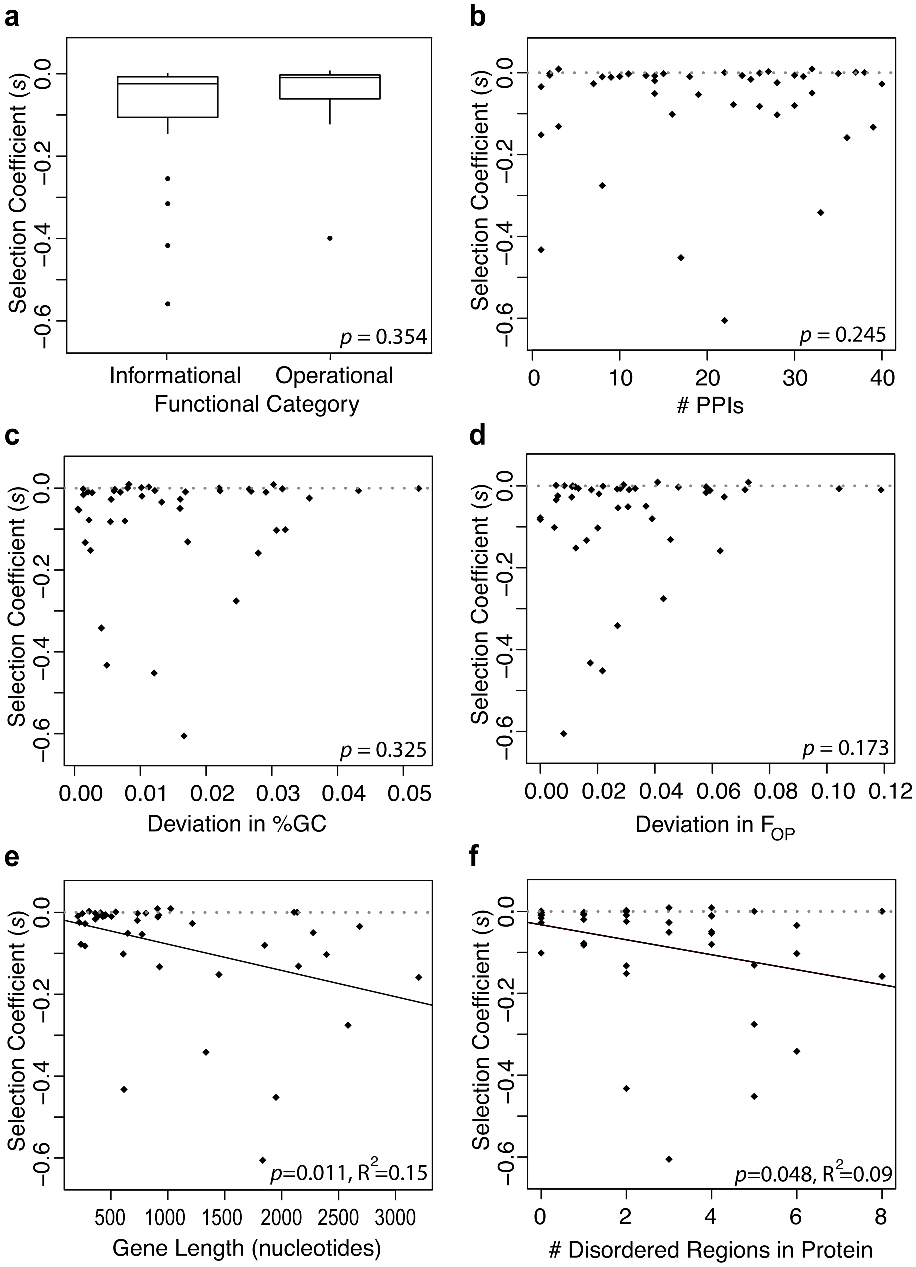

Fig. 3 Analyses of the selective barriers on HGT. The mean selective effects of 44 transferred S. Typhimurium genes expressed in *E. coli* background: **a.** divided into informational and operational genes based on functional categories; **b.** plotted against predicted number of PPIs; **c.** absolute deviation in %GC between orthologs; **d.** absolute deviation in codon usage between orthologs; **e.** gene length; **f.** number of disordered regions in the amino-acid sequences. All 5 factors given in **a.** to **e.** are used as explanatory variables in a multiple regression model (single *p*-values at the bottom right of each panel)**. e.** We repeated the simple linear regression for the gene length as it was the only factor with a significant effect in the multiple regression. Black lines in **e** and **f** show the best fit for the simple linear regression between the two variables; gray dashed lines show the zero line

is unlikely to arise from the costs of DNA replication and protein synthesis, as these costs are negligible for the small differences in length investigated here, with any effect likely to be below the experimental limit of detection [14, 34].

One explanation is intrinsically disordered protein regions — regions of proteins that lack a fixed tertiary structure and are inherently unstable or flexible — which are correlated with gene length (simple linear regression, $p < 0.001$, $R^2 = 0.67$, Sup. Figure 4). While little is understood about disordered regions of proteins, such regions potentially give rise to promiscuous molecular interactions, misfolding, and aggregation [35]. We identified the number of disordered regions within the protein sequences of the 44 transferred genes using Globplot (see Methods section for details, Sup. Table 2), [36], and found that the *S.* Typhimurium orthologs with more disordered regions in their protein sequences have significantly higher fitness costs (Fig. 3f, simple linear regression, $p = 0.048$, $R^2 = 0.090$).

