

# Inference of the Demographic Histories and Selective Effects of Human Gut Commensal Microbiota Over the Course of Human History

Jonathan C. Mah <sup>1</sup>, Kirk E. Lohmueller <sup>2,3,\*</sup>, Nandita R. Garud <sup>2,3,\*</sup>

<sup>1</sup>Bioinformatics Interdepartmental Program, University of California, Los Angeles, USA

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA

<sup>3</sup>Department of Human Genetics, University of California, Los Angeles, USA

\*Corresponding authors: Emails: [ngarud@ucla.edu](mailto:ngarud@ucla.edu); [klohmueller@ucla.edu](mailto:klohmueller@ucla.edu).

Associate editor: Crystal Hepp

## Abstract

Despite the importance of gut commensal microbiota to human health, there is little knowledge about their evolutionary histories, including their demographic histories and distributions of fitness effects (DFEs) of mutations. Here, we infer the demographic histories and DFEs for amino acid-changing mutations of 39 of the most prevalent and abundant commensal gut microbial species found in Westernized individuals over timescales exceeding human generations. Some species display contractions in population size and others expansions, with several of these events coinciding with several key historical moments in human history. DFEs across species vary from highly to mildly deleterious, with differences between accessory and core gene DFEs largely driven by genetic drift. Within genera, DFEs tend to be more congruent, reflective of underlying phylogenetic relationships. Together, these findings suggest that gut microbes have distinct demographic and selective histories.

**Keywords:** microbiome, DFE, selection, population genetics, demography, evolution

## Introduction

Diversity in the human gut microbiome varies significantly across geographic regions around the world. For example, rural African microbiomes display greater levels of diversity than urban Western microbiomes at multiple levels including at the species (Sonnenburg et al. 2016), gene (Tett et al. 2019), and nucleotide (Nayfach and Pollard 2015; Tett et al. 2019) levels. A number of demographic and selective forces may be responsible for these differences in diversity, including urbanization (Obregon-Tito et al. 2015), antibiotic usage (Clemente et al. 2015; Sonnenburg and Sonnenburg 2019), shifts in diet (Filippo et al. 2010; Schnorr et al. 2014; Blaser 2018), and human migration (Yatsunen et al. 2012). For example, reductions in dietary fiber in mice lead to irreversible reductions in genetic diversity in the gut microbiome and are thought to be a contributing factor to the depleted diversity of rural agrarian populations compared to hunter-gatherer populations (Sonnenburg et al. 2016). Additionally, some microbes such as *Helicobacter pylori* (Falush et al. 2003) and *Prevotella copri* (Tett et al. 2019) display phylogenetic patterns mirroring human migration events around the world, suggesting that human demographic events themselves may have shaped bacterial genomic variation.

To date, there has not been a systematic inference of the demographic histories and distributions of fitness effects (DFEs) of new mutations of human commensal gut microbiota, nor has there been any investigation into how these quantities might vary across species. Such inferences are important for a number of reasons. Demographic inference provides an understanding of pre-historical events, such as

continental migrations as well as population contractions and expansions. Additionally, demographic inference enables the ability to detect genomic regions subject to selective forces by providing an expectation of genetic diversity under neutral conditions. Inference of the DFE, which quantifies the proportion of new mutations that are deleterious, neutral, or beneficial, is necessary for addressing fundamental questions such as the fate of colonizing lineages within a human host and the evolutionary capacity of a population to respond to novel selection pressures (Chevin et al. 2010; Dapa et al. 2023).

A key population genetics approach (Wakeley and Hey 1997; Nielsen 2000; Gutenkunst et al. 2009) for the inference of demographic histories and DFEs among eukaryotes leverages a summary statistic known as the site-frequency spectrum (SFS) (Fig. 1b). The SFS describes the distribution of minor allele frequencies from a given sample of DNA sequences and is highly sensitive to the impact of demography and selection (Nielsen 2000). For example, relative to the SFS of a population in demographic equilibrium, a population undergoing a demographic expansion and/or experiencing purifying selection would harbor a higher proportion of rare variants, i.e. single nucleotide polymorphisms (SNPs) present in a small number of individuals relative to the overall sample. Conversely, a population undergoing a demographic contraction and/or experiencing balancing selection would harbor a higher proportion of intermediate frequency variants.

While the SFS has been extensively used to infer demography and selective effects in eukaryotes (Eyre-Walker et al. 2006; Boyko et al. 2008; Gutenkunst et al. 2009; Beichman et al. 2018), it has been under-utilized for bacteria

Received: February 6, 2024. Revised: November 7, 2024. Accepted: January 7, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



While bacteria reproduce asexually, most commensal gut bacteria experience extensive recombination (Smith et al. 1993; Vos and Didelot 2009; Garud et al. 2019; Lin and Kussell 2019; Sakoparnig et al. 2021; Liu and Good 2024). This recombination plays a critical role in shuffling genetic material and breaking up correlations between loci in the genome on long timescales exceeding within-host evolution. The effects of recombination on bacterial genomic diversity is evident in the rapid decay in linkage disequilibrium (LD) with genomic distance in samples of lineages from human hosts around the world (Garud et al. 2019; Lin and Kussell 2019; Sakoparnig et al. 2021; Liu and Good 2024), as well as phylogenetic inconsistencies at the SNP and gene level with trees built from genome-wide data (Garud et al. 2019, Sakoparnig et al. 2021). It has been estimated that recombination contributes on the order of 10 times more to variation than mutations to human commensal gut bacteria, indicating that recombination overwrites any semblance of clonality among circulating lineages (Liu and Good 2024). Given the relatively high rates of recombination in gut microbiota, we propose to use the SFS constructed using a collection of diverse lineages across hosts to gain insights into the evolutionary history of commensal bacteria.

In this study, we perform an SFS-based inference of the demographic history and fitness effects acting on nonsynonymous mutations in 39 prevalent commensal gut species using a panel of healthy hosts from North America. Concretely, we examine the evolutionary dynamics operating across hosts on timescales exceeding a human lifetime by building SFSs composed of a single strain per individual microbiome for each species. We infer that human gut commensal bacteria have experienced a range of demographic histories, including contractions and expansions in effective population size over the past  $\sim 10^3$  to  $10^6$  years of human history. Additionally, we infer DFEs from the SFS, and find that bacterial DFEs are more congruent within versus between-genera, indicative of underlying phylogenetic relationships. Finally, we find that both genetic drift and natural selection play a role in driving differences in the DFE between core and accessory genes, with drift likely being a stronger driving force. Taken together, these findings suggest that common commensal gut microbiota have distinct evolutionary histories.

## Results

### Data

To infer the demographic histories and DFEs of new mutations of gut commensal bacteria found in Western urbanized populations, we analyzed shotgun metagenomic data from a panel of 693 healthy hosts from four datasets from North America (Human Microbiome Consortium 2012; Lloyd-Price et al. 2017), Europe (Xie et al. 2016; Korpela et al. 2018), and China (Qin et al. 2010) that we previously analyzed (Garud et al. 2019). To identify SNPs and gene copy number variants (CNVs) from these metagenomic samples, reads were mapped to reference genomes using a standard pipeline (Nayfach et al. 2016) (See Methods). Since each host may be colonized by multiple strains of the same species, to obtain a sample of lineages across hosts, we leveraged a previous approach we developed to “quasi-phase” haplotypes corresponding to the dominant strain present in each host for each species (Garud et al. 2019) (See Methods). In total, we obtained 5,416 “quasi-phased” (QP) genomes from 43 different species across the hosts in our datasets (Fig. 1a). This

collection of genomes provided the ability to infer the demographic histories of bacteria in Westernized individuals.

## Gut Microbiota Experience a Range of Demographic Changes on Time Scales Spanning Human History

### Building SFSs From Metagenomic Data From a Panel of Hosts

To infer demographic histories of human gut commensal microbiota, we analyzed SFSs composed of lineages from different hosts as opposed to within-host SFSs. SFSs composed of diverse lineages across hosts may contain information about population histories on timescales exceeding a human lifetime as opposed to demographic changes occurring within a host (Fig. 1). Moreover, lineages found across hosts experience extensive levels of recombination, while lineages within a host may be derived from a recent clonal ancestor.

SFSs were composed of one QP genome per host for each of 39 of the most prevalent bacterial species that had a minimum of 14 QP genomes (See Methods; Fig. 1a). We required a minimum sample size of 14 QP genomes to allow for a large number of species to be included in our analyses, while maintaining enough bins in the SFS to infer population demographic histories. Additionally, we initially focused our analysis on the core genome of each species, i.e. genes present in at least 95% of samples (See Methods).

Since population structure can result in false inferences of demography and selection by generating an excess of intermediate frequency SNPs (Städler et al. 2009; Gazave et al. 2014), we controlled for population structure at two levels. First, while different hosts typically harbor their own genetically distinct set of strains, occasionally there are some exceptionally closely related lineages (divergence  $< 2 \times 10^{-4}$ /bp) circulating in the population (Garud et al. 2019). To avoid structure arising from potential clonal expansions, we excluded lineages with divergences  $< 2 \times 10^{-4}$ /bp to any other lineage in the dataset, as we previously did to control for population structure (Garud et al. 2019) (See Methods).

Second, previous work has found that several gut bacterial species comprise deeply diverged clades lacking gene flow between clades (Costea et al. 2017; Garud et al. 2019; Liu and Good 2024). In other words, lineages belonging to the same clade exhibit significantly higher amounts of recombination than lineages part of different clades, as demonstrated by more rapid rates of LD decay and increased phylogenetic inconsistencies of individual SNPs with a genome-wide phylogeny (Garud et al. 2019; Liu and Good 2024). Thus, to control for population structure as a potential confounder for demographic inference, we analyzed SFSs belonging to lineages comprising the largest clade for each species, based on our previous clade inferences (Garud et al. 2019). In supplementary fig. S1, Supplementary Material online we show the impact of population structure control and subsequent projection on the shape of the empirical SFS.

Finally, selective forces acting on nonsynonymous sites may confound inference of demography by changing the shape or skew of the SFS. For example, purifying selection typically results in an excess of rare variants, thus skewing the nonsynonymous SFS toward low frequency SNPs. By contrast, the synonymous SFS is thought to be more neutrally evolving, thereby allowing for more sensitive inference of the underlying demographic history (Williamson et al. 2005; Boyko et al. 2008). To control for the potential effects of selection, we constructed SFSs using only synonymous SNPs.

