
Supplementary information

**Prehistoric global migration of vanishing
gut microbes with humans**

In the format provided by the
authors and unedited

Supplementary Methods

Contents

	Page
1 Estimating the age of microbial isolation using the molecular clock	3
1.1 Estimating neutral divergence from ANI measurements	3
1.2 Estimating the yearly mutation rate	4
1.3 Application to the ANI measurements in Fig. 1	4
2 Assessing genetic differentiation with F_{ST}	4
3 Estimating the divergence times of clonal strains	5
4 Estimating population split times from recent HGT events	7
4.1 Identity-by-state (IBS) tracts	7
4.2 Effects of population isolation	8
4.3 Generating <i>in silico</i> population splits	9
4.4 Application to empirical data	9
4.5 Mapping-based validation of IBS tract analysis	11
4.6 Assessing the impact of within-host strain heterogeneity	12
5 Demographic inference using Moments	13
5.1 Demographic model of a historical population split	13
5.2 Identifying species where the model can be applied	14
5.3 Visualizing and assessing model fits	15
5.4 Generalizability of the Moments analysis and comparisons to related work	16

List of Figures

S1	Lack of clonal strain sharing between Tsimane and Hadza is robust to the definition of clonality.	20
S2	IBS tracts of simulated microbial genomes using FastSimBac	21
S3	Comparison of MAG N50 values across populations.	22
S4	Validation of isolation-time inference using simulated data	23
S5	Comparison of within-population L99 scores for different populations	24
S6	Comparison of within-population L99 scores without clonal-strain filtering	25
S7	Comparison of MAG-based and mapping-based estimates of longest IBS tracts	26
S8	Mapping-based L99 estimates reproduce population-level patterns of isolation	27
S9	Effect of within-host strain heterogeneity on L99-based isolation estimates	28
S10	Flowchart summarizing the filtering steps used for the Moments analysis	29
S11	Effects of clade structure in the Moments analysis	30
S12	Residuals from the Moments analysis	31
S13	Summary of Moments fits (1/2)	32

S14 Summary of Moments fits (2/2) 34

S15 Population-specific SNVs for simulated data 35

S16 Analogous versions of Fig. 4B for the remaining species. 36

Supplementary Methods

1 Estimating the age of microbial isolation using the molecular clock

To convert the average nucleotide identity (ANI) measurements in Fig. 1 into a corresponding timescale, we employed a standard molecular clock approach informed by bacterial population genetics. We assume that the genetic differences in homologous regions of the genome emerge from a mixture of two processes: (i) point mutations, which alter individual sites; and (ii) recombination events, which replace longer stretches of DNA with a corresponding fragment sampled from another strain (Extended Data Fig. 5A). For a given pair of MAGs, the expected nucleotide divergence at neutral sites can then be expressed as a linear combination,

$$\langle d_{\text{neutral}} \rangle = 2\mu \cdot T_{\text{MRCA}}\lambda + \Delta_r(T_{\text{MRCA}}), \quad (\text{S1})$$

where μ is the spontaneous mutation rate (measured in units of mutations per bp per generation), T_{MRCA} is the time to the most recent common ancestor of the two strains (measured in years), λ is the average growth rate (measured in units of generations per year), and $\Delta_r(T_{\text{MRCA}})$ is a correction term that accounts for recombination events with other strains. The presence of this correction term means that ANI values cannot naively be used for dating based on the molecular clock alone. However, while the precise form of $\Delta_r(T_{\text{MRCA}})$ depends on the nature of the underlying recombination events [1], one can show that under mild assumptions, this term is always greater than zero. Neglecting recombination therefore yields an upper bound on the expected time to the most recent common ancestor,

$$T_{\text{MRCA}} \leq \frac{\langle d_{\text{neutral}} \rangle}{2\mu\lambda}, \quad (\text{S2})$$

which is illustrated in Extended Data Fig. 5a. We used this approach to estimate the upper bounds on the strain-level divergence times reported in the main text.