Dosage sensitivity

Gene dosage effect, or the sensitivity to the change in the concentration of a gene product, is another potential barrier to HGT. Transfer events can result in additional copies of the gene within the recipient cell, potentially yielding a change in the protein concentration and lowering fitness by inducing an imbalance in the stoichiometry of the cell [15, 37, 38]. Sorek et al. [16] showed that, at least for some genes, an increase in dosage results in toxicity. Accordingly, we asked whether dosage sensitivity contributes to the observed fitness effects of transferred genes. Dosage sensitivity has classically been studied by modulating the number of gene copies, and hence the level of expression, then measuring the effects on the desired phenotype [39]. To test whether dosage sensitivity acts as a barrier to HGT, we measured the fitness effects of transferring additional copies of the native *E. coli* orthologs of the 44 genes into *E. coli* background using the same experimental setup (Sup. Figure 5). Using this design, any observed changes in fitness must arise from an imbalance in cellular protein levels, i.e., dosage sensitivity — rather than the effects of divergent function, number of interacting proteins, or codon usage — as the transferred genes were exact copies of existing genes.

We compared the fitness of each *E. coli* gene (grey bars, Fig. 4) with that of their orthologous *S.* Typhimurium gene that showed significant deleterious effects ($n = 36$, white bars, Fig. 4). If the *E. coli* copy was equal to or more deleterious than the *S.* Typhimurium copy, we attributed dosage sensitivity as the dominant factor determining the fitness effects of the transferred gene. Alternatively, if the *S.* Typhimurium copy is more deleterious, then factors other than dosage sensitivity drive the fitness effects of the gene. For simplicity, we refer to these groups as dosage sensitive and insensitive, respectively.

For 16 orthologous pairs, the *E. coli* copy was more deleterious than *S.* Typhimurium copy and 2 orthologous pairs had similar fitness effects. Thus 18 out of 36 *S.* Typhimurium genes were dosage sensitive. The remaining 18 genes, where the *S.* Typhimurium copy was significantly more deleterious than the *E. coli* copy, were dosage insensitive (Fig. 4).

We asked whether the previously identified barriers to HGT— functional category, number of PPIs, GC content, codon usage, gene length, and the number of disordered regions — determined the fitness effects of dosage sensitive and dosage insensitive genes independently. For dosage sensitive genes, only gene length (simple linear regression: $p = 0.002$, $R^2 = 0.456$) and the number of disordered regions (simple linear regression: $p = 0.006$, $R^2 = 0.391$) were significant predictors of fitness (Fig. 5).

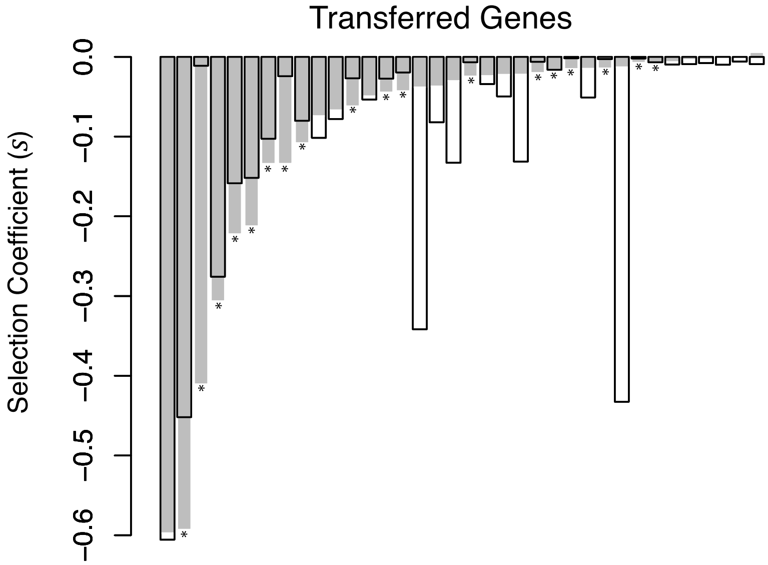

Fig. 4 Dosage sensitivity. Selection coefficients of newly transferred and deleterious genes from S. Typhimurium (white bars) expressed in *E. coli* background overlaid with those of their orthologs from *E. coli* (grey bars) expressed in *E. coli* background. On the x-axis genes are sorted according to their selection coefficients for the transfer of *E. coli* orthologs. Genes marked with asterisks show dosage sensitivity, as the fitness cost of the additional copies of an *E. coli* gene is the same or greater than the fitness cost of its *S.* Typhimurium ortholog