### Demographic Inferences With *dad*

To infer the demographic histories of gut commensals, we used a statistical inference package known as *dad* (Gutenkunst et al. 2009), which is a maximum-likelihood inference method based on the SFS. We explored three different putative demographic models in this framework for each species, permitting the inference of population expansions versus contractions (Fig. 1c). Selection of the best-fitting model was performed using the Akaike information criterion (AIC) (See Methods; [supplementary fig. S2, Supplementary Material online](#)).

Demographic inference performed with *dad* recovers two main evolutionary parameters:  $\rho$ , which is the ratio of the current effective population size ( $N_{\text{Curr}}$ ) to the ancestral effective population size ( $N_{\text{Anc}}$ ), and  $t$ , the time since the most recent instantaneous size change in units of generations scaled by  $2N_{\text{Anc}}$  (Gutenkunst et al. 2009). For each species, we found the maximum-likelihood estimates (MLEs) of  $\rho$  and  $t$  by computing the expected SFS under a given set of parameters and evaluating its fit to the observed SFS (See Methods, [Fig. 2c and d, supplementary fig. S3, Supplementary Material online](#)).

As an example, we highlight the demographic inference of two species, *Bacteroides xylanisolvens* (Fig. 2a) and *Ruminococcus bromii* (Fig. 2b). For both species, a two-epoch demographic model best fits the data (Fig. 2e and f). Analysis of the SFS of *Bacteroides xylanisolvens* reveals an effective population size contraction in the range of 0.02 to 0.66 times the ancestral effective population size on a time-scale of roughly  $\sim 58$  to  $\sim 9,000$  years ago (these values represent the respective 95% confidence intervals of  $\rho$  as well as  $t$  rescaled to years, [supplementary fig. S4, Supplementary Material online](#)) (See Methods). By contrast, *R. bromii* experienced a population expansion between 3.57 to 3.99 times the ancestral effective population size on a time-scale of roughly  $\sim 44,000$  to 63,000 years ago. Confirming these inferences, qualitatively, the observed SFS of *Bacteroides xylanisolvens* has an excess of common variants compared to a one-epoch model, consistent with a contraction, whereas *R. bromii* has an excess of rare variants, consistent with an expansion in effective population size.

Across the 39 species for which we were able to infer demographic models, we found that the two-epoch demographic model best fits the data in all analyzed bacteria except for *Bacteroides stercoris*, *Bacteroides plebeius*, and *Roseburia intestinalis*, all of which were instead better fit to a three-epoch model ([supplementary table S3, Supplementary Material online](#)). Across all the species best-fit to a two-epoch demographic model,  $\rho$  ranges from  $\sim 0.0685$  to  $\sim 5,933$ , indicating that some bacterial species have experienced a contraction ( $n=10$ ) while others have experienced an expansion ( $n=26$ ). In the three cases where  $\rho$  is extremely large ( $>100$  for the species *Dialister invisus*, *Alistipes shahii*, and *Bacteroides vulgatus*), the likelihood surface is very flat and confidence intervals exceptionally large, indicating substantial uncertainty in such large estimates of  $\rho$  for these species ([supplementary fig. S4, Supplementary Material online](#) and [supplementary table S4, Supplementary Material online](#)). Finally, when converted to years (See Methods), MLEs for the timing of population size changes approximately fall between  $5.0 \times 10^2$  to  $8.6 \times 10^5$  years ago ([supplementary table S4, Supplementary Material online](#)). This range of timescales span many important periods of human history, including industrialization ( $10^2$  years ago), the transition to agriculture ( $10^4$  years ago), and out of Africa migrations ( $10^4$  to  $10^5$  years ago).

Together, these findings suggest that commensal gut microbial populations have complex and diverse demographic histories, with some species experiencing substantial population size changes and others exhibiting more stable trends. Additionally, the approximate historical timescales of inferred demographic changes overlap with key events in human history, raising several interesting questions about how human behaviors may have selected for or against certain gut commensal microbiota, though this is an avenue for future work (See Discussion).

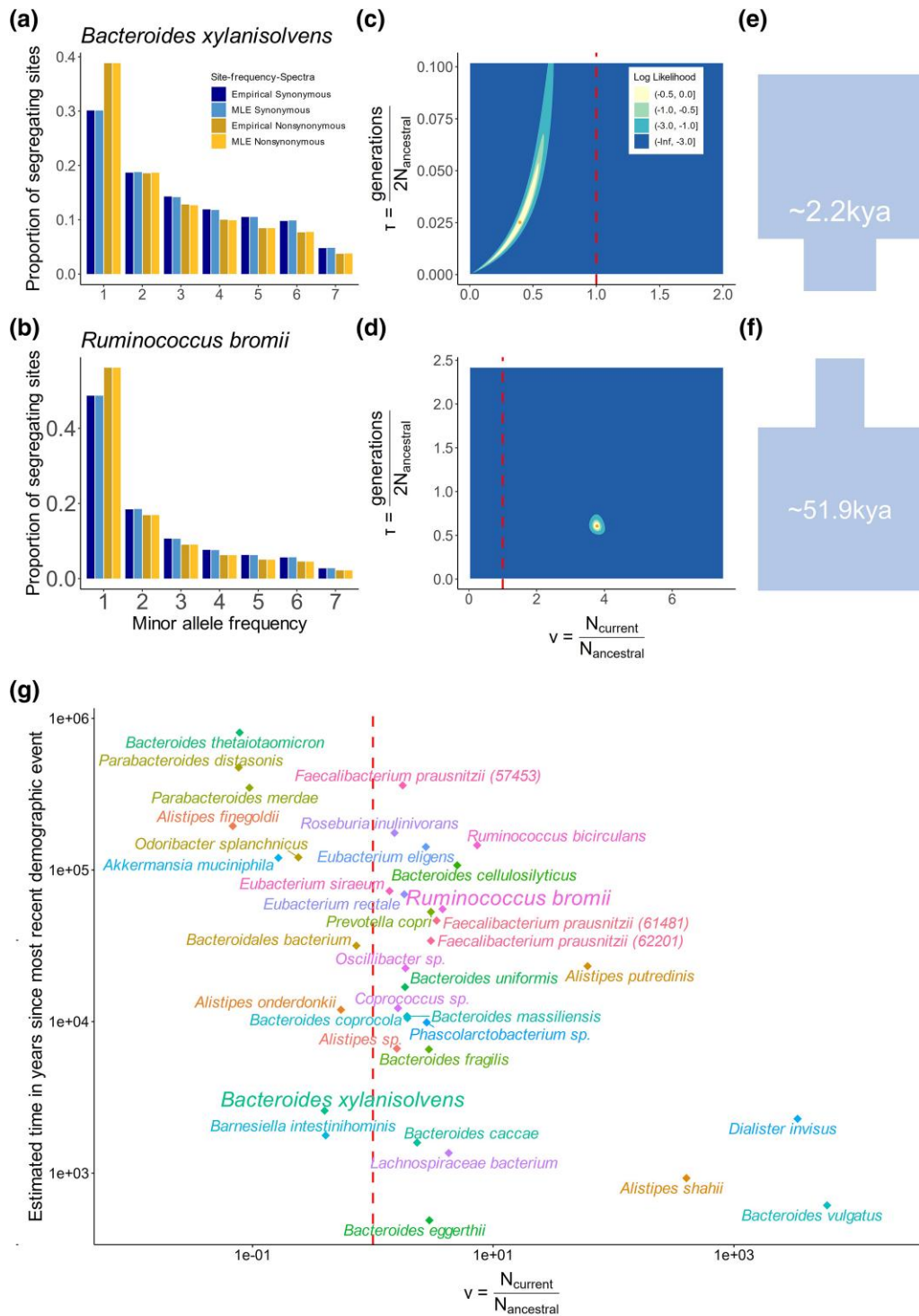
### DFEs of gut Commensal Microbiota Display Phylogenetic Trends

We next turned our attention to inferring the fitness effects of nonsynonymous mutations. As seen in [supplementary fig. S2a and b, Supplementary Material online](#), the observed SFS of synonymous SNPs differs substantially from that of nonsynonymous SNPs, with there being a higher proportion of singletons found among nonsynonymous SNPs compared to putatively neutral synonymous SNPs. This difference suggests the presence of strong purifying selection acting on nonsynonymous mutations (Bustamante et al. 2001; Williamson et al. 2005; Eyre-Walker et al. 2006; Eyre-Walker and Keightley 2007; Boyko et al. 2008) consistent with other bacterial populations (Hughes 2005; Cornejo et al. 2013). As such, we inferred the DFEs of nonsynonymous mutations in the 39 common commensal bacterial species for which we inferred demographic models (Fig. 3).

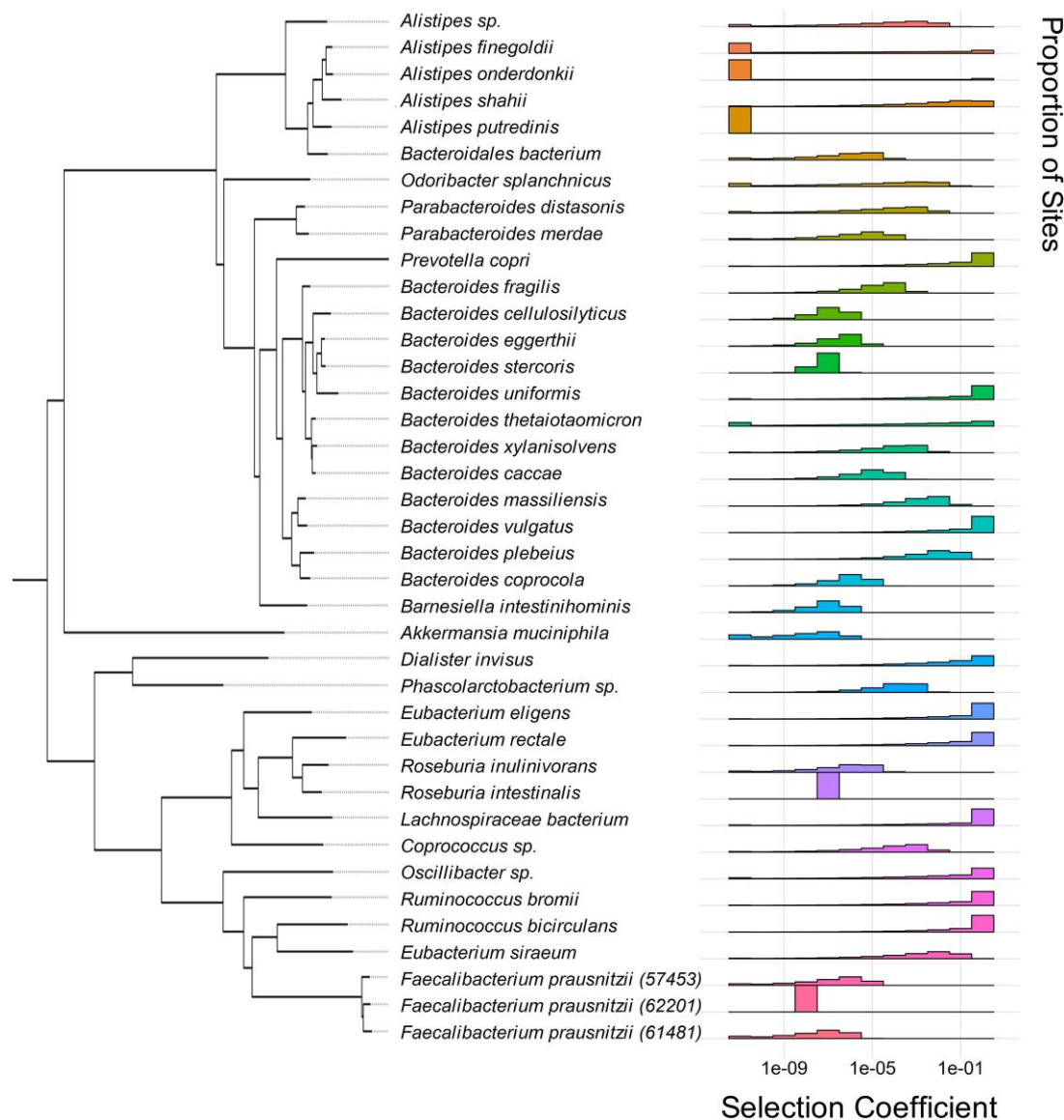
To infer the DFE, we constructed SFSs for nonsynonymous SNPs from the same species for which we constructed SFSs composed of synonymous SNPs. We then used *Fitdad* (Kim et al. 2017) to infer the DFE. Since demographic history can affect the shape or skew of the SFS, *Fitdad* infers selection conditional on the demographic history that was inferred from the synonymous SFS.