1.1 Estimating neutral divergence from ANI measurements

We estimated the numerator of Eq. S2 under the standard assumption that most synonymous mutations are effectively neutral. However, since the ANI is an average over all genomic sites, it is necessary to convert these measurements into a corresponding synonymous divergence rate. To do so, we decompose the ANI into a weighted average,

$$1 - \text{ANI} = \frac{d_S \cdot S + d_N \cdot N}{S + N} = d_S \left(\frac{1 + \frac{N}{S} \frac{d_N}{d_S}}{1 + \frac{N}{S}} \right) \quad (\text{S3})$$

where d_S and d_N represent the average divergence at synonymous and non-synonymous sites, respectively, and S and N are the corresponding numbers of synonymous and non-synonymous sites. While the d_N/d_S ratio generally varies with ANI [2, 3], previous work has shown that the d_N/d_S values for many species of gut bacteria are approximately ~ 0.1 when the ANI is $\lesssim 99\%$ [3]. Similarly, assuming equal codon frequencies and an unbiased mutational spectrum, the ratio of non-synonymous to synonymous sites is approximately 3:1. Substituting these ratios into Eq. S3 yields an approximate conversion between ANI and the neutral divergence in Eq. S2,

$$1 - \text{ANI} \approx d_S \left(\frac{1 + 3 \times 0.1}{1 + 3} \right) \approx \frac{d_{\text{neutral}}}{3}, \quad (\text{S4})$$

which applies for ANI values $\lesssim 99\%$.

1.2 Estimating the yearly mutation rate

We estimated the denominator of Eq. S2 using two complementary approaches. *In vitro* measurements of bacterial mutation rates (μ) range from 10^{-10} to 10^{-9} mutations per site per generation [4], while estimates of the growth rate (λ) in the human gut range from 1-10 generations per day [5, 6]. Combining these estimates leads to a midpoint value of $\mu\lambda \sim 10^{-6.5}$ per year, with a lower estimate of $\mu\lambda \sim 10^{-7.5}$. More direct estimates have been obtained in two recent studies in *E. coli* [7] and *B. fragilis* [8], which measured the wall-clock substitution rate in longitudinally sampled isolates from individual human subjects. These studies yielded estimates of $\mu\lambda \approx 7 \times 10^{-7}$ and $\mu\lambda \approx 2 \times 10^{-7}$ per year for *E. coli* and *B. fragilis* respectively, which approximately coincide with the midpoint value ($\mu\lambda \sim 10^{-6.5}$ per year) from the first-principles calculation above. A recent study based on reconstructed genomes from ancient fecal samples likewise estimated $\mu\lambda \approx 4 \times 10^{-7}$ per year for *Methanobrevibacter smithii* [9]. In the absence of specific information for most of the taxa in Fig. 1, we adopted the mid-point estimate above for all of our subsequent analyses, with the understanding that the actual mutation rates for any individual species may reasonably vary by ~ 2 -3-fold in either direction.

1.3 Application to the ANI measurements in Fig. 1

Substituting the above estimates into Eq. S2 yields a calibrated version,

$$T_{\text{MRCA}} \lesssim \frac{100 \times (1 - \text{ANI})}{2} \times 10^5 \text{ years}, \quad (\text{S5})$$

which is illustrated in Extended Data Fig. 5b. This approximate upper bound provides a benchmark for interpreting the ANI values in Fig. 1. Given the species definition employed by GTDB ($\text{ANI} \geq 95\%$), we estimate that two MAGs with the same species label likely shared a common ancestor within the last ~ 200 -300k years. The average ANI between Tsimane and Hadza MAGs is even larger ($\approx 98\%$, Fig. 1C), corresponding to a most recent common ancestor within the last ~ 100 k years.

We note that the ANI estimates in Fig. 1 were only computed for MAGs in the same GTDB species bin; MAGs from non-shared species will have a more distant evolutionary relationship across populations. However, since more than 80% of the Tsimane species had a conspecific relative in the Hadza (Fig. 1B), we conclude that these ANI-based estimates must apply to the vast majority of the Tsimane gut microbiome.

We also note that the timescale in Eq. S5 should be interpreted cautiously. Given the uncertainties in the *in situ* mutation rates and (μ) and growth rates (λ) in Eq. S2, the actual times may reasonably differ by several fold in either direction. However, these order-of-magnitude estimates still place strong constraints on the evolutionary history of the Tsimane strains. Taking the lower bound on the yearly mutation rate derived above ($\mu\lambda \sim 10^{-7.5}$ per year), we conclude that the Tsimane and Hadza MAGs must have shared a common ancestor within the last

$$T_{\text{MRCA}} = \frac{3}{2} \cdot \frac{2 \times 10^{-2}}{10^{-7.5}} \approx 10^6 \text{ years}. \quad (\text{S6})$$

This timescale is notably short compared to major geological events, or the origin of the *Homo* genus [10].