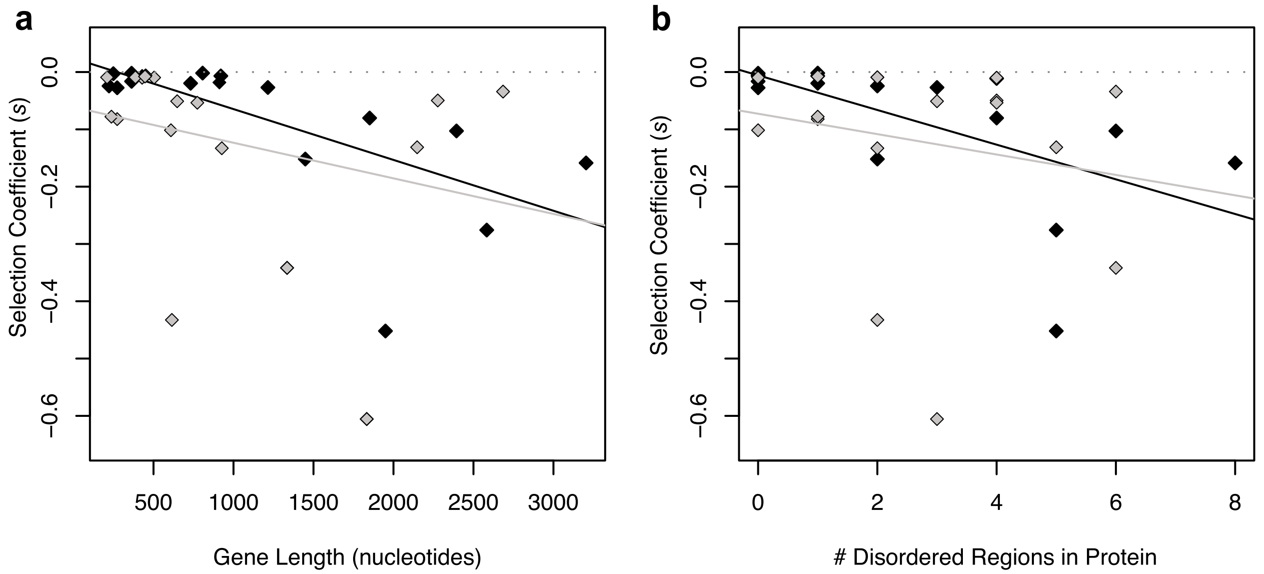

Fig. 5 Selection coefficients of 36 newly transferred and deleterious *S.* Typhimurium genes. Genes are plotted against **a**. gene length in nucleotides; **b**. number of disordered regions in their amino-acid sequence. Black data points represent the dosage sensitive genes, and grey data points with black frames represent the dosage insensitive genes. Lines are the simple linear regressions between the two variables with corresponding colors, gray dashed lines show the zero line

Surprisingly, these barriers did not predict the fitness effects of dosage insensitive genes (Fig. 5), even though the dosage sensitive and insensitive genes did not differ in terms of their length (Wilcoxon test: $W = 186$, $p = 0.451$, Sup. Figure 6a), the number of disordered regions in their amino-acid sequences (Wilcoxon test: $W = 145$, $p = 0.602$, Sup. Figure 6b), or their level of divergence (Wilcoxon test: $W = 138.5$, $p = 0.467$, Sup. Figure 6c).

We explored if the dosage effect was more likely to appear with greater fold-increase in expression resulting from the second *E. coli* copy of the gene. We used RNA-seq data to estimate the intrinsic expression level of each gene before the second copy is added. Interestingly, we observed high fitness costs ($s < -0.1$) only among those genes in which there was greater than a 10-fold increase in expression (Sup. Figure 7).

Discussion

Our understanding of HGT has relied primarily on comparative computational analyses of genomes, which can only study successful transfer events that have gone through the sieve of natural selection. Experimental approaches, which have been rare due to their labour-intensive nature and technological limitations, offer the potential to test the role of a different set of factors that could promote or hinder HGT [16, 40]. We relied on a very precise experimental approach for measuring fitness to better understand the role of different selective factors in HGT, which has allowed us to formulate and test new hypotheses about previously unconsidered barriers to HGT. We performed competition experiments during the exponential phase of growth where the difference in growth rates, which originated from the expression of the horizontally transferred genes, were most pronounced.

We explored the relative importance of several previously hypothesized factors that may affect the probability of an HGT event, by transferring orthologous genes from *S.* Typhimurium into *E. coli*. We wanted to understand the fitness effects of the transferred genes when they are liberated from their regulatory context, meaning they are not under the control of their endogenous promoters anymore. This allowed us to test, within the constraints of our experimental design, the intrinsic fitness costs of each transferred gene. We did not test for the role of expression level or regulatory context, which can alter the fitness effects of transferred genes. It is also possible that the fitness effects of transferred genes might differ if transferred together with their functional partners.

Unlike the mostly neutral effects of transferred DNA fragments of random size, for which the expression state was not known [40], we find that the majority of transferred genes that are expressed in the recipient cell impose a significant fitness cost. Consistent with Hao and Golding [41], our findings suggest that, if expressed, a large fraction of transferred genes would be rapidly eliminated by selection – a likely scenario given the ease of evolving constitutive promoters in bacteria [42]. Yet, despite most transfer events resulting in fitness costs, HGT is still one of the major sources of novel genetic material in microbes [8, 43] . The disparity between this observation and the DFE we report could be reconciled in a couple of ways. Firstly, if the bacterial effective population sizes in nature are smaller than believed, due to spatial structure, recurrent bottlenecks, or recurrent selective sweeps, genetic drift will dominate over selection [7, 44]. Secondly, the ubiquitous mechanisms for gene silencing may also explain this discrepancy as silencing shields the recipient cell from deleterious effects [45]. Thus, deleterious genes may persist long enough to be rescued by beneficial compensatory mutations or a change in the selective environment. Such silencing mechanisms may give enough time for the evolution of more complex regulation of the transferred gene, enabling its conditional expression.

Interestingly, we did not find evidence for many of the previously hypothesized barriers to HGT: functional gene categories, the number of PPIs, GC content, or optimal codon usage [14]. Note that in this study, we opted to measure the fitness effects of a modest number of genes with very high precision. Further studies that extend the number of genes might provide evidence for some of these barriers. In addition, with increased divergence between donor and recipient species, these factors may become increasingly important, and potential new factors might play a previously unappreciated role. In our dataset, the relatively modest differences in codon usage and GC content between *S.* Typhimurium and *E. coli* orthologs may have prevented us from detecting their role as selective barriers to HGT. Nevertheless, among closely related species these intrinsic properties do not appear to be strong selective barriers.