We examined the fit of several functional forms of the DFE to the SFS (Kim et al. 2017). One such functional form is a gamma distribution, characterized by *shape* and *scale* parameters that generate flexible distributions. Another such functional form is the neu + gamma distribution, which is a mixture distribution of a gamma distribution and a point-mass on neutrality. The neu + gamma distribution has one additional parameter compared to the gamma distribution, describing the proportion of neutral mutations. The model parameters that produce a nonsynonymous SFS most closely resembling the observed nonsynonymous SFS are considered representative of the underlying evolutionary parameters. We compared the AICs of the gamma-distributed DFE (Fig. 3) with the neu + gamma-distributed DFE ([supplementary fig. S5, Supplementary Material online](#)), and found that there was no significant improvement in fit for the neu + gamma compared to the gamma, except in the case of a single species, *Prevotella copri* ([supplementary table S5, Supplementary Material online](#)). Given a general absence of improved model fit for the neu + gamma model over the gamma model, below we further analyze the gamma-distributed DFEs.

There is considerable variation in the degree of deleterious selection coefficients ( $s$ ) expected for each of the species analyzed. For example, in *R. bromii*, we inferred that 7.0% of nonsynonymous mutations were weakly deleterious or neutral ( $|s| < 10^{-6}$ ), 21.4% were moderately deleterious ( $10^{-6} < |s| < 10^{-2}$ ), 22.5% were highly deleterious ( $10^{-2} < |s| < 0.5$ ), and



**Fig. 2.** Demographic inference of common commensal gut microbes. Site-frequency spectra from the empirical data are compared to those predicted from the best-fit demographic models for two species: a) *Bacteroides xylanisolvens* and b) *Ruminococcus bromii*. “MLE synonymous” shows the expected SFS produced by the ML demographic parameter estimates. “MLE nonsynonymous” shows the expected SFS produced by the ML demographic and selection parameters from a gamma-distributed DFE. Two-dimensional likelihood surfaces of population size, (in units of  $N_{\text{anc}}$ ) and time since the onset of the most recent demographic event, (in units of  $\frac{\text{generations}}{2N_{\text{anc}}}$ ) are shown for c) *B. xylanisolvens*, and d) *R. bromii*. A dashed red line separates the parameter space between contractions and expansions. The maximum-likelihood demographic parameters are shown with an orange dot, and colored intervals denote the decrease in log-likelihood from the MLE. The light cyan regions (LL—3) correspond to the asymptotic 95% confidence interval from the log-likelihood surface. See [supplementary fig. S3, Supplementary Material](#) online for similar plots for all species analyzed. Cartoon schematics (approximately to scale) describing the inferred demography of e) *B. xylanisolvens*, and f) *R. bromii*. g) A summarizing two-dimensional scatterplot of  $v$  and approximate time in years for ML demographic parameters for the 36 of the 39 species analyzed in this paper which were best-fit to a two-epoch demographic model. There are three species named *Faecalibacterium prausnitzii*, and so their distinct species IDs as per the MIDAS (Nayfach et al. 2016) database used to map read are indicated in parentheses.



**Fig. 3.** Phylogenetically sorted distributions of fitness effects. DFEs were estimated for each species assuming a gamma distribution. For each species, the x-axis denotes the  $\log_{10}$  scaled discrete bin of selective effect, while the y-axis denotes the proportion of nonsynonymous mutations which fall into each bin.

49% were lethal ( $|\text{sl}| \geq 0.5$ ). By contrast, in *B. fragilis*, we inferred that 18.4% of mutations were weakly deleterious or neutral ( $|\text{sl}| < 10^{-6}$ ), 81.6% were moderately deleterious ( $10^{-6} < |\text{sl}| < 10^{-2}$ ), and no mutations were found to be highly deleterious ( $10^{-2} < |\text{sl}| < 0.5$ ) or lethal ( $|\text{sl}| \geq 0.5$ ).

We phylogenetically ordered the 39 inferred gamma-distributed DFEs (Fig. 3) and neu + gamma-distributed DFEs (See Methods; [supplementary fig. S5, Supplementary Material](#) online). This visualization revealed some evidence of phylogenetic trends in the DFE of the gut microbiome. For example, both species in the *Ruminococcus* genus were inferred to have DFEs skewed toward highly deleterious and lethal mutations (on average,  $\sim 76.4\%$  of new mutations are inferred to have  $|\text{sl}| > 10^{-2}$ ). By contrast, many of the species in the *Bacteroides* genus have DFEs skewed toward weakly or moderately deleterious mutations (on average,  $\sim 82.5\%$  of sites are inferred with  $|\text{sl}| < 10^{-2}$ ).

To quantify whether gamma-distributed DFEs between all pairs of the 39 species in our dataset are distinct versus the same, we implemented a likelihood ratio test (LRT) similar

to that in [Huber et al. \(2017\)](#) (See Methods). Specifically, we compared the log-likelihood of two models: (1) a full model where pairs of species have independently inferred *shape* and *scale* parameters and (2) a constrained (null) model in which pairs of species have the same *shape* and *scale* parameters.

Using this framework, we found that the DFE is significantly different in 334 out of 741 ( $\sim 45\%$ ) pairs of species, after Bonferroni correction. Comparisons of DFEs between pairs of species belonging to the same genera versus different genera reveal some evidence of phylogenetic congruence of the DFE among closely related species. Concretely, of 334 significantly different DFEs, 284 ( $\sim 84\%$ ) of such tests were between species of different genera, while the remaining 50 ( $\sim 16\%$ ) were between species within the same genera, suggesting that differences in the DFE often arise between species that are more distantly related.

Supporting the observation that DFEs may be more different among distantly related species, we found that LRT statistics for between-genera comparisons are significantly elevated compared to within-genera comparisons ( $P = 0.008$ ,

two-sided Wilcoxon rank-sum test, [supplementary fig. S6, Supplementary Material](#) online). After permuting species labels (See Methods), there is no longer a significant difference. We also found that the difference in mean selection coefficient ( $E[s]$ ) is significantly higher for between-genera comparisons versus within-genera comparisons, suggesting that  $s$  truly differs rather than there being a difference in statistical power ([supplementary fig. S6, Supplementary Material](#) online).

To more formally test whether the DFE parameters covary with phylogeny following Brownian motion, we estimated Pagel's  $\lambda$  (See Methods). We considered four parameters related to the DFE: the proportion of neutral mutations (i.e.  $|s| < 10^{-6}$ ), the mean selection coefficient ( $E[s]$ ), the ancestral effective population size ( $N_{Anc}$ ), and the *shape* parameter of the best-fit gamma-distributed DFE. These analyses yielded some evidence of phylogenetic signal for the *shape* parameter (Pagel's  $\lambda \sim 1.032$ ,  $P = 4.9 \times 10^{-10}$ ); however, this phylogenetic signal is supported by only two outlier species: *F. prausnitzii* (62201), and *R. intestinalis*. Removing these two species from our dataset resulted in no evidence of phylogenetic signal (Pagel's  $\lambda = 7.17 \times 10^{-5}$ ,  $P = 1.0$ ). Thus, while there are greater differences in the DFE between species more distantly related compared to those more closely related, more generally, the DFE does not seem to have evolved via Brownian motion across the entire phylogeny.

We also considered if there are phylogenetic trends for DFEs when the selection coefficient,  $s$ , is scaled by  $2N_{Anc}$  ([supplementary fig. S7a, Supplementary Material](#) online) (Huber et al. 2017; Kim et al. 2017; Huang et al. 2021).  $2N_{Anc}s$  provides insight into how selection and drift act together, whereas  $s$  provides insight into the relative fitness of individual mutations. Thus, we next compared whether the distribution of  $2N_{Anc}s$  differs across species ([supplementary fig. S7b, Supplementary Material](#) online). We found that the LRT comparing pairs of DFEs is significant in 458 instances (~62%), after Bonferroni correction ([supplementary fig. S7b, Supplementary Material](#) online). Of these 458 significantly different DFEs, 408 (~89%) were between species of different genera, consistent with the results above performed with  $s$  showing that DFEs are more similar among closely related species. However, similar to our analyses of DFEs parameterized with  $s$ , Pagel's  $\lambda$  yields no evidence that the population-scale DFE covaries with phylogeny.

### Demographic and DFE Inference is Robust to the Removal of Low-Recombination Regions and Selective Sweeps

One of the main assumptions of SFS-based analyses is free recombination between sites. Given that rates of recombination can vary across the genome, we asked whether excluding regions of low recombination would result in different demographic and DFE inferences. To do so, we leveraged recent estimates of recombination rates (Liu and Good 2024) inferred from the same datasets analyzed in our study. For 18 species for which there were available recombination maps, we removed sites from the SFS found in genomic regions with recombination rates less than the inferred genome-wide median rate (See Methods).