2 Assessing genetic differentiation with F_{ST}

To quantify the extent of population differentiation between the Tsimane and Hadza microbial populations, we applied the population genetic statistic F_{ST} , also known as the fixation index. F_{ST} quantifies the proportion of

total genetic variation that is attributable to differences between populations, as opposed to within-population variation. Values of F_{ST} range from 0 to 1, where 0 indicates no differentiation (all genetic diversity is shared equally between populations), and 1 indicates complete differentiation (populations are entirely distinct). We estimated F_{ST} using the formula:

$$F_{ST} \equiv \frac{d_{tot} - d_w}{d_{tot}} \equiv \frac{(1 - \text{ANI}_{tot}) - (1 - \text{ANI}_w)}{1 - \text{ANI}_{tot}}, \quad (\text{S7})$$

where d_{tot} and d_w represent the average pairwise genetic differences within the total population and within sub-populations, respectively. The observed F_{ST} values typically ranges from 0.05 to 0.25, with a mean of 0.1 (Extended Data Fig.6c). For comparison, the observed F_{ST} between African and non-African human populations typically ranges from 0.1 to 0.3 [11]. This indicates that the differentiation observed between the microbial populations is relatively small.

In simplest model of a population split with no migration or external recombination, the expected value of F_{ST} is given by

$$F_{ST} \approx \frac{T_{\text{split}}}{T_{\text{split}} + T_{\text{MRCA}}}, \quad (\text{S8})$$

where T_{MRCA} is the average time the most recent common ancestor for a pair of strains in the same population, and T_{split} is the corresponding time since the two populations first diverged. While the mapping between F_{ST} and T_{split} is more complicated in more realistic demographic models (see Section 5.1 below), this result suggests that the genetic isolation of the Tsimane and Hadza microbial populations is relatively recent compared to the T_{MRCA} of each species (100-200kya).

One implication of this relatively recent genetic isolation is that the divergence times estimated for MAGs from different populations cannot be directly interpreted as the population split time. Rather, the pairwise ANI values primarily reflect the ancestral diversity that already existed before the host populations diverged, as indicated by the broad overlap between the intra- and inter-population ANI values in Fig 1C. This contrasts with comparisons among much more divergent bacterial genomes (e.g., those differing by >10–20% ANI across distinct clades or species). In those cases, the within-species coalescence time is negligible relative to the between-species divergence time, allowing the inferred divergence to more faithfully approximate the true split between microbial populations. To precisely estimate the split time between the Tsimane and Hadza microbial populations, we turned to the additional methods described in Sections 3-5 below.

3 Estimating the divergence times of clonal strains

To quantify the timing of recent strain sharing between the Tsimane and the Hadza, we focused on pairs of clonal strains where the time of their most recent common genealogical ancestor could be inferred. Extended Data Fig. 5 shows that the genomes of these closely related strains will contain a mixture of clonal regions, which are nearly identical, and diverged segments imported from homologous recombination events with other strains. In the clonal regions of the genome, the recombination term in Eq. S1 vanishes,

$$\langle d_{\text{neutral}} \rangle \approx 2\mu \cdot T_{\text{MRCA}}\lambda, \quad (\text{S9})$$

allowing us to directly estimate the divergence time of the pair of strains using the molecular clock.

To distinguish the clonal regions of the genome from the recombined segments, we applied a recombination detection method (CP-HMM) from our previous study [1]. Briefly, this method models the spatial distribution of synonymous SNVs along the genome using a Hidden Markov Model. Because recombination

introduces many nucleotide differences simultaneously, it results in localized clusters of SNVs. This elevated density of SNVs is the key signature exploited by the method to identify recombination events.

We applied this HMM-based method to the 29,204 pairs of MAGs in our dataset with more than 50% identical genes. Below this threshold, it becomes harder to resolve individual recombination events using this method because of a higher chance of overlapping recombination events. We then computed the average synonymous divergence within the clonal regions of each pair (d_S^{clonal}), and estimated the time to the most recent common ancestor using Eq. S9,

$$T_{\text{MRCA}} \approx \frac{d_S^{\text{clonal}}}{2\mu\lambda}, \quad (\text{S10})$$

using the yearly mutation rate of $\mu\lambda \sim 10^{-6.5}$ mutations per site per year estimated in Section 1.2. The inferred T_{MRCA} values for each analyzed pair are shown in Extended Data Fig.7 and are provided in Supplementary Table 4.