In contrast, computational studies found limited support for dosage sensitivity as a major barrier to HGT, in part because it is difficult to infer dosage from genomic data [37]. Several genes have been identified previously as either dosage sensitive or insensitive through experimental altering of their native expression levels [15, 46]. For instance, Sorek et al. identified dosage sensitivity as a barrier to HGT along with toxicity for a set of nearly lethal but very divergent genes from a variety of bacterial species [16]. Our findings provide further evidence that dosage sensitivity is a general barrier to HGT by focusing on transfer of non-lethal genes between two closely related species.

Divergence between the newly transferred gene and its native counterpart could result in both, dosage sensitive (foreign ortholog is divergent enough that it does not interfere with the original function of the gene anymore) and insensitive (mutations cause a dominant negative effect whereby interfering with the original function of the gene) effects. We did not see a significant difference between the dosage sensitive and insensitive genes in terms of divergence. The dosage sensitivity, however, can also be related to the intrinsic disorder of proteins, which predicts the fitness effects of dosage sensitive, but not of the dosage insensitive genes in our data set. Here we show that gene dosage and regions of protein disorder are significant predictors of the success of HGT, and may arise due to pathological molecular interactions, misfolded toxic configurations, or deleterious aggregates.

Taken together with the disparities in previous reports on the topic, our results show that there might be a broad variety of factors that limit or promote HGT in a context specific manner, especially during the early stages following a transfer event. The immediate (short-term) effect of a transferred gene on host fitness plays a critical role in the long-term success of a horizontal transfer event, but is not the only factor determining the long-term fate of a transferred gene. Our findings suggest that the short-term evolutionary dynamics of HGT and the barriers that determine the early success of a transferred gene might differ from their long-term evolutionary dynamics, emphasizing the need for further experiments to elucidate these differences.

Methods

Chromosomal modifications in the recipient strain

To differentiate two cell types, we inserted the fluorescent markers *cfp* and *yfp* under control of the constitutive $P_{\lambda R}$ into the phage p21 attachment site on the *E. coli* MG1655 (DSM18039) chromosome using pAH95 from CRIM system and protocol given previously [47]. Briefly, *E. coli* cells containing pAH121 were electroporated with the pAH95, suspended in SOC without antibiotics, incubated at 37 °C for 1 h and at 42 °C for 30 min to get rid of the pAH121, spread onto LB agar plates with kanamycin 10 μg/mL and incubated over night at 37 °C. Colonies were streak-purified once nonselectively and then tested for antibiotic resistance for stable integration and loss of the pAH121 and by PCR for single integration of the cassette. The insertion is sequence-verified. The *cfp* gene was derived from an *E. coli* strain used in Elowitz et al. [48], and *yfp* gene is the *Venus* gene derived from the pZS123 used in Cox et al. [49] .

We integrated the repressor protein gene *tetR* that suppresses the expression of transferred genes into the phage λ attachment site on the *E. coli* MG1655 att-p21::(CFP/YFP-Kn^R) chromosome using the plasmids and protocol given previously [50]. pZS4Int plasmid contained *tetR* gene under P_{N25} promoter, spectinomycin resistance gene, and origin of replication pSC101. Briefly, the origin of replication was cut out of the plasmid pZS4Int and the rest of the plasmid was ligated back. *E. coli* cells, containing pLDR8 [51] were then electroporated with this ligated DNA. Cells were incubated first at 42 °C for 2 h to get rid of pLDR8 and then at 37 °C overnight on agar plates supplemented with spectinomycin

50 μg/mL to select resistant clones. Colonies were streak-purified once non-selectively and then tested for antibiotic resistance for stable integration and loss of the pLDR8 and by PCR for single integration of the *tetR* cassette. The insertion is sequence-verified.

Selection of genes

S. Typhimurium and *E. coli* are genetically similar and share ecological environments [52, 53] As we are interested in the role of protein-protein interactions (PPIs) and functional categories, this similarity ensures that the majority of gene products transferred from *S.* Typhimurium are functional and most of their functional partners exist in the *E. coli* genetic background. Although the exact number of interacting partners may differ in *E. coli*, this difference is expected to be minimal. In addition, the two species are sufficiently divergent to allow us to systematically test the effects of several factors of the introduced genes.

More specifically, *Salmonella enterica* serovar Typhimurium LT2 (DSM18522, Genbank AE006468.1) [17] was used as the gene donor. We excluded genes that are known to be mobile genes, such as phage related proteins, transposable elements or insertion sequences, as well as ribosomal and transfer RNAs that are known to be related to the mobile genes. We applied an arbitrary selection method for the 45 genes as follows. We first obtained the information of protein-protein interactions and functional modules for *E. coli* genes from Hu et al. [26]. In that study, genes were grouped into 507 functional modules, largest containing 297 genes and smallest with 2 genes, through a clustering algorithm based on their interaction partners (Additional Data File 1). We only selected genes with interactions validated in that study using LCMS and MALDI after SPA tagging of proteins. Since a random selection of genes with this setting would be biased towards large functional modules, and we expected specific functions of the genes to have an effect on fitness, we ensured that the sampled genes were from different functional modules. In addition, from the estimated number of interaction partners as reported in that same study [26], we ensured that the sampling included the widest possible range of PPIs — we sampled uniformly from a range of 1 to 40 physical PPI. We kept our sampling independent from the information related to the origins of the genes and ended up having 5 genes that are predicted to have been horizontally transferred, 36 as part of the core genome and 3 without known origins (ecogene.org, [54], Sup. Table 2, Additional Data File 1).