Unaccounted-for selective sweeps may also confound inferences made from the SFS, as linked selection can result in skews in the SFS (Braverman et al. 1995; Fay and Wu 2000). In the 18 species for which we controlled for recombination, we excluded sites inferred to be undergoing a selective

sweep (Wolff and Garud 2023), as well as 1,000 base pairs flanking the selective sweeps. These sweeps were inferred from the same datasets analyzed in this study. After removing low-recombination regions and selective sweeps, on average, the “filtered SFSs” retained ~47.2% of SNPs compared to the “full SFS”.

Figure 4 shows the comparison of demographic and DFE parameters inferred with the full versus filtered SFSs ([supplementary tables S8 and S9, Supplementary Material](#) online). The parameters  $\lambda$ ,  $E[s]$ , *shape*, and *scale* inferred from both SFSs are correlated ( $R^2 = 0.27, 0.27, 0.41, 0.43, 0.60$ , respectively); however, one parameter,  $N_{Anc}$ , has a lower  $R^2$  of 0.20. This lower  $R^2$  is largely driven by the ~5 species with the largest  $N_{Anc}$ , whereby  $N_{Anc}$  in the filtered SFS is inferred to be smaller than in the full SFS. The reduction in  $N_{Anc}$  among these species could be due to recombination acting more efficiently in larger populations. In recombining bacterial populations, diversity may be reduced in regions where recombined fragments are transferred from one genome to another (Liu and Good 2024). Understanding the full impact of recombination on SFS inferences in greater detail is an avenue for future research. Despite this, the high correlation of inferred parameters with and without control for low recombination and selective sweeps suggests that demography and DFE inferences are largely robust to violations of independence in the SFS.

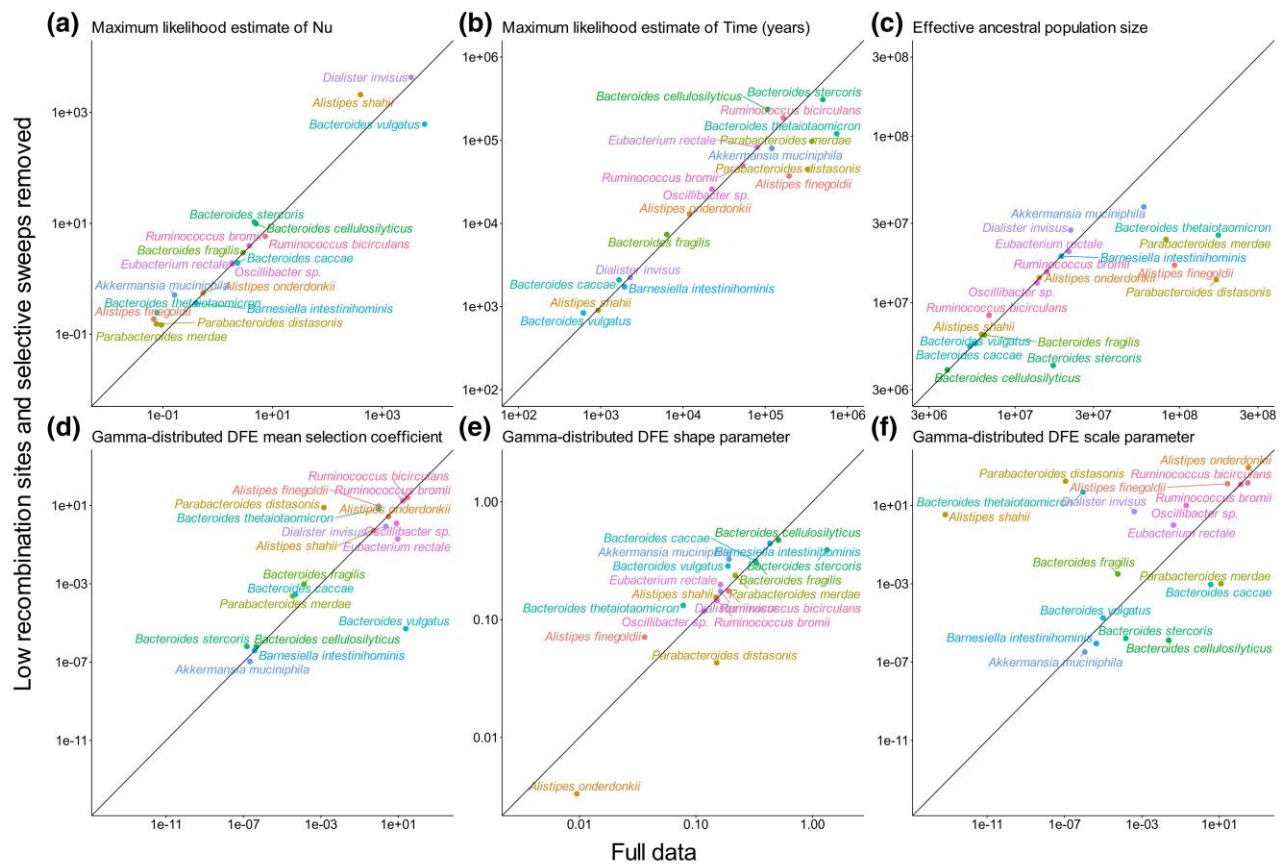
### Differences in the DFE Between Core Versus Accessory Genes are Largely Driven by Drift

Thus far, our analyses only have considered core genes present in  $\geq 95\%$  of samples. Next, we evaluated whether accessory genes present in 30% to 70% of samples have different demographic histories or DFEs from core genes. To ensure that at least 14 genomes could be sampled even for the most rare genes considered (30% prevalence), we analyzed the 18 out of 39 species with (1) at least 47 QP samples (Fig. 1a) and (2) with models of demography and selection inferred from the accessory genome that adequately fit the SFS (Fig. 5c, [supplementary fig. S7, Supplementary Material](#) online).

By visual inspection, the synonymous SFSs constructed from the accessory genes typically exhibit a proportional depletion of rare variants relative to the SFSs constructed from core genes of the same species (Fig. 5a and b and [supplementary fig. S8, Supplementary Material](#) online). This signature is consistent with possible contractions in effective population size due to demography and/or reduced linked selection.

We first investigated if accessory versus core genes have distinct demographic histories. For 17 of the 18 species analyzed, we found that two-epoch demographic models best fit the data ([supplementary table S6, Supplementary Material](#) online). Moreover, we found that on average,  $\lambda$  was depressed for accessory genes compared to core genes, suggesting stronger population contractions and/or weaker population expansions in the accessory genome relative to the core genome ([supplementary table S6, Supplementary Material](#) online, [supplementary fig. S9, Supplementary Material](#) online). Additionally, for four of the species,  $N_{Anc}$  was found to be higher for accessory genes, while for nine of the species,  $N_{Anc}$  was found to be higher for core genes, suggesting overall differences in effective population sizes, and differing impacts of drift on core versus accessory genes.

Next we inferred the DFE for accessory genes ([supplementary table S7, Supplementary Material](#) online). Like core genes,



**Fig. 4.** Comparison of evolutionary parameters inferred from the full SFS (x-axis) versus after low recombination and selective sweeps are removed (filtered SFS) (y-axis). The solid black-line represents the  $y = x$  trend for each panel. a) MLE of  $\nu$ ; b) MLE of time; c) MLE of the ancestral effective population size; d) MLE of the mean selection coefficient of Gamma-distributed DFE; e) MLE of the *Shape* parameter of Gamma-distributed DFE; f) MLE of the *Scale* parameter of Gamma-distributed DFE.

accessory genes frequently harbor an excess of rare nonsynonymous variants compared to rare synonymous variants. However, core genes have a greater excess of rare nonsynonymous variants compared to rare synonymous variants suggesting that the DFEs of core genes may be more deleterious (Fig. 5a and b and supplementary fig. S8, Supplementary Material online). To test for a systematic difference in the DFE between core and accessory genes, we performed a LRT (See Methods) of similarity of DFEs inferred from accessory and core genes at two levels of selection: individual selective effects ( $s$ ) and population-scaled selective effects ( $2N_{Anc}s$ ). We found that the DFE parameterized by  $s$  was significantly different in six species, with core genes being more deleterious in four species, and accessory genes being more deleterious in two species. All of these six species also had an even greater significant difference in  $2N_{Anc}s$ . Specifically, when comparing the DFE between accessory and core genes with  $2N_{Anc}s$ , we found that 10 species had a statistically significant difference in  $2N_{Anc}s$  between accessory and core genes with six species having a more deleterious  $2N_{Anc}s$  in the core genome and four species having a more deleterious  $2N_{Anc}s$  in the accessory genome.

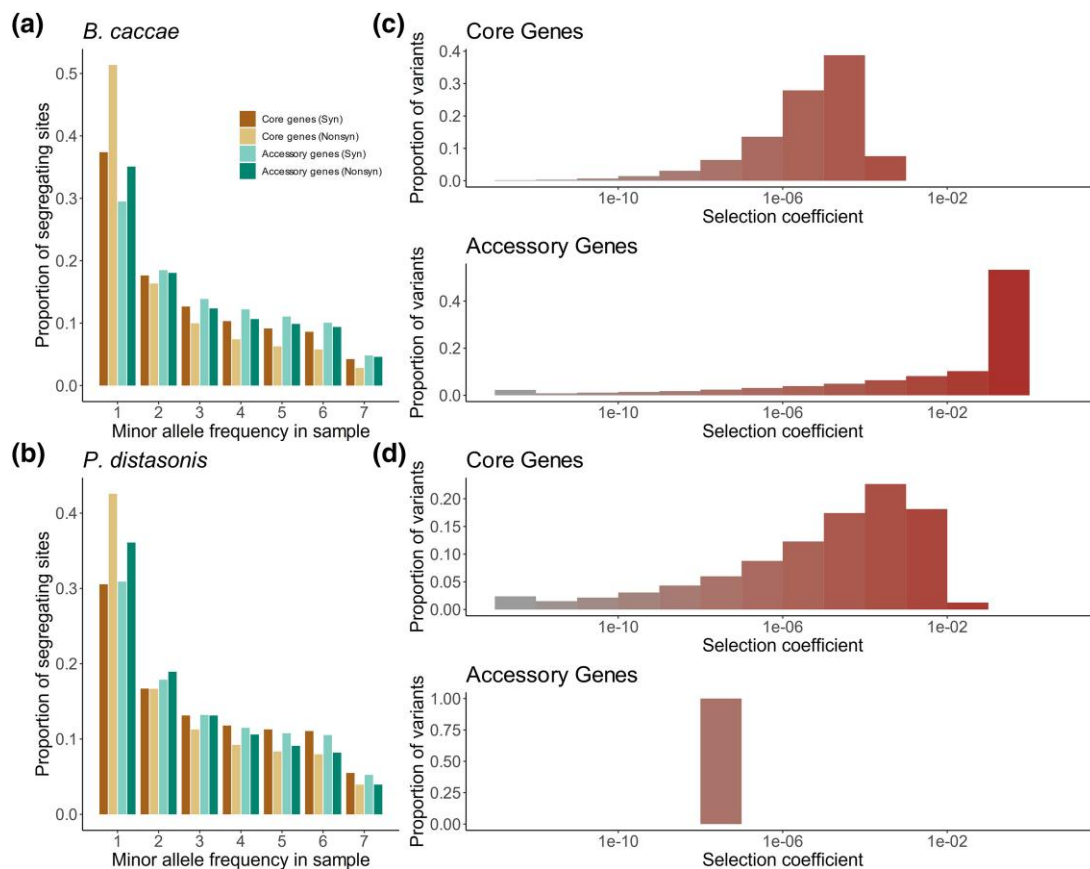
For the six cases where  $2N_{Anc}s$  was higher in the core genome,  $N_{Anc}$  for the core genome was also inferred to be greater than or equal to  $N_{Anc}$  for the accessory genome. Similarly, for the four cases where  $2N_{Anc}s$  was higher in the accessory genome,  $N_{Anc}$  for the accessory genome was also inferred to be greater than or equal for the core genome. This suggests that  $N_{Anc}$  may be driving differences in the DFE at the resolution