While this method only directly infers the divergence times of strains with >50% identical genes, we extended this approach to more diverged clonal strains (10-50% identical genes) by extrapolating the observed relationship between divergence time and the fraction of identical genes. As a lower bound, we used the fact that the fraction of identical genes appears to decrease approximately linearly with clonal divergence for the data in Extended Data Fig.7c. This linear relationship suggests a linear extrapolation to pairs of strains pairs with fewer than 50% identical genes:

$$f \approx 1 - cT_{\text{MRCA}}, \quad (\text{S11})$$

where the constant of proportionality is equal to $c \approx 5 \times 10^{-4} \text{ year}^{-1}$. Under this model, the threshold we used for defining clonal strains – 10% identical genes – corresponds to an estimated T_{MRCA} of roughly 2,000 years.

As an upper bound, we used the fact that simple models of bacterial divergence [1] predict an exponentially declining relationship,

$$f \approx e^{-cT_{\text{MRCA}}}, \quad (\text{S12})$$

which reduces to Eq. S11 when $1 - f \ll 1$. This exponential dependence accounts for the saturation that occurs when successive transfers start to overlap with each other. Fitting this model to the data yields an extrapolated divergence time that is $\approx 2.5\times$ longer than Eq. S11 when $f \approx 0.1$. However, we note that in practice, the linear model appears to provide a better fit to the data even when $f \approx 0.5$, suggesting that the “true” extrapolation time may be closer to this lower bound.

We also performed the extrapolation separately for each microbial species (Extended Data Fig.7b). We found that the estimated divergence time at a threshold of 10% identical genes varies between approximately 1,000 and 4,000 years under the linear model, with a median of about 2,000 years (Extended Data Fig.7d). Together, these results support the conclusion that pairs of MAGs with >10% identical genes likely shared a common ancestor within the last $\sim 2,000$ years. Since this timescale is more recent than the estimated closing of the land bridge between Asia and North America [12], we reasoned that the levels of clonal strain sharing could provide a sensitive test for microbial isolation between the Tsimane and the Hadza.

Note on the definition of clonal strains. In the strain-sharing analysis above, we used the threshold of >10% identical genes as an operational definition of clonal strains. Changing this threshold will naturally affect the numerical values of strain-sharing rates reported in Fig. 2, and it will also shift the approximate

age boundary of strains that are defined as clonal, with higher thresholds corresponding to more recent age boundaries.

Although the precise threshold is to some degree arbitrary, several considerations guided our choice. First, even between otherwise unrelated strains, very recent horizontal gene transfer events can generate small numbers of identical genes. Therefore, the clonal pair threshold cannot be set too low in order to avoid conflating these short recombination-derived tracts (analyzed in Section 4) with true vertically transmitted regions. Previous work has shown that typical recombination events in the gut microbiome span on the order of ~ 10 genes [1], whereas our 10% cutoff corresponds to hundreds of genes for most species, providing a conservative buffer against this potential confounder.

Second, the empirical distribution of percent identical genes shows that nearly all Tsimane–Hadza strain pairs exhibit identical gene fractions close to zero (Fig. 2b, right panel). This indicates that our key finding of reduced Tsimane–Hadza strain sharing in the recent past is insensitive to the specific threshold. To demonstrate this point directly, we scanned the threshold across a broad range (5–50%) and repeated the strain-sharing analysis (Fig. S1). Across this full range, we observed little to no clonal strain sharing between Tsimane and Hadza in almost all microbial species, demonstrating that this qualitative result is highly robust to changes in the underlying definition.

Finally, because mutation and recombination rates vary across bacterial species, the precise age boundary associated with a given threshold also varies. Thus, the strain-sharing results in Fig. 2 should be interpreted as reflecting reduced sharing over an approximate timescale of $\lesssim 5,000$ years, rather than as precise species-specific estimates of the onset of genetic isolation.

4 Estimating population split times from recent HGT events

Since the maximum age of a pair of clonal strains is still comparatively recent ($\lesssim 5\text{kya}$), it is difficult to use these strain-sharing patterns to estimate the timing of the corresponding population splits. We therefore developed an alternative approach to date the onset of the genetic isolation between the Tsimane and Hadza strains, which leverages recent recombination events between more distantly related strains.

