Horizontal gene transfer rates decrease during transitions to new ecosystems

Swastik Mishra[1] **and Martin J. Lercher**[1] ✉

[1]Institute for Computer Science and Department of Biology, Heinrich Heine University, Düsseldorf, Germany

Horizontal gene transfer (HGT) is a major driver of bacterial evolution, allowing bacteria to acquire new traits from other organisms. It enables the transfer of advantageous traits, such as antibiotic resistance, which can rapidly spread across bacterial populations, affecting public health. Several studies have estimated how widespread HGT can be and suggested that environmental changes are associated with changes in the rate of HGT. It is often considered that environmental changes lead to an increase in HGT rates due to the adaptive benefits of HGT. However, previous studies have been limited to one or a few species and have often focused on specific genes of interest. Here, we systematically investigate the effect of environmental changes on the rate of HGT in a large number of bacterial genomes along a deep phylogeny. We extend the scope of use of existing state-of-the-art methods for HGT inference to infer the transition into new ecosystems and the gain of pathogenicity. Transitions to new environments are associated with a decrease in the rate of HGT, along with the strengthening of purifying selection. This decrease in HGT rates is consistent regardless of genomic relatedness between the donor and recipient. Our results are robust to the use of different methodological frameworks and have implications for understanding the evolution of bacteria and HGT rates in natural microbial communities.

Horizontal Gene Transfer, Bacterial Evolution, Ecosystem Transitions, Pathogenicity Gains, Purifying Selection

Correspondence: ***martin.lercher@hhu.de***

Introduction

Horizontal gene transfer (HGT) is an important driver of bacterial evolution (1). Instead of relying only on nucleotide substitutions, bacterial adaptations are driven by the acquisition of genomic fragments on the scale of entire genes or operons (2, 3), or by novel genes originating from recombination events (4). Bacteria can in this way acquire new traits from other organisms, allowing them to adapt to new environments. HGT rates are however, not constant across bacterial genomes. The rate of HGT is influenced by the rate of mutation and the rate of fixation of new genes in the recipient genome, through genetic drift and/or natural selection (5, 6). This includes the rate of gene loss, which can be higher in bacteria that have recently transitioned to a new environment (7, 8). Large-scale genomic studies have revealed that HGT events are frequent and widespread, with significant variation in transfer rates between different functional gene categories (9). Thus, the environment can strongly influence selection strengths, and hence the fixation rate.

Several studies have tried to show that environmental changes are associated with strong transient increases in HGT rate (10–13). However, these studies have been limited to a single or a few species, or have focused on specific genes or gene families, e.g., the rapid acquisition of antibiotic resistance genes via HGT. Despite the focus on antibiotic resistance genes due to legitimate public health concerns, HGT is not limited to antibiotic resistance genes and can generally involve all types of genes in a bacterial genome (14, 15). Most studies focus on laboratory experiments. The assembly and tracking of the entire mobilome in a natural microbial community is a challenging task, and the genomic HGT rate is difficult to measure directly (16). As a result, to date, no large-scale studies have systematically investigated the effect of environmental changes on the rate of bacterial HGT in natural ecosystems.

On the one hand, environmental changes could lead to an increase in the rate of HGT, since bacteria can adapt to new environments by acquiring new genes. New environments may have new pangenomes that can be accessed via HGT (17), and the selection pressures in the new environment may favor the acquisition and fixation of new genes (18). On the other hand, environmental changes could lead to a decrease in the rate of HGT. Several factors could contribute to this decrease, including changes in population size (environmental transitions often lead to bottlenecks) (19), changes in selection pressures on existing genes (20, 21), and a possible decrease in the availability of donors closely related to the recipient for transfers to be successful (17). This raises the question of what the rates of HGT are when there is a transition to a new environment, with respect to a context where there is no such transition.

Here, we address this question by analyzing bacterial genomes from different environments. Advances in metagenomics have made available the data for hundreds of bacterial genomes from different environments, as available in the GOLD database (22). We used these data to infer the number of transitions to new ecosystems, as well as the rates of HGT, across a deep phylogeny of bacterial genomes. We find that transitions to new ecosystems are associated with a decrease in the rate of HGT, suggesting that the adaptive benefits of HGT are outweighed by the selection pressures during transitions to new environments. Purifying selection is significantly stronger when bacteria make such transitions. Our results are robust to the use of different methods for inferring HGT and are consistent regardless of genomic relatedness between the HGT donor and the recipient.

Results

For HGT inferences, we used the orthologous gene families dataset of Bacteria from the EggNOG database v6 (23). EggNOG terms these gene families as non-supervised orthologous groups or *NOGs*. For inference of ecosystem transitions, we used the GOLD database. We limited our analyses to bacterial genomes where enough information is available such that they have at least 3 'Ecosystem Type' labels as defined in GOLD. Examples of these labels include 'fermented vegetables', 'fetus', 'fish products', 'freshwater', 'industrial wastewater', etc. GOLD follows a heirarchical labeling system, and this level provides the most fine-grained and diversified set of labels such that most genomes have a label at this level, unlike, for example, the 'Ecosystem Subtype level. Crucially, the data from GOLD span ecosystems ranging from natural microbial ones, including those found around humans, to laboratory ecosystems. These labels were cleaned and corrected for spelling mistakes, which resulted in 80 unique labels. Nucleotide sequences were retrieved for all genomes for which the EggNOG taxon ID could be reliably mapped to a genome in NCBI using the NCBI Datasets API (24). This resulted in a dataset of 8197 NOGs across 159 genomic taxa, 3718 of them being single-copy NOGs, such that each NOG is present in at least 4 taxa. 233 of these NOGs contained at least 95% of all taxa in the dataset, and were used to estimate a genome tree using ASTRAL-Pro 2 (25). The branch lengths of this tree were estimated with IQ-TREE 2 (26), using a concatenated alignment of multiple sequence alignments retrieved from EggNOG. The complete set of 8197 NOGs and the corresponding gene trees of EggNOG were used for HGT inferences.