of population-scale selection; however, differences in selection also arise at the level of individual mutations ( $s$ ), suggesting that while drift is likely a driving force for differences in the DFE between core and accessory genes, selection still plays an important role.

## Discussion

In this study, we infer the population demographic histories and DFEs of new mutations of almost 40 prevalent human commensal gut bacteria. To do so, we used a SFS-based approach that is commonly used in eukaryotes (Gutenkunst et al. 2009; Beichman et al. 2018) but has been rarely used for bacterial populations (Cornejo et al. 2013; Pepperell et al. 2013; Mortimer et al. 2017). We found that commensal microbiota have experienced a range of effective population size changes over the course of human history. Additionally, we inferred finer distributions of fitness effects compared to previous work, which estimated the proportion of sites experiencing a single selection coefficient rather than a range of coefficients (Cornejo et al. 2013; Garud et al. 2019). Finally, we found that closely related species exhibit more similar DFEs than those from different genera, and that differences in the DFE between core versus accessory genes may largely be driven by differences in effective population size.

In our inferences, we assume that the SFS for synonymous SNPs accurately reflects demographic history. However, a myriad of evolutionary processes can affect the SFS, and thus have potential confounding effects (O’Fallon 2010;



**Fig. 5.** Differences in selective effects between core and accessory genes. Empirical site-frequency spectra of synonymous and nonsynonymous SNPs from core genes and accessory genes for two species: a) *Bacteroides caccae* and b) *Parabacteroides distasonis*. Gamma-distributed selection coefficients (s) of new mutations are shown for c) *Bacteroides caccae* and d) *Parabacteroides distasonis*. For each species, the x-axis denotes the log<sub>10</sub> scaled discrete bin of selective effect, while the y-axis denotes the proportion of mutations which fall into each bin. Bins are color coded from gray to dark red according to the inferred lethality on the x-axis. See [supplementary fig. S10, Supplementary Material](#) online for similar plots for all species analyzed.

Lapierre et al. 2017). One such process is purifying selection acting on synonymous mutations, which can reduce diversity and skew the SFS toward rare variants (Chen et al. 2017; Johri et al. 2021; Murphy et al. 2022), potentially giving rise to false inferences of expansion. However, recently we found via a simulation study (Martinez I Zurita et al. 2024) that only extreme scenarios of strong negative selection ( $s = -0.001$ ) acting on all synonymous mutations results in false inferences of severe ( $>100\times$ ) expansions. When fewer synonymous mutations are subjected to selection, the falsely-inferred expansions are smaller in magnitude than most of the expansions inferred in our study. Further, the fact that 10 out of 39 species we analyzed underwent population contractions suggests widespread selection on synonymous mutations is not driving the demographic signals we uncover.

Another evolutionary force that can create confounding signatures is the presence of selective sweeps, which can also generate an excess of rare variants (Braverman et al. 1995; Fay and Wu 2000). Additionally, while recombination seems to be common and frequent in human gut microbiota, local dips in recombination rates along the genome could result in linked effects between neutral and non-neutral sites. To our knowledge, here we perform the first systematic empirical investigation across multiple species on the impact of recombination and selective sweeps on demographic inference. The finding that demographic histories and DFEs are robust to the presence of selective sweeps as well as inclusion of heterogeneous recombining regions is important beyond our study,

as SFS-based analyses are frequently employed across the tree of life without full knowledge of potential confounders.

Finally, false population contractions may be inferred if there is some force that inflates intermediate frequency synonymous variants, such as balancing selection or cryptic population substructure (Voight and Pritchard 2005; Lapierre et al. 2017). Balancing selection has been previously inferred using Tajima's  $D$  to act on a very small number of genes but not genome-wide (Moeller 2021). Thus, it is highly unlikely that the signals observed here for population contractions stem from balancing selection as we use genome-wide polymorphism data. While we controlled for population substructure (supplementary fig. S1, Supplementary Material online, Methods), both by removing closely related lineages and also by analyzing clades that display evidence of recombination among member lineages, we acknowledge there still could be residual cryptic population structure confounding analyses. Future work, potentially leveraging orthogonal signals like haplotype sharing, may be useful for uncovering different features of the demographic history and population structure that our current approach misses.

Despite their simplicity, basic aspects of our inferred demographic models concur with previous estimates of  $N_{\text{Curr}}$  made in the literature (supplementary fig. S11, Supplementary Material online). We found that MLEs for  $N_{\text{Curr}}$  typically fall between  $10^6$  to  $10^{10}$ , in agreement with a previous study that estimated  $N_{\text{Curr}}$  using dN/dS to be in the range of  $10^6$  to  $10^9$  for  $>150$  commensal, pathogenic, and free-living bacterial species

(Bobay and Ochman 2018). Three of the species that we analyzed have  $N_{\text{Curr}}$  on the order of  $>10^9$ , but the likelihood surfaces for these species are extremely flat, yielding almost equivalent log likelihoods for a wide range of values as low as  $10^8$  (supplementary table S4, Supplementary Material online, supplementary fig. S3, Supplementary Material online).

It is striking that many of the demographic changes we infer coincide with key events in human history, such as the industrial revolution, the transition to agriculture, and human migration out of Africa. Even with a lower assumed mutation rate of  $10^{-10}$  per base pair per generation (Barrick and Lenski 2013) (See Methods), our estimate of time in years still would span these significant human epochs. It is important to note that when setting up our analyses, we did not specify any particular time frame for the events to occur—rather, that the majority of demographic events occurred between  $10^3$  to  $10^6$  years ago is the outcome of our inferences. Specifically, we searched the parameter space of demographic events ranging from the present day to a maximum of  $0.15 \times 2N_{\text{Anc}}$  generations ago (corresponding to a maximum of  $\sim 7.5$  million years ago on average), and could have inferred population size change events at any point in that time frame.

It is tempting to speculate that events like the agricultural revolution may have been causal of changes in effective population sizes among gut commensals. For example, Cornejo et al. inferred a population expansion for the cavity-causing pathogen *Streptococcus mutans*  $\sim 10,000$  to 20,000 years ago, potentially due to new available ecological niches opening up as a consequence of changes in diet during the agricultural revolution (2013). While the direction of inferred population size changes concur with changes in relative abundances for some species in industrialized versus traditional populations (e.g. we infer expansions for eight of ten *Bacteroides* species that have been previously shown to be higher in abundance in industrialized populations due to dietary shifts (Sonnenburg and Sonnenburg 2019)), this is not universally true. For example, we inferred a population expansion for *Prevotella copri* despite it being at markedly lower prevalence and abundance in industrialized versus traditional populations due to reductions in plant-based carbohydrates in western diets (Wu et al. 2011; Sonnenburg et al. 2016; Tett et al. 2019). It is possible that more sophisticated models can provide more biological insight into the population size changes experienced by gut microbiota as our models are necessarily simple and likely do not reflect the full set of demographic changes that have transpired for each species. The resolution of our ability to distinguish more sophisticated models is limited, however, by our sample size. For example, our investigation into two-epoch versus three-epoch models generally resulted in very similar AICs and LLs, making it difficult to distinguish between the simple models we inferred versus more complex models (supplementary fig. S2, Supplementary Material online). One-epoch models (models lacking any population size change), however, tended to have significantly worse AICs and thus were easier to reject. Future research leveraging larger sample sizes and possibly also time series data will be needed to infer more sophisticated models for human gut commensals and to make more explicit connections to human history and human health.

Beyond inferring a range of demographic histories across species, we also found differences in their DFEs. Interestingly, we found that phylogenetically related species have more similar DFEs than distantly related species

(supplementary fig. S6, Supplementary Material online, Methods) potentially implying functional similarities among closely related species. We note that both differences in  $s$  and  $N_{\text{Anc}}$  could be driving differences in the DFE across species, though we find a greater number of differences between pairs of species when comparing the DFE in terms of  $2N_{\text{Anc}s}$  (458 pairs) rather than  $s$  (334 pairs), suggesting that variation in ancestral population size plays a major role in driving differences in the DFE across species. Similarly, the differences in the DFEs for core versus accessory genes are less likely due to differences in  $s$  and instead largely due to differences in  $N_{\text{Anc}}$ .

Finally, it is worth noting that our DFE and demographic history inferences rely on genotypes supported by a majority of reads within hosts (See Methods), mitigating the impact of sequencing errors on our inferences. Consequently, our inferred DFEs represent what could be considered an “average across hosts and may not fully capture within-host DFE variations, which can potentially differ among individuals and environments (Dapa et al. 2023). Future research utilizing experimental evolution techniques (Shiver et al. 2023; Wong and Good 2024) will be pivotal in understanding differences between within-host and across-host DFEs.

In summary, this work advances our understanding of the evolutionary dynamics of human gut microbiota, highlighting the influence of demographic history and selective forces on genetic variation. Our findings underscore the complexity of these interactions and their implications for the human microbiome, which future work will undoubtedly further elucidate.

## Methods

### Data

Raw whole-genome shotgun sequencing reads for North American metagenomic samples analyzed in this study were downloaded from the Human Microbiome Project Consortium (Human Microbiome Consortium 2012; Lloyd-Price et al. 2017). Accession numbers associated with this data are in supplementary table S1, Supplementary Material online.