Cloning of selected genes

Selected genes were introduced into the recipient *E. coli* cells by transformation of a modified version of the pZS* class of plasmids [50] Sup. Figure 2). This plasmid is maintained in the recipient cell at 3–4 copies allowing us to keep the expression of the introduced gene at moderate levels. The coding regions of the selected *S.* Typhimurium genes (or their endogenous *E. coli* orthologs) were cloned into the pZS*-HGT plasmids under the control of the inducible promoter $P_{LtetO-1}$ [50]. Each gene was cloned at the *Avr*II site at 5′-end to ensure that start codon was located at the exact position relative to the promoter and ribosomal binding site. We failed to clone one gene (STM4381, *ulaR*, transcriptional repressor for the L-ascorbate utilization divergent operon, 756 bps) out of 45 selected genes, therefore, rest of the analyses were performed on the 44 successfully cloned genes.

The plasmids were then transferred into *E. coli* MG1655 att-λ::(tetR-Sp^R) att-p21::(CFP/YFP-Kn^R) cells and successful transformants were selected on LB agar with ampicillin 50 μg/mL. After two rounds of streak purification on ‘rich M9 medium’ (1x M9 salts (Sigma-Aldrich, M6030), 1% CAA (Sigma-Aldrich, A2427), 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) agar plates with ampicillin 50 μg/mL, single colonies were grown overnight in liquid rich M9 medium with ampicillin 50 μg/mL and stored at − 80 °C. All the cloned genes are sequence-verified.

Same modified version of the pZS* backbone was used for the experiments of 44 *E. coli* orthologs in the *E. coli* background (DNA synthesized and cloned into our pZS* backbone by Epoch Life Science Inc., Sugar Land, Texas, USA).

Competition assays

We performed competition assays using *E. coli* MG1655 att-λ::(tetR-Sp^R) att-p21::(CFP/YFP-Kn^R) strains, CFP strain carrying the plasmid pZS*-HGT with the transferred gene (referred to as the ‘mutant’ in this study) while the YFP strain carrying the same plasmid without an insert (referred to as the ‘wild type’ in this study). In total 32 replicate competitions were performed across 4 different days for each gene.

All competition assays were done in ‘rich M9 medium’ with ampicillin 50 μg/mL. We determined the inducer concentration as 5 ng/mL anhydrotetracycline (ATC, Sigma-Aldrich, 37,919) with a titration assay. For each competition assay, first day, frozen stocks were streaked on rich M9 agar plates. Second day, a colony was picked and grown in rich M9 medium for 16 h. Third day, overnight cultures were diluted 1000x and grown initially for 60 min, followed by the addition of 5 ng/mL ATC to initiate the induction of inserted genes, and then grown for another 60 min, until OD ≃ 0.12. Then the two cell types (wild type and mutant) were mixed at equal ratios and competed with each other for 120 min (~ 3 generations) in 96-well plates. An initial sample (t_0) was taken at the beginning of the competition and three more samples

(t_1, t_2, t_3) were taken every 40 mins ~ each generation. Mutant to wild type ratio was determined by counting 50, 000 cells at each sampling with BD FACSCanto II flow cytometer. Using these four ratios, the fitness costs of genes (s) were estimated using the regression model $\ln(1+s) = (\ln R_t - \ln R_0)/t$, where R is the ratio of mutant to wild type and t is the number of generations [18].

By conducting competition assays during the exponential phase of growth and using time-series data, we were able to detect very small differences in selection coefficients, avoiding random fluctuations that occur during the lag and stationary phases. We performed a post-hoc power analysis to test the difference between two independent 30 replicates of selection coefficients using a two-tailed test, power of 0.80, alpha of 0.01 and obtained $\Delta s \approx 0.005$ as the potential effect size of our measurements (Sup. Figure 3). The standard deviation of this test is calculated as the average standard deviation observed during the selection coefficient calculations of all genes (including the within plate control competitions) performed in this study.

Accounting for the fitness effects of fluorescent markers

As we wished to control for any fitness differences that might be the result of introducing two different fluorescent markers, we compared the fitness of the two 'wild type' strains, i.e., strains carrying *cfp* vs *yfp* (both with empty pZS*-HGT plasmids) using the competition assay protocol described in the Methods section. We detected a small but significant difference between the fitness of CFP strain and YFP strain ($s_{\mathrm{CFP} > \mathrm{YFP}} = 0.004$, SD = 0.010, $t_{(314)} = 7.118$, $p < 0.001$). Therefore, we accounted for this difference in the estimation of selection coefficients of introduced genes during the competition assays. We did that by running a set of competitions between these two 'wild type' cells during every competition assay in parallel as a control. Since we did the competition assays in the deterministic phase of the growth under pure haploid selection, this difference in the fitness costs of different fluorescent markers is a constant that we subtracted from the estimated selection coefficient of transferred gene. Such that, each estimation of selection coefficient was corrected for with the fitness difference of the two 'wild types' in the control wells of corresponding experiments.

To determine whether the introduced genes might show different fitness effects on the different fluorescent backgrounds (CFP strain and YFP strain) we did a reciprocal introduction by cloning a subset of 8 randomly selected genes out of our 44 *S.* Typhimurium genes into the YFP strain and repeated the competition assays. The regression between the selection coefficients of genes (mean of 32 replicates) in CFP strain and YFP strain was highly significant ($F_{1,6} = 117$, $p < 0.001$, $R^2 = 0.943$, slope = 1.007), indicating different fluorescent backgrounds do not interact with this subset of genes, and all measured effects are solely due to the introduced genes.