For HGT inferences we used asymmetric Wagner parsimony (a maximum parsimony (MP) method) in Count (27), the maximum likelihood (ML) reconciliation model ALE (28), and the ML and MP options in GLOOME (29), to check for the robustness of our results. Note that ALE is different from any of the other methods in that it aims to infer the sets of Duplication, Transfer, and Loss events (DTL) that maximize the likelihood of the topology of a gene tree given that of the corresponding genome tree. It can therefore also infer the donor of a transferred gene in a recipient genome. We estimated the HGT rate by dividing the number of inferred HGT events on each branch by the length of the branch. The rate of HGT is therefore the number of HGT events per unit substitution rate.

For the inference of ecosystem transitions, we prepared a presence-absence profile matrix of the 'Ecosystem Type' labels for each genome of GOLD in our dataset. We used a numerical profile of the distribution of ecosystems across the genomes to infer the number of transitions to new ecosystems using the asymmetric Wagner parsimony model in Count. We also used a corresponding binary profile for the MP model in GLOOME to check for consistency in the results.

This method therefore extends the scope of the use of existing state-of-the-art methods for HGT inference. Note that a maximum parsimony method like this simply infers the gain and loss of a character across the genomes in a genome tree. Neither Count nor GLOOME (MP) makes any assumptions about the nature of that character. Hence, whether it is the presence/absence of a gene in a genome or the ecosystem mapped to a genome, is immaterial.

Inferences of HGT and ecosystem transfers were made independently.

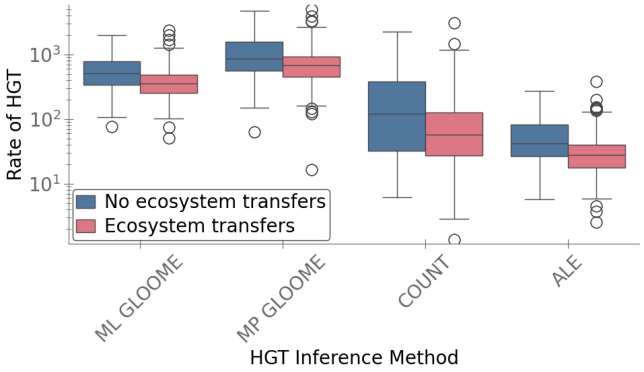

Fig. 1. HGT rates are lower at transitions to new ecosystems. Each pair of boxes in this figure represents the distribution of HGT rates on branches that have ecosystem transfers and those that do not have ecosystem transfers, when ecosystem transfers were inferred using Count. For all methods of HGT inference, i.e. all pairs of boxes shown here, the HGT rates are significantly lower at transitions to new ecosystems (two-tailed p-values $< 1.5 \times 10^{-4}$, CLES $\geqslant 0.630$, across independent Mann-Whitney U tests). For a summary of the effect sizes and mean HGT rate values and mean values, see Supplementary Table S1.

HGT rates are lower at transitions to new ecosystems. We compared branches in the genome tree where Count inferred ecosystem transfers to those where there were zero ecosystem transfers. The former branch set represents instances where there were transitions to new ecosystems, and the latter were without such transitions. Across all tested methods of HGT inference, the rate of HGT was found to be significantly lower in the set of branches with ecosystem transfers (Figure 1, two-tailed p-values $< 1.5 \times 10^{-4}$, CLES $\geqslant 0.630$, in independent Mann-Whitney U tests). The Supplementary Table S1 provides a summary of the Mann-Whitney U tests and the mean HGT rate values for all pairs of HGT inference methods and ecosystem transfer inference methods that were tested. These results are also consistent when the analysis is limited to single-copy NOGs (Supplementary Figure S1, Supplementary Table S4).

Decrease in HGT rates is consistent regardless of the genomic distance between donor and recipient. In new ecosystems, the compatibility of transferred genes with the existing genetic architecture can be reduced, thus decreasing HGT rates (30). This could be due to the lack of closely related donors in the new ecosystem. To test whether the observed decrease in HGT rates is present only in cases where the donor and recipient are distantly related, we binned the decrease in HGT rates by the genomic distance between the donor and recipient. The donors in this case could be inferred by using ALE.

Across all bins of genomic distance where there were branches with and without ecosystem transfers, the decrease in HGT rates was statistically significant (Figure 2). We per-

formed independent randomization tests of 10^4 iterations for each bin, and the decrease is at least this extreme with a p-value $< 5 \times 10^{-3}$ corrected for multiple testing using the false discovery rate (FDR) method (31)). The Supplementary Table S2 provides a summary of the randomization tests for each bin.