This dataset included 471 fecal samples from 250 healthy individuals. A subset of these 250 individuals were sampled at two or three time points. We analyzed data from a single time point per host per species in the North American dataset. Additionally, since previous work has shown that technical and sample replicates of the same fecal sample have little genomic variability (Nayfach et al. 2016; Maghini et al. 2024), we merged FASTA files for technical and sample replicates from the same time point prior to calling species, SNPs and CNVs to increase coverage as in Garud et al. (2019).

For European gut microbiomes, there were 250 fecal samples available from 125 pairs of human adult twins (Xie et al. 2016). We analyzed data from a single randomly selected twin to better approximate a random sampling of the population. Additionally, we analyzed one family member chosen at random from Korpela et al. (2018).

For Chinese gut microbiomes (Qin et al. 2010), we analyzed 185 fecal samples from 185 healthy individuals.

Relevant data, including accession numbers are in supplementary table S1, Supplementary Material online.

### Identification of Bacterial Species, Genes, and SNPs

We used a standard reference-based approach called MIDAS to quantify bacterial species abundances and gene and SNP content (Nayfach et al. 2016).

The first step of the MIDAS pipeline is to determine the species present in each set of sample(s) for each host. MIDAS quantifies the relative abundance of species in a given sample by mapping sequencing reads to a reference database of single-copy “marker” genes unique to each species. We used database version 1.2 (Nayfach et al. 2016), downloaded on November 21st, 2016. A species was considered present if it had an average marker gene coverage of at least  $3 \times$  for a given sample.

Second, MIDAS quantifies SNPs for each species for each sample. Similar to the “species” step, this “SNP” step leverages a standard reference-based approach in which reads are mapped to a single genome per species. To avoid reads mapping spuriously to species not present in the dataset, reads were mapped only to species truly present in the sample as per the species step defined above. Specifically, the reference panel consisted of the union of species found in a host at any time point, as per Garud et al. (2019). Mapping was done with Bowtie2 (Langmead and Salzberg 2012) with the following default MIDAS mapping thresholds: global alignment, MAPID  $\geq 94.0\%$ , READQ  $\geq 20$ , ALN\_COV  $\geq 0.75$ , and MAP  $\geq 20$ .

MIDAS annotates SNPs as either 1D or 4D, indicating codon degeneracy. For example, 1D means that any nucleotide change will result in an amino acid change, resulting in a non-synonymous site. By contrast, 4D means that any nucleotide change will result in a synonymous site. When constructing synonymous SFSs, only 4D sites were considered, and when constructing nonsynonymous SFSs, only 1D sites were considered.

Finally, MIDAS identifies gene CNVs by aligning reads to a pangenome, i.e. the union of all genes found in any strain present in the MIDAS database for a given species. Once again, reads were mapped to pangenomes only for those species considered to be present. Standard MIDAS read coverage parameters were used for calling CNVs and as described in Garud et al. (2019).

Since there may be orthologs of the same gene present in multiple species, these genes can potentially result in spurious read donating and recruiting. Therefore, any gene with  $\geq 95\%$  ANI between different species’ pangenomes was excluded from further analysis. Genes with  $\geq 95\%$  ANI with another gene in multiple species’ pangenomes were previously identified in Garud et al. (2019).

To identify which genes to include in our analysis, we considered the copy number of each gene. Genes were included in our data set if their estimated copy number was  $\geq 0.3$  and  $\leq 3.0$  (Garud et al. 2019), to avoid including genes with insufficient read support and genes with abnormally elevated copy numbers, potentially due to spurious read mapping. Additionally, any gene with copy number values  $> 3.0$  in any host was excluded from the entire dataset to once again control against spurious read mapping. Next, to differentiate between core and accessory genes, we considered the prevalence of each gene: core genes were identified as genes with prevalence  $\geq 95\%$ , while accessory genes were identified as genes with prevalence between 30% and 70%.

Once core and accessory genes were identified, sites were included using a minimum sample depth of  $20 \times$  coverage, and a minimum site depth of 20 reads.

## Quasi-Phasing

To estimate quasi-phased haplotypes from individual hosts, we identified hosts in which the lineage structure of their

bacteria was simple enough to assign alleles to the dominant lineage with high confidence as described in Garud et al. (2019). Since both our paper and Garud and Good et al. utilize the same data for analyses, we used the same sample  $\times$  species identified in Garud and Good et al. that were classified as quasi-phaseable (2019). We then estimated the genotype of the dominant lineage by assigning the major allele of each site only if it had a frequency of 0.8 or greater. Any site with a major allele frequency between 0.5 and 0.8 was considered missing data.

## Phylogenetic Ordering

To order species phylogenetically (such as in Fig. 3), we used the phylogenetic tree from Nayfach et al. (2016). We abstracted a subtree of relevant species using the APE package (version 5.8) in R (Paradis and Schliep 2019).

## Calculating the SFS

We compute “folded” site-frequency spectra (SFS) for both synonymous and nonsynonymous sites from quasi-phased bacterial genomes. The folded SFS describes the distribution of minor allele frequencies across the genome. “Folded” indicates that since the ancestral versus derived state of alleles are unknown, any allele frequency ( $f$ ) that is  $> 50\%$  becomes  $1-f$ . Folded SFSs have been shown to yield accurate inferences of the DFEs of deleterious mutations (Keightley and Eyre-Walker 2007; Boyko et al. 2008; Kim et al. 2017).

The SFS can be represented as a vector,  $X = [X_1, X_2, \dots, X_{n-1}]$ , in which each element of the vector describes the number of SNPs at frequency  $i$  given  $n$  chromosomes. In other words,  $X_1$  is the number of singletons,  $X_2$  is the number of doubletons, etc. Thus, the SFS describes the number of QP samples a SNP appears in. A site is considered to be variable if at least one quasi-phased genome has a different allele from other quasi-phased genomes in our dataset.

## Projection of the SFS

The empirical SFS may show irregular peaks and valleys due to missing data (supplementary fig. S1b, Supplementary Material online), which may confound inference from and modeling of the SFS, as some sites are not called for all individuals. To diminish the confounding effects of missing data, we projected the empirical SFS down from a larger sample size to a smaller sample size of 14 haplotypes for all species. In other words, for all sites present in  $\geq 14$  hosts, we computed the expected frequency of each SNP in a sample of 14 haplotypes using a hypergeometric distribution (Gutenkunst et al. 2009). Sites with data in  $< 14$  hosts were omitted.

## Controlling for Recombination

A major assumption of SFS-based analyses is free recombination between sites. To control for low recombination regions, we inferred recombination maps using recent work from Liu and Good (2024). Liu and Good provide the first estimate to date of recombination rates across bacterial genomes for the 29 most prevalent and abundant species present in Westernized gut microbiomes using the same datasets analyzed in this paper (2024). To generate this recombination map, they inferred recombination transfer events between pairs of strains residing within the same host. To convert these transfer events to a rate of recombination per site, we computed the mean number of detected transfer events occurring

in 1,000-bp windows across the core genome using any pair of strains with a transfer event in that region. We then removed sites from the SFS found in regions with a mean recombination rate less than the median recombination rate for each of the 18 species overlapping our dataset.

### Controlling for Selective Sweeps

Unaccounted-for selective sweeps and patterns of linked selection may confound inferences made from SFSs, as evolutionary signals from these forces may be incorrectly attributed to population demographic history. To control for selective sweeps, we masked regions of the genome that we previously inferred using the same dataset to have undergone selective sweeps (Wolff and Garud 2023). We also masked 1,000 bps flanking the selective sweeps.

### Inference of Demographic Models

We consider three model specifications: (1) a one-epoch model consisting of a constant effective population size, (2) a two-epoch model consisting of two epochs, i.e. two periods of time, separated by one effective population size change, and (3) a three-epoch model consisting of three epochs separated by two effective population size changes (Fig. 1c).

We use the statistical inference package, *dadadi* (Gutenkunst et al. 2009) to infer two main demographic parameters for each effective population size change,  $\theta$  and  $\phi$ . The time of the population size change is represented by  $t$ , in units of  $\frac{\text{generations}}{2N_{Anc}}$ , where  $N_{Anc}$  is the ancestral effective population size. The parameter  $\phi$  denotes the ratio of effective population size after the population size change relative to  $N_{Anc}$ .

To find the maximum-likelihood parameter estimates, we defined a grid of points over a likelihood surface parameterized by  $\theta$  and  $\phi$ . For a given point on the likelihood surface, *dadadi* estimates the model spectrum expected under those evolutionary parameters using diffusion theory. The multinomial log-likelihood function is then used to assess the fit of the model spectrum to the empirical data as described in Gutenkunst et al. (2009). We then performed a gradient-descent based approach to evaluate points along the likelihood surface to find the MLEs. The pair of parameters which yielded the highest log-likelihood were assumed to represent the underlying evolutionary history. We did not force parameter bounds on the likelihood surface, allowing the gradient-descent search to evaluate the entirety of the log-likelihood surface. In addition, at least 25 initial parameter guesses, spanning a large range across the parameter surface, were evaluated to ensure that the likelihood surface did not converge at a local maximum specific to a set of starting parameters. We used the chi-square approximation to the log-likelihood ratio to obtain the critical values for the asymptotic 95% confidence intervals of the demographic parameters. The CIs included all parameter values within three log-likelihood units ( $df = 2$ , reflecting the two parameters) from the MLE.

### Estimation of Ancestral Population Size

We compute  $N_{Anc}$  from  $s$  as:

$$N_{Anc} = \frac{s}{4\mu L_{syn}} \quad (1)$$

where  $\mu$  is the mutation rate per site per generation and  $L_{syn}$  is the number of synonymous sites.  $s$  is the population-scaled mutation rate, optimized as the scaling factor between the

empirical data and the model spectrum that best fits the data.  $\mu$  is assumed to be  $4.08 \times 10^{-10}$  per site per generation, based on experimental estimation for neutral evolution in bacteria (Drake 1991).

### Computation of Estimated Times of Demographic Size Changes

We convert the time parameter,  $t$ , which is in units of  $\frac{\text{generations}}{2N_{Anc}}$ , into units of time in years. We chose to assume a conservative generation time of one generation per day, based on estimates of bacterial growth (Gibbons and Kapsimalis 1967; Savageau 1983; Ghosh and Good 2022). Thus, we compute:

$$\text{Time (in years)} = \frac{2N_{Anc}}{365} \quad (2)$$

### Inference of the DFEs of Nonsynonymous Mutations

To infer the distribution of fitness effects (DFE), we utilize the SFS of nonsynonymous variants. We use a statistical inference package, *Fitdadi* (Kim et al. 2017). *Fitdadi* uses a maximum-likelihood approach to infer the DFE from site-frequency spectra. We note that  $s$  in *Fitdadi* refers to the change in relative fitness of the heterozygous genotype and is equivalent to the fitness in the haploid.