RNA-seq: sample preparation

To estimate the expression level of transferred genes in our experiment, we cloned the *mCherry* gene into an empty pZS*-HGT plasmid under the hybrid promoter P_{LtetO}-1 as an expression control and used RNA-seq to measure the expression level of *mCherry* gene. Cultures grown overnight in rich M9 medium were diluted 1000x and handled under the same conditions as the competition assays described above. When the OD of the cultures reached to ~ 0.12, growth was stopped by adding Qiagen RNA protect Bacteria Reagent (cat no. 76506) to 20 mL cultures (~ 6×10^8 cells). Total RNA was purified with Qiagen Rneasy Mini Kit (cat no. 74104). Quality and integrity of the total RNA samples were checked in Agilent 2100 Bioanalyzer and Agilent RNA 6000 Nano Kit (reorder number 5067–1511). The experiment was performed as 3 biological replicates. Library preparation (RiboZero, NEB), further quality checks and next-generation sequencing (HiSeq2500-v4, SR100 mode) were performed at the VBCF NGS Unit (www.vbcf.ac.at).

RNA-seq: data processing

Sequence reads with an average read quality of > = 34 were retained for further analysis. After quality controls, fastq files were mapped against the *E. coli* MG1655 genome (Genbank U00096.3) using the Bowtie2 aligner using default settings in RSEM [55]. The reference genome was modified in silico to contain the chromosomal modifications of *tetR* and fluorescent protein gene cassettes. Expected counts were calculated by using the defaults in RSEM [56]. After between-sample normalization of the counts with the DESeq package of the R statistical software [57], TPM (transcript per million) values for each gene were calculated and used in further analyses [56] (Additional Data File 1).

RNA-seq data have also served to inspect whether all of the listed interaction partners of the 44 selected genes were expressed under our experimental conditions. After obtaining the expression levels for the whole transcriptome, to determine whether a gene is expressed or not, we needed a threshold level of expression below which a gene would have been eliminated from further analyses. To this end, we inspected the expression levels of genes that are known as being repressed under our experimental conditions, i.e., lactose operon and arabinose regulon genes. Expression level of these genes ranged from 0.5–10 TPM in our RNA-seq data. Based on this observation and previous similar studies, we set a threshold value of TPM ≥ 10 when counting a PPI partner as expressed [58].

To confirm the reproducibility of our RNA-seq based transcriptomic data, we have performed 3 biological replicates. Correlation coefficients calculated for the expression values of the genes above the TPM_10 threshold between replicates were ≥ 0.99 for all combinations.

With this RNA-seq analysis we also identified the expression levels of the transferred genes by inserting mCherry gene under the control of the pLtet0–1 promoter and found that the transferred genes were highly expressed – within the top 0.5% of expressed genes (~ 3300 TPM).

Testing the protein expression with fusion-GFP

To confirm the translational expression of transferred genes in our experiments, we prepared GFP fusion protein cloned at the 3′-end of the genes. We randomly chose a subset of genes, and on each of the corresponding plasmids (pZS*-HGT) we first removed the stop codons of the transferred genes, added a GGSGGS linker and the *GFP* gene without its start codon. In this setting, level of GFP expression indicates the expression level of the transferred genes. Cultures grown overnight in rich M9 medium were diluted 1000x and handled under the same conditions as the competition assays described above. One h after adding the inducer (5 ng/mL ATC), OD_{600} and GFP_{540} emissions were measured using 96 well plate reader, and GFP_{540} readings were normalized with the corresponding OD_{600} readings (Sup. Figure 1).

Statistical analysis

To determine whether the fitness effects of genes were significantly different from zero or not, one-sample t-tests were done for the 32 replicates of each gene, with one-tailed ($\mu_0 > 0$ for beneficial, $\mu_0 < 0$ for deleterious genes) or two-tailed ($\mu_0 = 0$ for neutral genes) settings. $\alpha = 0.05$ was used as the significance level after Benjamini and Hochberg false discovery rate (BH-FDR) corrections for multiple testing [59] (Fig. 1). The data for all 32 replicates of 44 orthologous genes of *S.* Typhimurium and *E. coli* are given in Additional Data File 2.

After dividing the 44 orthologous genes from *S.* Typhimurium to *E. coli* into two groups according to their functional categories, one-sided Wilcoxon rank sum test (Mann-Whitney U test) was used to decide if the fitness effects of the informational genes were less than that of the operational genes. Similarly, Levene's test was used to decide if the variance of the distribution of fitness effects of the informational genes was less than that of the operational genes. Analyses were done on the mean selection coefficients of genes for the 32 replicate measurements (Additional Data File 1, Fig. 3a). Note that, in the experiments in which we transferred the *E. coli* orthologs into the *E. coli* background, there is a significant difference in the means and variances of fitness effects between informational and operational genes (Wilcoxon rank sum test: median s_{Info} = – 0.039, median s_{Oper} = – 0.015, W = 158, p = 0.045, and Levene's test: σ^2_{Info} =0.033, σ^2_{Oper} =0.001, $F_{1,41}$ = 5.335, p = 0.026, respectively).

We investigated a number of intrinsic genetic properties —Protein-protein interaction level (explained under 'selection of genes' section), GC content, codon usage, gene length, and disordered amino acid regions. GC content was calculated as the absolute deviation between the introduced *S.* Typhimurium gene and the *E. coli* ortholog. Codon usage was calculated as the absolute deviation of the frequency of optimal (F_{OP}) usage in the introduced *S.* Typhimurium sequence using the *E. coli* F_{OP}. Gene length was quantified as the number of base pairs from the start to stop codon of the *S.* Typhimurium gene (i.e., cds).

We used the web service GlobPlot (http://globplot.embl.de) to identify the number of disordered regions within the protein sequences of the transferred genes from both donors [36]. GlobPlot is a CGI (common gateway interface) for exploring disorder and globular segments. The user can paste a sequence and then the GlobPlot server displays any obtained domain predictions as colored boxes layered on a graph. Residue ranges for found disordered segments and globular regions are shown at the bottom of the output page.