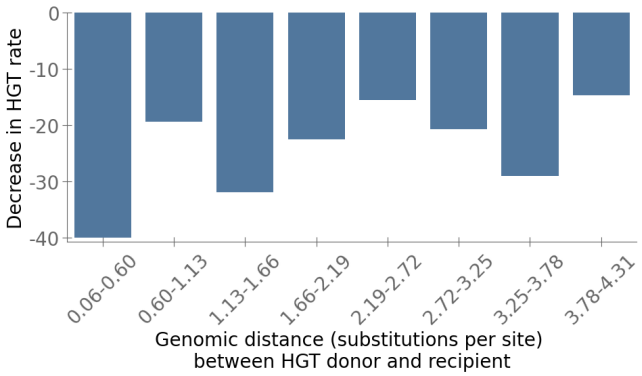

Fig. 2. HGT rates decrease at transitions to new ecosystems, irrespective of genomic distance between donor and recipient. For all bins of genomic distance for which there were branches with and without ecosystem transfers, we tested using randomization tests whether the decrease in HGT rates was statistically significant or not. The y-axis shows the mean HGT rate of branches without ecosystem transfers, minus the mean HGT rate of branches with ecosystem transfers. HGT rates were significantly lower at transitions to new ecosystems in all bins (two-tailed p-value $< 3.4 \times 10^{-3}$, corrected for multiple testing using the false discovery rate method).

Transitions to new ecosystems are associated with stronger purifying selection. Environmental transitions can significantly influence bacterial evolution through positive and purifying selection mechanisms. Certain cases of positive selection have been observed, for example, in the case of niche adaptation (32), and transitions to nutrient-depleted environments (33). However, genes generally undergo purifying selection on average. This selective pressure removes deleterious mutations that could impair protein function or cellular processes. This can be strengthened by environmental transitions. There is evidence for purifying selection under environmental stress (such as acidity, aridity, heat, or salinity) for genomic streamlining (34), as well as under nutrient depletion (35). The evidence for whether selection is more positive or negative after environmental transitions is therefore mixed and not generalizable, since the studies are limited to specific bacteria and/or specific environments.

We estimated the ratio ω (dN / dS) for the two branch sets, using HyPhy (36). The value of ω is a measure of the strength of selection on a gene, with $\omega < 1$ indicating purifying selection, $\omega = 1$ indicating neutral evolution and $\omega > 1$ indicating positive selection. As expected, the ω for both branch sets, for all gene families, is generally lower than 1. Branches with ecosystem transfers have a slightly lower ω compared to branches without ecosystem transfers (Figure 3; mean ω = 0.041 (σ= 0.042) for branches with ecosystem transfers and 0.090 (σ= 0.169) for branches without ecosystem transfers, in 2527 NOGs). A paired t-test showed that the difference in means was statistically significant (p-value $< 10^{-40}$). In

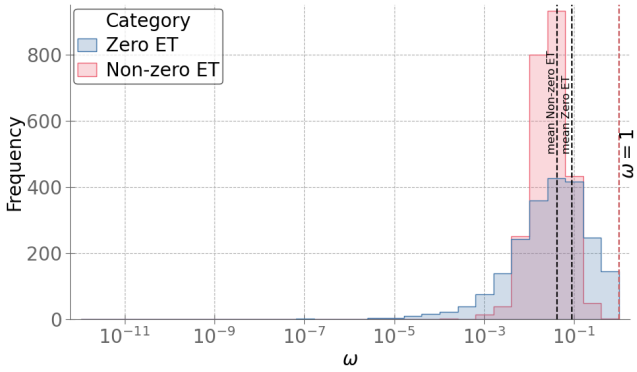

Fig. 3. Purifying selection is stronger at transitions to new ecosystems. The distribution of ω values for branches with ecosystem transfers (Non-zero ET) and branches without ecosystem transfers (Zero ET) are shown here. The ω values are significantly lower at transitions to new ecosystems (two-tailed p-value $< 10^{-40}$, paired t-test).

particular, the branch set without ecosystem transfers had a much larger variance (p-value $< 10^{-5}$ based on 10^5 iterations in a randomization test of variance ratios, where no iteration produced a ratio as extreme as the observed ratio). Therefore, purifying selection is significantly stronger in transitions to new ecosystems, leading to a much narrower distribution of ω values.

HGT rates are lower at transitions to pathogenic lifestyles. Bacterial pathogenicity has been associated with changes in environmental conditions, such as transition to a host environment, but also transitions within different ecosystems within the host, as well as changes in environmental conditions, with or without the host (37–39). A change in environmental conditions may not have a direct correlation with a change to a different ecosystem and may not be the result of the gain of virulence genes.

The analyses so far are limited to the framework of ecosystem transitions as a proxy for environmental changes. We performed a separate analysis with a similar pipeline to infer changes in HGT rates at transitions to pathogenic lifestyles.

For this analysis, we used a set of 151 *E.coli* and *Shigella* genomes retrieved from NCBI, with 98 being commensal and the rest pathogenic. As a result, instead of using the 'Ecosystem Type' labels, we have a binary label for each genome, indicating whether it is pathogenic or not. The gain and loss of pathogenicity was then inferred using Count. Independently, we inferred the HGT rates of 5503 orthologous groups of genes inferred using Orthofinder (40).

Unlike analysis of ecosystem transitions, very few branches were inferred to have pathogenicity gains (only 14 of 300). Although we still found a decrease in HGT rates at transitions to pathogenic lifestyles, this decrease is statistically significant only when ALE is used for inference of HGT (p-value $< 10^{-2}$, U= 2820, CLES= 0.704296, Mann-Whitney U test; mean HGT rates of 1.751×10^4 and 1.830×10^7, respectively, for branches with and without pathogenicity gains). Note that the branch lengths in this tree are extremely small compared to the tree of 159 taxa, and the inferred HGT rates are therefore much higher. The Supplementary Table S3 pro-

vides a summary of the Mann-Whitney U tests for all pairs of HGT inference methods and pathogenicity gain inference methods that were tested.