We examined two different probability distributions for the DFE: a gamma-distributed DFE, and a neutral+gamma-distributed DFE. Under a gamma-distributed DFE, the selective effect of a given mutation, denoted as  $s$ , is randomly drawn from a gamma distribution, parameterized by *shape*, denoted as  $\alpha$  and *scale*, denoted as  $\beta$ . Under a neutral+gamma-distributed DFE,  $s$  is drawn from a gamma distribution, but a proportion of sites is assumed to be neutral, generating a point-mass at  $s = 0$ , parameterized as  $p_{neu}$ .

The parameters for the DFEs described above were inferred using *Fitdadi*. Importantly, because demography can impact the SFS, we condition on the demographic models described above that we inferred from synonymous variants. *Fitdadi* models the expected nonsynonymous SFS for a given combination of DFE parameters and a demographic history. Then, the fit to the data is computed using a Poisson log-likelihood function. In contrast with the multinomial likelihood, the Poisson likelihood additionally requires an a priori assumption about the population-scaled mutation rate of nonsynonymous mutations, denoted as  $\theta_{ns}$ .  $\theta_{ns}$  is obtained by multiplying the best-fitting  $\theta_s$  from demographic inference by the expected ratio of nonsynonymous to synonymous mutation rates. We assumed this ratio to be  $\theta_{ns}/\theta_s = 2.31$  (Huber et al. 2017). We found the DFE parameters that maximized the log-likelihood by gradient descent over a two-parameter likelihood surface parameterized by  $\alpha$  and  $\beta$ . Similar to inference of demography, we did not force parameter bounds on the likelihood surface, allowing the gradient-descent search to evaluate the entirety of the log-likelihood surface. Furthermore, at least 25 initial parameter guesses, spanning a large range across the parameter surface, were evaluated to ensure that the likelihood surface did not converge at a local maximum specific to a set of starting parameters.

### Comparing DFEs Across Species

To quantify the similarities and differences between the DFEs of two species, we performed a LRT comparing the fit of models of the DFEs of nonsynonymous mutations assuming that

the *shape* and *scale* parameters are the same in both species (the null model) versus a model where each species can have its own *shape* and *scale* parameters (the alternative model).

Consider two species, *A*, and *B*, for which we have independently inferred demographic histories, respectively denoted as  $D_{A,A}$  and  $D_{B,B}$ . Conditional on these best-fitting demographic histories, we then independently fit the DFE for both species under a gamma-distributed model specification, yielding a *shape* parameter and a *scale* parameter for each species. To treat species *A* and *B* as independent (Boyko et al. 2008; Lawrie and Petrov 2014), we used a Poisson composite likelihood function identical to the likelihood function used in Huber et al., reproduced here (2017):

$$L(\hat{\theta}_A, \hat{\theta}_B | D_{A,A}, D_{B,B}) = \prod_{i=1}^{n-1} \frac{E(X_{i,A} | \hat{\theta}_A, D_{A,A})^{X_{i,A}}}{X_{i,A}!} \cdot e^{-E(X_{i,A} | \hat{\theta}_A, D_{A,A})} \cdot \prod_{i=1}^{m-1} \frac{E(X_{i,B} | \hat{\theta}_B, D_{B,B})^{X_{i,B}}}{X_{i,B}!} \cdot e^{-E(X_{i,B} | \hat{\theta}_B, D_{B,B})} \quad (3)$$

where  $n$  and  $m$  denote the sample size of species *A* and *B* respectively, and  $X_{i,A}$  denotes the number of SNPs at frequency  $i$  for a given species, *A*.

To test whether the *shape* ( $\theta$ ) and *scale* ( $\omega$ ) parameters of species *A* differ from those in species *B*, we utilized the LRT similar to that proposed in Huber et al. (2017):

$$= \frac{L(\hat{\theta}_A, \hat{\theta}_B, \hat{\omega}_A, \hat{\omega}_B | \hat{D}_{A,A}, \hat{D}_{B,B})}{L(\hat{\theta}_A, \hat{\theta}_B | \hat{D}_{A,A}, \hat{D}_{B,B})} \quad (4)$$

The null hypothesis for this test is the likelihood of the constrained model, i.e. the case in which  $\theta_A = \theta_B$  and  $\omega_A = \omega_B$ , while the full model allows for  $\theta_A \neq \theta_B$  and  $\omega_A \neq \omega_B$ . We optimized the likelihood function under both the constrained model and the full model. We optimized the constrained model by considering a grid of values for  $\theta$  and  $\omega$ . We then used *Fitdadi* to calculate a SFS of nonsynonymous variants for each pair of parameters for each species, and computed the log-likelihood for the fit of each model SFS to the data. Then, for each pair of parameter values, we summed the log-likelihood over both species. The pair of parameters with the highest log-likelihood was the MLE for this constrained null model.

Asymptotically,  $2 \ln L$  follows a  $\chi^2$  distribution with two degrees of freedom, as the full model has two additional free parameters relative to the constrained model. The critical value for the LRT is approximately 6 for a critical region of 0.05 for a singular test; however, given that we perform 471 pairwise comparisons, we approximate the critical value as 18.30 to correct for multiple hypothesis testing.

### Comparing DFEs Between Core and Accessory Genes

Similar to the method by which we compare DFEs across species, we performed a LRT comparing DFEs of core versus accessory genes, whereby the null hypothesis is a constrained model which assumes that core and accessory genes have the same  $\theta$  and  $\omega$  parameters. The alternative hypothesis is a full model in which  $\theta$  and  $\omega$  may be different for core and accessory genes.

We perform comparisons of the DFE between core and accessory genes for eighteen species; thus, the critical value

for this LRT is approximately 11.77, after Bonferroni correction.

### Phylogenetic Analyses of the DFE

To assess the phylogenetic signal of the DFE while controlling for the phylogenetic non-independence of species we considered four quantities: the proportion of neutral mutations (i.e.  $|s| < 10^{-6}$ ), the mean selection coefficient ( $E[s]$ ), the ancestral effective population size ( $N_{Anc}$ ), and the *shape* parameter of the best-fit gamma-distributed DFE.

To assess the phylogenetic signal of these four quantities, we computed Pagel's Lambda ( $\lambda$ ) using the R package "phylosignal" (Keck et al. 2016). This analysis was performed with 'phylosignal' version 1.3.1.

### Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

### Acknowledgments

The authors would like to thank Daisy Chen and Richard Wolff for their computational help, Jesse Shapiro and Pleuni Pennings for reading the manuscript and providing feedback, and Benjamin Good, Emily Ebel, and all members of the Garud and Lohmueller labs for helpful discussions.

### Funding

This work was funded in part through the NIH T32 5T32GM008185 (Systems in Integrative Biology) training grant to JM and NIH awards R35GM151023 to NRG and R35GM119856 to KEL. NRG also received support from an NSF CAREER award (no. 2240098), a Paul G. Allen Frontiers Group award, a University of California Hellman Fellows award, and a UCLA faculty research grant.

### Competing Interests

The authors declare no competing interests.

### Code Availability

All code for the bioinformatics pipeline, downstream analysis, and figure generation are available at Github: [https://github.com/garudlab/microbiome\\_demography\\_manuscript](https://github.com/garudlab/microbiome_demography_manuscript).

### Data Availability

All data analyzed in this paper is publicly available. Accessions for the data analyzed are available in [supplementary table S1, Supplementary Material](#) online.

### References

- Barrick JE, Lenski RE. Genome dynamics during experimental evolution. *Nat Rev Genet.* 2013;14(12):827–839. <https://doi.org/10.1038/nrg3564>.
- Beichman AC, Huerta-Sanchez E, Lohmueller KE. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu Ecol Evol Syst.* 2018;49(1):433–456. <https://doi.org/10.1146/annurev-ecolsys-110617-062431>.
- Blaser MJ. The past and future biology of the human microbiome in an age of extinctions. *Cell.* 2018;172(6):1173–1177. <https://doi.org/10.1016/j.cell.2018.02.040>.