To investigate the effect of these intrinsic factors we employed multiple linear regression. After investigating interactions and more complicated models, we used the following model: '*S.* Typhimurium selection coefficients' ~ 'Functional Category (as dummy variable)' + 'Protein-protein interaction level' + 'Deviation in GC% between orthologs' + 'Deviation in codon usage between orthologs' + 'Gene length in nucleotides'. The analysis was done on the mean selection coefficients of genes for the 32 replicate measurements (Fig. 3a, b, c, d, Sup. Table 3). Extended tables with all the relevant information are attached in Sup. Table 1 and 2, and Additional Data File 1. As gene length was the only factor with a significant effect in this multiple regression analysis, a simple regression between gene length and *S.* Typhimurium selection coefficients was performed separately (Fig. 3e). Additional simple linear regression was performed between '*S.* Typhimurium selection coefficients' and 'the number of disordered regions in their amino-acid sequence' (Fig. 3f).

After dividing the 44 orthologous genes from *S.* Typhimurium to *E. coli* into two groups — 18 genes where dosage is the dominant factor and the remaining 18 genes where it is not, simply by comparing 32 replicates of *S.* Typhimurium selection coefficients to *E. coli* selection coefficients for each ortholog pairs using one-sided t-tests — we performed simple linear regression analyses for the rest of the intrinsic genetic properties (protein-protein interaction level, GC content, codon usage, gene length, and disordered regions in amino-acid sequence) on these two groups separately (Fig. 5).

Additional two-sided Wilcoxon rank sum tests (Mann-Whitney U test) were performed to test if these two categories of genes are different in their mean gene length, mean number of disordered regions in amino-acid sequence, and level of divergence (Sup. Figure 6). A final simple linear regression was performed to see the relationship between gene length and the number of disordered regions in amino-acid sequence of *S.* Typhimurium genes (Sup. Figure 4).

All the statistical analyses were done using the R software package (version 3.1.1).

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12866-020-01983-5.

Additional file 1. An extended table with all the relevant information on the transferred genes.

Additional file 2. The selection coefficients data calculated from flow cytometry outputs for all 32 replicates of 44 orthologous genes of S. Typhimurium and *E. coli*.

Additional file 3. Supplementary figures and tables that accompany the paper.

Abbreviations

HGT: Horizontal gene transfer; DFE: Distribution of fitness effects; PPI: Protein-protein interactions; CFP: Cyan fluorescent protein; YFP: Yellow fluorescent protein; *s*: Selection coefficient

Acknowledgements

We thank Katharina M. Pöcher for her guidance during the RNA-seq analysis; John F. Baines, Andrea J. Betancourt, Călin C. Guet and Tugba Keskin for valuable discussions and contributions; Rama P. Bhatia and Jay C. D. Hinton for helpful feedback on the manuscript.

Authors' contributions

HAK and JPB conceived the study and designed the experiments together. HAK carried out the experiments. HAK and ML analyzed the data. HAK performed RNA-seq analysis. HAK wrote the initial draft of the manuscript and revised it together with ML and JPB. All authors have read and approved the last version of the manuscript.

Funding

This work was supported by the H2020 European Research Council EVOLHGT No. 648440) to J.P.B. Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 216779/Z/19/Z) to M.L.

Availability of data and materials

Authors confirm that all relevant data are included in the article and/or its supplementary information files. Furthermore, the RNA-seq data is available in the GEO repository with GEO accession number GSE148719, [www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148719].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Institute of Integrative Biology, Functional and Comparative Genomics, University of Liverpool, Liverpool L69 7ZB, UK. [2]Present Address: Stephenson School of Biomedical Engineering, University of Oklahoma, Norman 73019, USA. [3]Faculty of Biology, Medicine and Health, School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK.

Received: 21 April 2020 Accepted: 21 September 2020
Published online: 28 October 2020

References

1. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405(6784):299–304 Nature Publishing Group.
2. Popa O, Dagan T. Trends and barriers to lateral gene transfer in prokaryotes. Curr Opin Microbiol. 2011;14(5):615–23.
3. Polz MF, Alm EJ, Hanage WP. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet. 2013;29(3):170–5 Elsevier Ltd.
4. Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, et al. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proc Natl Acad Sci U S A. 2012;109(11):4197–202.
5. Weigel LM, Clewell DB, Gill SR, Clark NC, LK MD, Flannagan SE, et al. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. Science. 2003;302(5650):1569–71 American Association for the Advancement of Science.
6. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany. N Engl J Med. 2011;365(8):709–17 Massachusetts Medical Society.
7. Gillespie JH. Genetic drift in an infinite population: the Pseudohitchhiking model. Genetics. 2000;155(2):909–19.
8. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet. 2015;16(8):472–82.
9. Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes. Proc Natl Acad Sci. 1998;95(11):6239–44 National Academy of Sciences.
10. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: The complexity hypothesis. Proc Natl Acad Sci. 1999;96(7):3801–6 National Academy of Sciences.
11. Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol. 2011;28(4):1481–9.
12. Tuller T, Girshovich Y, Sella Y, Kreimer A, Freilich S, Kupiec M, et al. Association between translation efficiency and horizontal gene transfer within microbial communities. Nucleic Acids Res. 2011;39(11):4743–55 Oxford University Press.
13. Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet, Nat Publ Group. 2009;10(10):715–24.
14. Baltrus DA. Exploring the costs of horizontal gene transfer. Trends Ecol Evol. 2013;28(8):489–95.
15. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003;424(6945):194–7.
16. Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide experimental determination of barriers to horizontal gene transfer. Science. 2007;318(5855):1449–52 American Association for the Advancement of Science.
17. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature. 2001;413(6858):852–6 Nature Publishing Group.
18. Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. Genetica. 1998;102–103(1–6):349–58.
19. Kinney JB, Murugan A, Callan CG, Cox EC. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proc Natl Acad Sci U S A. 2010;107(20):9158–63.
20. Lagator M, Sarikas S, Acar H, Bollback JP, Guet CC. Regulatory network structure determines patterns of intermolecular epistasis. Elife. 2017;13:6.
21. Sanjuán R, Moya A, Elena SF. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc Natl Acad Sci. 2004;101(22):8396–401 National Acad Sciences.
22. Shultzaberger RK, Maerkl SJ, Kirsch JF, Eisen MB. Probing the Informational and Regulatory Plasticity of a Transcription Factor DNA–Binding Domain. PLoS Genet. 2012;8(3):e1002614–3 Madhani HD, editor. Public Libr Sci.
23. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8(8):610–8.
24. Nakamura Y, Itoh T, Matsuda H, Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat Genet. 2004;36(7):760–6.

25. Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, et al. Multidimensional annotation of the Escherichia coli K-12 genome. Nucleic Acids Res. 2007;35(22):7577–90.
26. Hu P, Janga SC, Babu M, Díaz-Mejía JJ, Butland G, Yang W, et al. Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. PLoS Biol. 2009;7(4):e96.
27. Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JCD. H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog. 2006;2(8):e81–7 Public Library of Science.
28. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, et al. Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science. 2006;313(5784):236–8.
29. Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. Mol Syst Biol. 2011;7:1–13 Nature Publishing Group.
30. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-limiting steps in yeast protein translation. Cell. 2013;153(7):1589–601 Elsevier Inc.
31. Roller BRK, Stoddard SF, Schmidt TM. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. Nat Microbiol. 2016;1:16160 Nat Publ Group.
32. Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol. 1981;151(3):389–409 Academic Press.
33. Quandt EM, Traverse CC, Ochman H. Local genic base composition impacts protein production and cellular fitness. PeerJ. 2018;6:e4286 PeerJ Inc.
34. Lynch M, Marinov GK. The bioenergetic costs of a gene. Proc Natl Acad Sci U S A. 2015;112(51):15690–5.
35. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell. 2009;138(1):198–208.
36. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003;31(13):3701–8.
37. Park C, Zhang J. High expression hampers horizontal gene transfer. Genome Biol Evol. 2012;4(4):523–32.
38. Bershtein S, Serohijos AWR, Bhattacharyya S, Manhart M, Choi J-M, Mu W, et al. Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria. PLoS Genet. 2015;11(10): e1005612 Achtman M, editor. Public Library of Science.
39. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci U S A. 2012; 109(37):14746–53 National Academy of Sciences.
40. Knöppel A, Lind PA, Lustig U, Näsvall J, Andersson DI. Minor fitness costs in an experimental model of horizontal gene transfer in bacteria. Mol Biol Evol. 2014;31(5):1220–7.
41. Hao W. The fate of laterally transferred genes: life in the fast lane to adaptation or death. Genome Res. 2006;16(5):636–43.
42. Yona AH, Alm EJ, Gore J. Random sequences rapidly evolve into de novo promoters. Nat Commun. 2018;9(1):1530 Nature Publishing Group.
43. Boto L. Horizontal gene transfer in evolution: facts and challenges. Proc R Soc B Biol Sci. 2010;277(1683):819–27.
44. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217(5129): 624–6 Nature Publishing Group.
45. Navarre WW. The impact of gene silencing on horizontal gene transfer and bacterial evolution. Adv Microb Physiol. 2016;69:157–86 Elsevier.
46. Omer S, Kovacs A, Mazor Y, Gophna U. Integration of a foreign gene into a native complex does not impair fitness in an experimental model of lateral gene transfer. Mol Biol Evol. 2010;27(11):2441–5.
47. Haldimann A, Wanner BL. Conditional-replication, integration, excision, and retrieval plasmid-host Systems for Gene Structure-Function Studies of Bacteria. J Bacteriol. 2001;183(21):6384–93.
48. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297(5584):1183–6.
49. Cox RS, Dunlop MJ, Elowitz MB. A synthetic three-color scaffold for monitoring genetic regulation and noise. J Biol Eng. 2010;4:10.
50. Lutz R, Bujard H. 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 Res. 1997;25(6):1203–10.
51. Diederich L, Rasmussen LJ, Messer W. New cloning vectors for integration in the lambda attachment site attB of the Escherichia coli chromosome. Plasmid. 1992;28(1):14–24.
52. Winfield MD, Groisman EA. Role of nonhost environments in the lifestyles of Salmonella and Escherichia coli. Appl Environ Microbiol. 2003;69.7(July):3687–94.
53. Mugnai R, Sattamini A, Albuquerque dos Santos JA, Regua-Mangia AH. A Survey of *Escherichia coli* and Salmonella in the Hyporheic Zone of a Subtropical Stream: Their Bacteriological, Physicochemical and Environmental Relationships. PLoS One. 2015;10(6):e0129382.
54. Zhou J, Rudd KE. EcoGene 3.0. Nucleic Acids Res. 2013;41(Database issue):D613–24.
55. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–9.
56. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323 BioMed Central.
57. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106 BioMed Central.
58. Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, et al. An infection-relevant Transcriptomic compendium for Salmonella enterica Serovar Typhimurium. Cell Host Microbe. 2013;14(6):683–95 Elsevier Inc.
59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B. 1995;57:289–300.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.