The fact that the gain of pathogenicity is associated with a decrease in HGT rates suggests that the decrease in HGT rates when bacteria experience environmental changes is not limited to consideration of ecosystems only. Furthermore, even if only in the case of ALE is it statistically significant, this decrease in the HGT rate is generally agreed upon by all of the methods of HGT inference we tested.

Discussion

Since HGT can rapidly introduce new genes into bacterial genomes, even with minimal systematic evidence, it is often assumed that HGT rates increase during changes in environmental conditions. However, it is difficult to measure the rate of HGT directly, especially in natural microbial communities without a priori knowledge of the mobilome and when there are no environmental changes. Experimental studies are often limited in the scope of the bacteria and environments they can study, and the results may not be generalizable.

In such a context, we have performed a large-scale genomic study to systematically investigate the effect of environmental changes on HGT rates. Our results show that transitions to new ecosystems or gain of pathogenicity are associated with a decrease in the rate of HGT. We have shown that the inference of this decrease in HGT rates is robust to the use of different methods.

This decrease in HGT rates is not limited to cases where the donor and recipient are distantly related. Strong, purifying selection pressures in the new environment may be a stronger factor in determining the rate of HGT, rather than the availability of closely related donors. Our results have implications for understanding the evolution of bacteria. They suggest that such evolutionary processes in bacteria may be less driven by adaptive benefits of HGT and more limited by selection pressures or other factors.

Although we infer that environmental changes *cause* a drop in HGT rates, technically it is possible that the causal link is the other way around. For example, bacteria that have a lower HGT rate at any given period of time may be more likely to move into new ecosystems or gain pathogenicity, and propagate their lineage to be sampled in the present day. This seems unlikely to us, given that experimental evidence suggests changes in the HGT rate at least for specific bacteria and genes under study, in response to environmental changes (10–13). We would like to note here several limitations of our study. First, the 'Ecosystem Type' labels in the GOLD database are not standardized. The scope of a label such as 'Fermented vegetables' may not be comparable to that of 'Freshwater' aquatic environments. This is a major problem, as the inference of ecosystem transfers is based on the assumption that the labels are comparable. A broad ecosystem label may result in fewer inferred transitions, missing out on the narrower niches that may be a part of the broader ecosystem.

Secondly, even if sampling issues are less of a concern in the gene families we analyzed (given the scale and dense taxon sampling of the dataset), the same cannot be said for the ecosystem labels. The number of genomes in the dataset for each ecosystem label is not uniform, and most genomes in GOLD are sampled in only one or a few ecosystems. This may result in certain ecosystem transfers not being inferred or in the inference of false positives.

Finally, the effect size of stronger selection during transitions to new ecosystems is quite small. It should be interpreted with caution, as it may not be biologically significant.

In conclusion, leveraging the large-scale genomic data available today, as well as the novel use of maximum parsimony inference, we have systematically inferred changes in environmental conditions and their effect on HGT rates. Importantly, the use of metagenomics data enables us to make inferences about bacterial evolution in natural microbial communities, as opposed to laboratory experiments. Our results suggest that selection pressures in new environments may be a stronger factor in determining the rate of HGT, rather than the availability of closely related donors, and generally outweigh the adaptive benefits of HGT. Our analyses suggest the role of selection in shaping the rate of HGT in bacteria. However, we have not characterized the exact factors that contribute to the decrease in HGT rates. This decrease could be due to changes in population size and may differ between different types of genes depending on their function, connectivity, or other characteristics. Genes related to bacteria survival in the new environment may be more likely to be transferred and the rate of transfer of these specific genes may be higher, while the average rate of HGT across the genome may be lower. These are directions in which future investigations could be directed.

Methods

Data retrieval and processing. We retrieved NOGs from the EggNOG database v6, and the corresponding gene trees. The genome tree topology was estimated using broad-distribution NOGs (at least 95% of taxa represented in each NOG) with ASTRAL-Pro 2. Branch lengths on this topology were estimated using IQ-TREE 2 using 'Q.pfam+I+R8' model on a concatenated set of multiple sequence alignments retrieved from EggNOG. The genome tree was rooted using the Minimum Ancestor Deviation method (41), which has been shown to be more accurate than other rooting methods (42). The 'Ecosystem Type' labels for each genome were retrieved from the GOLD database (v9). We used the NCBI Datasets API to retrieve nucleotide sequences for each genome. Nucleotide sequences were used to prepare codon alignments for each NOG, translated from the amino acid multiple sequence alignments provided by EggNOG using PAL2NAL (43).

Pathogenic and commensal *E.coli* and *Shigella* genomes were retrieved from NCBI RefSeq using the NCBI Datasets API. Strain genomes with at least 95% CheckM completeness and without suppressed assembly status were retained. The pathogenicity of each genome was inferred from the 'Pathogenicity' label in the assembly attributes. OrthoFinder

was used to infer orthologous groups of genes, and gene trees were estimated using IQ-TREE 2 with the 'JTT' substitution model.