- Bobay L-M, Ochman H. Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evol Biol.* 2018;18(1):1–12. <https://doi.org/10.1186/s12862-018-1272-4>.
- Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 2008;4(5):e1000083. <https://doi.org/10.1371/journal.pgen.1000083>.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics.* 1995;140(2):783–796. <https://doi.org/10.1093/genetics/140.2.783>.
- Bustamante CD, Wakeley J, Sawyer S, Hartl DL. Directional selection and the site-frequency spectrum. *Genetics.* 2001;159(4):1779–1788. <https://doi.org/10.1093/genetics/159.4.1779>.
- Chen J, Glémin S, Lascoux M. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol Biol Evol.* 2017;34(6):1417–1428. <https://doi.org/10.1093/molbev/msx088>.
- Chevin L-M, Lande R, Mace GM. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol.* 2010;8(4):e1000357. <https://doi.org/10.1371/journal.pbio.1000357>.
- Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, Magris M, Hidalgo G, Contreras M, Noya-Alarcón Ó, et al. The microbiome of uncontacted Amerindians. *Sci Adv.* 2015;1(3):e1500183. <https://doi.org/10.1126/sciadv.1500183>.
- Cornejo OE, Lefébure T, Bitar P, Lang P, Richards VP, Eilertson K, Do T, Beighton D, Zeng L, Ahn S-J, et al. Evolutionary and population genomics of the cavity causing bacteria *Streptococcus mutans*. *Mol Biol Evol.* 2013;30(4):881–893. <https://doi.org/10.1093/molbev/mss278>.
- Costea PI, Coelho LP, Sunagawa S, Munch R, Huerta-Cepas J, Forslund K, Hildebrand F, Kushugulova A, Zeller G, Bork P. Subspecies in the global human gut microbiome. *Mol Syst Biol.* 2017;13(12):960. <https://doi.org/10.15252/msb.20177589>.
- Dapa T, Wong DP, Vasquez KS, Xavier KB, Huang KC, Good BH. Within-host evolution of the gut microbiome. *Curr Opin Microbiol.* 2023;71:102258. <https://doi.org/10.1016/j.mib.2022.102258>.
- Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A.* 1991;88(16):7160–7164. <https://doi.org/10.1073/pnas.88.16.7160>.
- Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 2007;8(8):610–618. <https://doi.org/10.1038/nrg2146>.
- Eyre-Walker A, Woolfit M, Phelps T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics.* 2006;173(2):891–900. <https://doi.org/10.1534/genetics.106.057570>.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, et al. Traces of human migrations in *Helicobacter pylori* populations. *Science.* 2003;299(5612):1582–1585. <https://doi.org/10.1126/science.1080857>.
- Fay JC, Wu C-I. Hitchhiking under positive Darwinian selection. *Genetics.* 2000;155(3):1405–1413. <https://doi.org/10.1093/genetics/155.3.1405>.
- Filippo D, Cavalieri D, Paola D, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A.* 2010;107(33):14691–14696. <https://doi.org/10.1073/pnas.1005963107>.
- Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS Biol.* 2019;17(1):e3000102. <https://doi.org/10.1371/journal.pbio.3000102>.
- Gazave E, Ma L, Chang D, Coventry A, Gao F, Muzny D, Boerwinkle E, Gibbs RA, Sing CF, Clark AG, et al. Neutral genomic regions refine models of recent rapid human population growth. *Proc Natl Acad Sci U S A.* 2014;111(2):757–762. <https://doi.org/10.1073/pnas.1310398110>.
- Ghosh OM, Good BH. Emergent evolutionary forces in spatial models of luminal growth and their application to the human gut microbiota. *Proc Natl Acad Sci U S A.* 2022;119(28):e2114931119. <https://doi.org/10.1073/pnas.2114931119>.
- Gibbons R, Kapsimalis B. Estimates of the overall rate of growth of the intestinal microflora of hamsters, Guinea pigs, and mice. *J Bacteriol.* 1967;93(1):510–512. <https://doi.org/10.1128/jb.93.1.510-512.1967>.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 2009;5(10):e1000695. <https://doi.org/10.1371/journal.pgen.1000695>.
- Huang X, Fortier AL, Coffman AJ, Struck TJ, Irby MN, James JE, León-Burguete JE, Ragsdale AP, Gutenkunst RN. Inferring genome-wide correlations of mutation fitness effects between populations. *Mol Biol Evol.* 2021;38(10):4588–4602. <https://doi.org/10.1093/molbev/msab162>.
- Huber CD, Kim BY, Marsden CD, Lohmueller KE. Determining the factors driving selective effects of new nonsynonymous mutations. *Proc Natl Acad Sci U S A.* 2017;114(17):4465–4470. <https://doi.org/10.1073/pnas.1619508114>.
- Hughes AL. Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics.* 2005;169(2):533–538. <https://doi.org/10.1534/genetics.104.036939>.
- Human Microbiome Project Consortium. A framework for human microbiome research. *Nature.* 2012;486(7402):215–221. <https://doi.org/10.1038/nature11209>.
- Johri P, Riall K, Becher H, Excoffier L, Charlesworth B, Jensen JD. The impact of purifying and background selection on the inference of population history: problems and prospects. *Mol Biol Evol.* 2021;38(7):2986–3003. <https://doi.org/10.1093/molbev/msab050>.
- Keck F, Rimet F, Bouchez A, Franc A. PhyloSignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol Evol.* 2016;6(9):2774–2780. <https://doi.org/10.1002/ece3.2051>.
- Keightley PD, Eyre-Walker A. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics.* 2007;177(4):2251–2261. <https://doi.org/10.1534/genetics.107.080663>.
- Kim BY, Huber CD, Lohmueller KE. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics.* 2017;206(1):345–361. <https://doi.org/10.1534/genetics.116.197145>.
- Korpela K, Costea P, Coelho LP, Kandels-Lewis S, Willemsen G, Boomsma DI, Segata N, Bork P. Selective maternal seeding and environment shape the human gut microbiome. *Genome Res.* 2018;28(4):561–568. <https://doi.org/10.1101/gr.233940.117>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 2012;9(4):357–359. <https://doi.org/10.1038/nmeth.1923>.
- Lapierre M, Lambert A, Achaz G. Accuracy of demographic inferences from the site frequency spectrum: the case of the Yoruba population. *Genetics.* 2017;206(1):439–449. <https://doi.org/10.1534/genetics.116.192708>.
- Lawrie DS, Petrov DA. Comparative population genomics: power and principles for the inference of functionality. *Trends Genet.* 2014;30(4):133–139. <https://doi.org/10.1016/j.tig.2014.02.002>.
- Lin M, Kussell E. Inferring bacterial recombination rates from large-scale sequencing datasets. *Nat Methods.* 2019;16(2):199–204. <https://doi.org/10.1038/s41592-018-0293-7>.
- Liu Z, Good BH. Dynamics of bacterial recombination in the human gut microbiome. *PLoS Biol.* 2024;22(2):e3002472. <https://doi.org/10.1371/journal.pbio.3002472>.
- Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, et al. Strains, functions and dynamics in the expanded human microbiome project. *Nature.* 2017;550(7674):61–66. <https://doi.org/10.1038/nature23889>.

- Lyulina AS, Liu Z, Good BH. Linkage equilibrium between rare mutations. *Genetics*. 2024;228(3). <https://doi.org/10.1093/genetics/iyae145>.
- Maghini DG, Dvorak M, Dahlen A, Roos M, Kuersten S, Bhatt AS. Quantifying bias introduced by sample collection in relative and absolute microbiome measurements. *Nat Biotechnol*. 2024;42(2):328–338. <https://doi.org/10.1038/s41587-023-01754-3>.
- Martinez I Zurita A, Kyriazis CC, Lohmueller KE. The impact of non-neutral synonymous mutations when inferring selection on non-synonymous mutations. bioRxiv. 8 February 2024. <https://doi.org/10.1101/2024.02.07.579314>, preprint: not peer reviewed.
- Moeller AH. Genomic expansions in the human gut microbiome. *Genome Biol Evol*. 2021;13(7):evab156. <https://doi.org/10.1093/gbe/evab156>.
- Mortimer TD, Annis DS, O'Neill MB, Bohr LL, Smith TM, Poinar HN, Mosher DF, Pepperell CS. Adaptation in a fibronectin binding autolysin of *Staphylococcus saprophyticus*. *MSphere*. 2017;2(6):e00511–17. <https://doi.org/10.1128/mSphere.00511-17>.
- Murphy DA, Eyal E, Amster G, Sella G. Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements. *Elife*. 2022;12:e76065. <https://doi.org/10.7554/eLife.76065>.
- Nayfach S, Pollard KS. Population genetic analyses of metagenomes reveal extensive strain-level variation in prevalent human-associated bacteria. bioRxiv. 031757. 14 November 2015. <https://doi.org/10.1101/031757>, preprint: not peer reviewed.
- Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res*. 2016;26(11):1612–1625. <https://doi.org/10.1101/gr.201863.115>.
- Nielsen R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*. 2000;154(2):931–942. <https://doi.org/10.1093/genetics/154.2.931>.
- Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, Xu Z, Treuren V, Knight R, Gaffney PM, et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun*. 2015;6(1):6505. <https://doi.org/10.1038/ncomms7505>.
- O'Fallon BD. A method to correct for the effects of purifying selection on genealogical inference. *Mol Biol Evol*. 2010;27(10):2406–2416. <https://doi.org/10.1093/molbev/msq132>.
- Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;35(3):526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
- Pepperell CS, Casto AM, Kitchen A, Granka JM, Cornejo OE, Holmes EC, Birren B, Galagan J, Feldman MW. The role of selection in shaping diversity of natural *M. tuberculosis* populations. *PLoS Pathog*. 2013;9(8):e1003543. <https://doi.org/10.1371/journal.ppat.1003543>.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59–65. <https://doi.org/10.1038/nature08821>.
- Sakoparnig T, Field C, van Nimwegen E. Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. *Elife*. 2021;10:e65366. <https://doi.org/10.7554/eLife.65366>.
- Savageau MA. *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *Am Nat*. 1983;122(6):732–744. <https://doi.org/10.1086/284168>.
- Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turrioni S, Biagi E, Peano C, Severgnini M, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun*. 2014;5(1):3654. <https://doi.org/10.1038/ncomms4654>.
- Shiver AL, Sun J, Culver R, Violette A, Wynter C, Nieckarz M, Mattiello SP, Sekhon PK, Friess L, Carlson HK, et al. A mutant fitness compendium in Bifidobacteria reveals molecular determinants of colonization and host-microbe interactions. bioRxiv. 29 August 2023. <https://doi.org/10.1101/2023.08.29.555234>, preprint: not peer reviewed.
- Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci U S A*. 1993;90(10):4384–4388. <https://doi.org/10.1073/pnas.90.10.4384>.
- Sonnenburg ED, Smits SA, Mikhail T, Higginbottom SK, Wingreen NS, Sonnenburg JL. Diet-induced extinctions in the gut microbiota compound over generations. *Nature*. 2016;529(7585):212–215. <https://doi.org/10.1038/nature16504>.
- Sonnenburg JL, Sonnenburg ED. Vulnerability of the industrialized microbiota. *Science*. 2019;366(6464):eaaw9255. <https://doi.org/10.1126/science.aaw9255>.
- Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics*. 2009;182(1):205–216. <https://doi.org/10.1534/genetics.108.094904>.
- Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, Armanini F, Manghi P, Bonham K, Zolfo M, et al. The Prevotella copri complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe*. 2019;26(5):666–679. <https://doi.org/10.1016/j.chom.2019.08.018>.
- Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet*. 2005;1(3):e32. <https://doi.org/10.1371/journal.pgen.0010032>.
- Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J*. 2009;3(2):199–208. <https://doi.org/10.1038/ismej.2008.93>.
- Wakeley J, Hey J. Estimating ancestral population parameters. *Genetics*. 1997;145(3):847–855. <https://doi.org/10.1093/genetics/145.3.847>.
- Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Natl Acad Sci U S A*. 2005;102(22):7882–7887. <https://doi.org/10.1073/pnas.0502300102>.
- Wolff R, Garud NR. Pervasive selective sweeps across human gut microbiomes. bioRxiv. 23 December 2023. <https://doi.org/10.1101/2023.12.22.573162>, preprint: not peer reviewed.
- Wong DP, Good BH. Quantifying the adaptive landscape of commensal gut bacteria using high-resolution lineage tracking. *Nat Commun*. 2024;15(1):1605. <https://doi.org/10.1038/s41467-024-45792-0>.
- Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Meenakshi B, Knights D, Walters WA, Rob K. Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 2011;334(6052):105–108. <https://doi.org/10.1126/science.1208344>.
- Xie H, Guo R, Zhong H, Feng Q, Lan Z, Qin B, Ward KJ, Jackson MA, Yan X, Chen X. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst*. 2016;3(6):572–584. <https://doi.org/10.1016/j.cels.2016.10.004>.
- Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222–227. <https://doi.org/10.1038/nature11053>.