Inference of HGT, Environmental transitions, and ω. HGT was inferred using Count, ALE, and GLOOME. Count was used in a maximum parsimony framework with asymmetric Wagner parsimony. For ALE, we used the 'undated' model, which does not require a dated species tree. For maximum parsimony inferences of HGT, ecosystem gains, and pathogenicity gains, the gain penalty ratio for Count and gain-loss ratio for GLOOME (MP) was set to 1.0. For 2527 NOGs for which we could reliably retrieve nucleotide sequences for all taxa, we estimated the ω ratio using HyPhy BUSTED (36), with the branches of the subtree where a NOG is mapped on the genome tree, partitioned into two sets of branches – with and without ecosystem gains – resulting in two values of ω for each NOG. NOGs with less than two branches in either set were excluded from the analysis. Limiting the analysis to NOGs with higher proportions of branches in the smaller set does not change the results of the selection analysis.

Code and data availability

The code used in the analysis of ecosystem transitions is available in the following Gitlab repository: https://gitlab.cs.uni-duesseldorf.de/mishra/hgt_rates_and_ecosystem_transfers For pathogenicity analysis, it is at: https://gitlab.cs.uni-duesseldorf.de/mishra/hgt_rates_and_pathogenicity_gains The data used or generated in this study can be retrieved from Zenodo: https://doi.org/10.5281/zenodo.14555205.

Competing interests

The authors have declared no competing interest.

Funding

This work was supported by a grant from Volkswagenstiftung in the "Life?" initiative (to M.J.L.)

Bibliography

1. Brian J. Arnold, I.-Ting Huang, and William P. Hanage. Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology*, pages 1–13, November 2021. ISSN 1740-1534. doi: 10.1038/s41579-021-00650-4.
2. Tin Yau Pang and Martin J. Lercher. Each of 3,323 metabolic innovations in the evolution of E. coli arose through the horizontal transfer of a single DNA segment. *Proceedings of the National Academy of Sciences*, 116(1):187–192, January 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1718997115.
3. Csaba Pál, Balázs Papp, and Martin J. Lercher. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genetics*, 37(12):1372–1375, December 2005. ISSN 1546-1718. doi: 10.1038/ng1686.
4. Maria Izabel A Cavassim, Stig U Andersen, Thomas Bataillon, and Mikkel Heide Schierup. Recombination Facilitates Adaptive Evolution in Rhizobial Soil Bacteria. *Molecular Biology and Evolution*, 38(12):5480–5490, December 2021. ISSN 1537-1719. doi: 10.1093/molbev/msab247.
5. Adam C. Retchless and Jeffrey G. Lawrence. Ecological Adaptation in Bacteria: Speciation Driven by Codon Selection. *Molecular Biology and Evolution*, 29(12):3669–3683, December 2012. ISSN 0737-4038. doi: 10.1093/molbev/mss171.
6. Christopher M. Thomas and Kaare M. Nielsen. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology*, 3(9):711–721, September 2005. ISSN 1740-1534. doi: 10.1038/nrmicro1234.
7. Sanna Koskiniemi, Song Sun, Otto G. Berg, and Dan I. Andersson. Selection-Driven Gene Loss in Bacteria. *PLOS Genetics*, 8(6):e1002787, June 2012. ISSN 1553-7404. doi: 10.1371/journal.pgen.1002787.
8. Bert Ely. Recombination and gene loss occur simultaneously during bacterial horizontal gene transfer. *PLOS ONE*, 15(1):e0227987, January 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0227987.
9. Michael Sheinman, Ksenia Arkhipova, Peter F Arndt, Bas E Dutilh, Rutger Hermsen, and Florian Massip. Identical sequences found in distant genomes reveal frequent horizontal transfer across the bacterial domain. *eLife*, 10:e62719, June 2021. ISSN 2050-084X. doi: 10.7554/eLife.62719.
10. Laura C. Woods, Rebecca J. Gorrell, Frank Taylor, Tim Connallon, Terry Kwok, and Michael J. McDonald. Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *Proceedings of the National Academy of Sciences*, 117(43):26868–26875, October 2020. doi: 10.1073/pnas.2005331117.
11. Ying-Xian Goh, Sai Manohar Balu Anupoju, Anthony Nguyen, Hailong Zhang, Monica Ponder, Leigh-Anne Krometis, Amy Pruden, and Jingqiu Liao. Evidence of horizontal gene transfer and environmental selection impacting antibiotic resistance evolution in soil-dwelling Listeria. *Nature Communications*, 15(1):10034, November 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-54459-9.
12. Katayoun Dadeh Amirfard, Momoko Moriyama, Satoru Suzuki, and Daisuke Sano. Effect of environmental factors on conjugative transfer of antibiotic resistance genes in aquatic settings. *Journal of Applied Microbiology*, 135(6):lxae129, June 2024. ISSN 1364-5072. doi: 10.1093/jambio/lxae129.
13. Hande Acar Kirit, Mato Lagator, and Jonathan P. Bollback. Experimental determination of evolutionary barriers to horizontal gene transfer. *BMC microbiology*, 20(1):326, October 2020. ISSN 1471-2180. doi: 10.1186/s12866-020-01983-5.
14. Bram van Dijk, Paulien Hogeweg, Hilje M Doekes, and Nobuto Takeuchi. Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. *eLife*, 9:e56801, May 2020. ISSN 2050-084X. doi: 10.7554/eLife.56801.
15. Charles Coluzzi, Martin Guillemet, Fanny Mazzamurro, Marie Touchon, Maxime Godfroid, Guillaume Achaz, Philippe Glaser, and Eduardo P C Rocha. Chance Favors the Prepared Genomes: Horizontal Transfer Shapes the Emergence of Antibiotic Resistance Mutations in Core Genes. *Molecular Biology and Evolution*, 40(10):msad217, October 2023. ISSN 1537-1719. doi: 10.1093/molbev/msad217.
16. Ilana Lauren Brito. Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology*, 19(7):442–453, July 2021. ISSN 1740-1534. doi: 10.1038/s41579-021-00534-7.
17. Marija Dmitrijeva, Janko Tackmann, João Frederico Matias Rodrigues, Jaime Huerta-Cepas, Luis Pedro Coelho, and Christian von Mering. A global survey of prokaryotic genomes reveals the eco-evolutionary pressures driving horizontal gene transfer. *Nature Ecology & Evolution*, pages 1–13, March 2024. ISSN 2397-334X. doi: 10.1038/s41559-024-02357-0.
18. Jan Engelstädter and Danesh Moradigaravand. Adaptation through genetic time travel? Fluctuating selection can drive the evolution of bacterial transformation. *Proceedings of the Royal Society B: Biological Sciences*, 281(1775):20132609, January 2014. doi: 10.1098/rspb.2013.2609.
19. Yevgeniy Raynes, Christina L. Burch, and Daniel M. Weinreich. When good mutations go bad: How population size can change the direction of natural selection, April 2021.
20. James P. J. Hall, David Williams, Steve Paterson, Ellie Harrison, and Michael A. Brockhurst. Positive selection inhibits gene mobilization and transfer in soil bacterial communities. *Nature Ecology & Evolution*, 1(9):1348–1353, September 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0250-3.
21. Gabriel Carvalho, David Fouchet, Gonché Danesh, Anne-Sophie Godeux, Maria-Halima Laaberki, Dominique Pontier, Xavier Charpentier, and Samuel Venner. Bacterial Transformation Buffers Environmental Fluctuations through the Reversible Integration of Mobile Genetic Elements. *mBio*, 11(2):10.1128/mbio.02443–19, March 2020. doi: 10.1128/mbio.02443-19.
22. Supratim Mukherjee, Dimitri Stamatis, Cindy Tianqing Li, Galina Ovchinnikova, Jon Bertsch, Jagadish Chandrabose Sundaramurthi, Mahathi Kandimalla, Paul A Nicolopoulos, Alessandro Favognano, I-Min A Chen, Nikos C Kyrpides, and T B K Reddy. Twenty-five years of Genomes OnLine Database (GOLD): Data updates and new features in v.9. *Nucleic Acids Research*, 51(D1):D957–D963, January 2023. ISSN 0305-1048. doi: 10.1093/nar/gkac974.
23. Ana Hernández-Plaza, Damian Szklarczyk, Jorge Botas, Carlos P Cantalapiedra, Joaquín Giner-Lamia, Daniel R Mende, Rebecca Kirsch, Thomas Rattei, Ivica Letunic, Lars J Jensen, Peer Bork, Christian von Mering, and Jaime Huerta-Cepas. eggNOG 6.0: Enabling comparative genomics across 12 535 organisms. *Nucleic Acids Research*, 51(D1):D389–D394, January 2023. ISSN 0305-1048. doi: 10.1093/nar/gkac1022.
24. Eric W. Sayers, Evan E. Bolton, J. Rodney Brister, Kathi Canese, Jessica Chan, Donald C. Comeau, Ryan Connor, Kathryn Funk, Chris Kelly, Sunghwan Kim, Tom Madej, Aron Marchler-Bauer, Christopher Lanczycki, Stacy Lathrop, Zhiyong Lu, Francoise Thibaud-Nissen, Terence Murphy, Lon Phan, Yuri Skripchenko, Tony Tse, Jiyao Wang, Rebecca Williams, Barton W. Trawick, Kim D. Pruitt, and Stephen T. Sherry. Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50(D1):D20–D26, January 2022. ISSN 1362-4962. doi: 10.1093/nar/gkab1112.
25. Chao Zhang and Siavash Mirarab. ASTRAL-Pro 2: Ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics*, 38(21):4949–4950, November 2022. ISSN 1367-4803. doi: 10.1093/bioinformatics/btac620.
26. Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, May 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa015.
27. Miklós Csűös. Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*, 26(15):1910–1912, August 2010. ISSN 1367-4803. doi: 10.1093/

bioinformatics/btq315.

28. Gergely J. Szöllősi, Adrián Arellano Davín, Eric Tannier, Vincent Daubin, and Bastien Boussau. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1678):20140335, September 2015. doi: 10.1098/rstb.2014.0335.
29. Ofir Cohen, Haim Ashkenazy, Frida Belinky, Dorothée Huchon, and Tal Pupko. GLOOME: Gain loss mapping engine. *Bioinformatics*, 26(22):2914–2915, November 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq549.
30. Jane Wiedenbeck and Frederick M. Cohan. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, 35 (5):957–976, September 2011. ISSN 0168-6445. doi: 10.1111/j.1574-6976.2011.00292.x.
31. Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, January 1995. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1995.tb02031.x.
32. Sha Cao, Gerrit Brandis, Douglas L. Huseby, and Diarmaid Hughes. Positive Selection during Niche Adaptation Results in Large-Scale and Irreversible Rearrangement of Chromosomal Gene Order in Bacteria. *Molecular Biology and Evolution*, 39(4):msac069, April 2022. ISSN 1537-1719. doi: 10.1093/molbev/msac069.
33. Ruben Props, Pieter Monsieurs, Peter Vandamme, Natalie Leys, Vincent J. Denef, and Nico Boon. Gene Expansion and Positive Selection as Bacterial Adaptations to Oligotrophic Conditions. *mSphere*, 4(1):10.1128/mspheredirect.00011–19, February 2019. doi: 10.1128/mspheredirect.00011-19.
34. Anna K Simonsen. Environmental stress leads to genome streamlining in a widely distributed species of soil bacteria. *The ISME Journal*, 16(2):423–434, February 2022. ISSN 1751-7362. doi: 10.1038/s41396-021-01082-x.
35. Liat Shenhav and David Zeevi. Resource conservation manifests in the genetic code. *Science*, 370(6517):683–687, November 2020. doi: 10.1126/science.aaz9642.
36. Ben Murrell, Steven Weaver, Martin D. Smith, Joel O. Wertheim, Sasha Murrell, Anthony Aylward, Kemal Eren, Tristan Pollner, Darren P. Martin, Davey M. Smith, Konrad Scheffler, and Sergei L. Kosakovsky Pond. Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution*, 32(5):1365–1371, May 2015. ISSN 0737-4038. doi: 10.1093/molbev/msv035.
37. Aaron Mischa Nuss, Franziska Schuster, Louisa Roselius, Johannes Klein, René Bücker, Katharina Herbst, Ann Kathrin Heroven, Fabio Pisano, Christoph Wittmann, Richard Münch, Johannes Müller, Dieter Jahn, and Petra Dersch. A Precise Temperature-Responsive Bistable Switch Controlling Yersinia Virulence. *PLOS Pathogens*, 12(12):e1006091, December 2016. ISSN 1553-7374. doi: 10.1371/journal.ppat.1006091.
38. Chui Yoke Chin, Kyle A. Tipton, Marjan Farokhyfar, Eileen M. Burd, David S. Weiss, and Philip N. Rather. A high-frequency phenotypic switch links bacterial virulence and environmental survival in Acinetobacter baumannii. *Nature Microbiology*, 3(5):563–569, May 2018. ISSN 2058-5276. doi: 10.1038/s41564-018-0151-5.
39. Nat F. Brown, Mark E. Wickham, Brian K. Coombes, and B. Brett Finlay. Crossing the Line: Selection and Evolution of Virulence Traits. *PLOS Pathogens*, 2(5):e42, May 2006. ISSN 1553-7374. doi: 10.1371/journal.ppat.0020042.
40. David M. Emms and Steven Kelly. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1):1–14, December 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1832-y.
41. Fernando Domingues Kümmel Tria, Giddy Landan, and Tal Dagan. Phylogenetic rooting using minimal ancestor deviation. *Nature Ecology & Evolution*, 1(1):1–7, June 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0193.
42. Taylor Wade, L. Thiberio Rangel, Soumya Kundu, Gregory P. Fournier, and Mukul S. Bansal. Assessing the accuracy of phylogenetic rooting methods on prokaryotic gene families. *PLOS ONE*, 15(5):e0232950, May 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0232950.
43. Mikita Suyama, David Torrents, and Peer Bork. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(suppl_2):W609–W612, July 2006. ISSN 0305-1048. doi: 10.1093/nar/gkl315.

Supplementary Information

Inference method HGT	Inference method Ecosystem Transfer	Mean$^{\alpha}$ HGT rate	Mean$^{\beta}$ HGT rate	Mann-Whitney U	p-value	CLES	#Branches$^{\alpha}$	#Branches$^{\beta}$
ALE	Count	37.023	59.576	15391.000	6.7×10^{-8}	0.685	208	108
Count	Count	128.615	314.920	14157.500	0.00015	0.630	208	108
ML GLOOME	Count	414.392	602.076	14987.000	1.1×10^{-6}	0.667	208	108
MP GLOOME	Count	804.285	1182.449	14246.000	9.2×10^{-5}	0.634	208	108
ALE	MP GLOOME	37.023	59.576	15391.000	6.7×10^{-8}	0.685	208	108
Count	MP GLOOME	128.615	314.920	14157.500	0.00015	0.630	208	108
ML GLOOME	MP GLOOME	414.392	602.076	14987.000	1.1×10^{-6}	0.667	208	108
MP GLOOME	MP GLOOME	804.285	1182.449	14246.000	9.2×10^{-5}	0.634	208	108

Table S1. Comparison of HGT rates at branches with transitions to new ecosystems using different HGT inference methods and ecosystem transfer inference methods.
$^{\alpha}$ Ecosystem(s) gained.
$^{\beta}$ No ecosystem gained.

Genomic distance bin	Mean HGT rate$^{\alpha}$	Mean HGT rate$^{\beta}$	Decrease in HGT rate	FDR corrected p-value
0.06-0.60	47.7507	87.6867	-39.9360	0.0001
0.60-1.13	29.4750	48.8658	-19.3908	0.0001
1.13-1.66	18.8310	50.6828	-31.8518	0.0001
1.66-2.19	22.0159	44.5591	-22.5431	0.0001
2.19-2.72	19.8674	35.4545	-15.5871	0.0001
2.72-3.25	20.4401	41.1379	-20.6978	0.0001
3.25-3.78	18.7013	47.6783	-28.9770	0.0001
3.78-4.31	16.2738	30.9129	-14.6392	0.0034
4.31-4.84	18.0784	NA	NA	NA

Table S2. Randomisation test results for the decrease in HGT rates at transitions to new ecosystems, binned by genomic distance between donor and recipient.

Inference method HGT	Inference method Pathogenicity Transfer	Mean$^{\alpha}$ HGT rate	Mean$^{\beta}$ HGT rate	Mann-Whitney U	p-value	CLES	#Branches$^{\alpha}$	#Branches$^{\beta}$
ALE	Count	1.751×10^{4}	1.830×10^{7}	2820.000	0.0099	0.704	14	286
Count	Count	4.584×10^{5}	4.845×10^{7}	2141.000	0.66	0.535	14	286
ML GLOOME	Count	1.544×10^{6}	7.167×10^{6}	1717.000	0.37	0.429	14	286
MP GLOOME	Count	6.273×10^{5}	7.811×10^{7}	2154.000	0.63	0.538	14	286
ALE	MP GLOOME	1.751×10^{4}	1.830×10^{7}	2820.000	0.0099	0.704	14	286
Count	MP GLOOME	4.584×10^{5}	4.845×10^{7}	2141.000	0.66	0.535	14	286
ML GLOOME	MP GLOOME	1.544×10^{6}	7.167×10^{6}	1717.000	0.37	0.429	14	286
MP GLOOME	MP GLOOME	6.273×10^{5}	7.811×10^{7}	2154.000	0.63	0.538	14	286

Table S3. Comparison of HGT rates at transitions to pathogenic lifestyles using different HGT inference methods and pathogenicity gain inference methods.
$^{\alpha}$ Pathogenicity gained
$^{\beta}$ Pathogenicity not gained

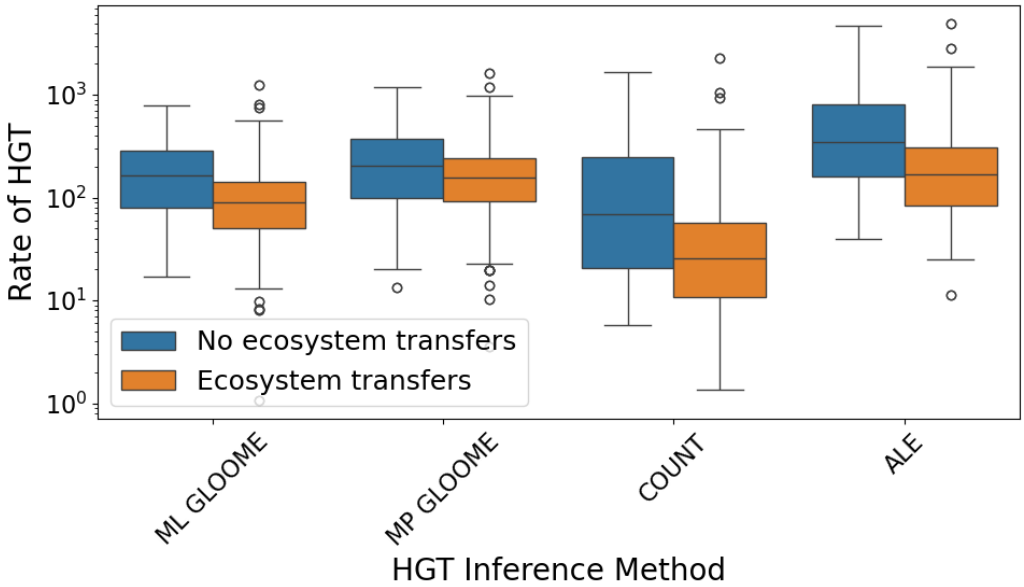

Fig. S1. HGT rates are lower at transitions to new ecosystems, for single-copy NOGs. For all methods shown here, the HGT rates are significantly lower in branches with ecosystem transfers (two-tailed p-values $< 2.14 \times 10^{-2}$, CLES $\geqslant 0.579$, across independent Mann-Whitney U tests). For a summary of the effect sizes and mean values, see Supplementary Table S4.

Inference method HGT	Inference method Ecosystem Transfer	Mean$^{\alpha}$ HGT rate	Mean$^{\beta}$ HGT rate	Mann-Whitney U	p-value	CLES	#Branches$^{\alpha}$	#Branches$^{\beta}$
ALE	Count	303.174	644.845	15972.000	7.6×10^{-10}	0.711	208	108
Count	Count	72.633	199.792	14841.500	2.8×10^{-6}	0.661	208	108
ML GLOOME	Count	124.299	204.700	14914.000	1.8×10^{-6}	0.664	208	108
MP GLOOME	Count	207.547	275.251	13004.000	0.021	0.579	208	108
ALE	MP GLOOME	303.174	644.845	15972.000	7.6×10^{-10}	0.711	208	108
Count	MP GLOOME	72.633	199.792	14841.500	2.8×10^{-6}	0.661	208	108
ML GLOOME	MP GLOOME	124.299	204.700	14914.000	1.8×10^{-6}	0.664	208	108
MP GLOOME	MP GLOOME	207.547	275.251	13004.000	0.021	0.579	208	108

Table S4. Comparison of HGT rates at branches with transitions to new ecosystems using different HGT inference methods and ecosystem transfer inference methods, for single-copy NOGs.
$^{\alpha}$ Ecosystem(s) gained.
$^{\beta}$ No ecosystem gained.