4.1 Identity-by-state (IBS) tracts

Identical segments between consecutive single nucleotide variants (SNVs) in a pair of genomes, also known as identity-by-state (IBS) tracts, contain valuable information about the demographic history of a population [13, 14]. IBS tracts arise from regions of recent common ancestry between the two genomes, either through direct vertical inheritance or recent recombination events between their ancestors. Regions with more recent ancestry tend to retain longer IBS tracts, as they have had less time for additional mutations and recombination events to accumulate. Consequently, the statistical distribution of IBS tract lengths reflects the underlying distribution of ancestry compositions, providing a means to infer past demographic processes such as population isolation [15].

Most previous studies of IBS tract length distributions have relied on models of crossover recombination in sexual organisms [16]. This makes it hard to apply existing IBS methods to the microbial populations studied in this work, where recombination events involve shorter DNA segments [17]. In this section, we develop methods to quantify the signal of population isolation encoded in IBS tract lengths and estimate the isolation time through simulations.

4.2 Effects of population isolation

Extended population isolation significantly affects the distribution of IBS tracts between populations. When a population splits into two isolated groups with no migration or gene flow, individuals sampled from different populations cannot share ancestry more recently than the population split time T_{split} . As a result, genome comparisons between the two populations will show fewer long IBS tracts compared to within-population comparisons, where genetic exchange continues. The magnitude of this reduction in IBS tract lengths provides quantitative insight into the duration of reproductive isolation between populations.

To focus on signals of recent demographic events, we recorded only the longest IBS tract between each pair of sequences. This approach improves computational efficiency and facilitates scaling to the full genome collection analyzed in this study. Throughout the remainder of this section, we use “IBS tract distribution” to refer to the distribution of these longest IBS tracts across genome pairs. We also excluded clonal pairs from this analysis (Section 3), since their longest IBS tracts are generated by vertical inheritance, rather than recent horizontal gene transfer. These latter events provide a more direct signal that two genomes are part of the same population unit, which is the goal of the present section.

To illustrate effect of population isolation on the IBS tract distribution, we utilized a population genetic simulator FastSimBac to simulate a simple neutral scenario of population split followed by complete isolation. FastSimBac generates genome sequences similar to our observed data, allowing us to compute IBS distributions within and between populations using the same procedures that we applied to the observed data. We selected simulation parameters designed to match the observed data of typical human gut commensals [1]. We set the genome length to 10^6 basepairs, the population-scaled mutation rate to $\theta = 0.01$, the population-scaled recombination rate to $R = 0.005$, and the typical recombination length to $\ell = 10^4$. The simulated demographic scenario assumes that an initially well-mixed population of size N splits into two populations (N each) T_{split} generations ago, after which they evolve in isolation. We simulated a sample of 50 individuals from each subpopulation across a range of scaled split times, defined as $\tilde{T}_{\text{split}} \equiv T_{\text{split}}\lambda/2N$, where λ is the average growth rate. The exact simulation commands can be found in the associated analysis source code.

Figure S2 shows that, compared to within-population comparisons, IBS distributions between populations exhibit a sharp cutoff at a characteristic length scale, similar to the pattern in observed data (e.g. Fig. 3B in the main text). As isolation time increases, this characteristic length scale decreases, reflecting the accumulation of mutations and recombination events that breaks apart long IBS tracts.

To systematically quantify the signal of population isolation, we focused on the 99th percentile of the IBS tract distribution (the L99 length). Since the cutoff in the tail of the IBS distribution is sharp in both simulations and observed data, this statistic is robust to the exact choice of percentile. The right panel of Fig. S2 shows how the between-population L99 length decreases as a function of isolation time in FastSimBac simulations. The clear decay of the L99 length as T_{split} increases suggests that we can infer the isolation time from the observed L99 lengths by inverting this functional relationship.

However, the precise form of the L99- T_{split} curve will vary across species due to differences in biological parameters such as recombination rate and tract length. Additionally, we expect that real populations may have more complex demographic histories than the simple scenario modeled above (e.g., a population bottleneck or expansion after the split). Generating species-specific L99- T_{split} curves through tailored FastSimBac simulations is therefore not only computationally demanding but also requires additional demographic assumptions for each species. Instead, we sought to derive these curves directly from observed genome sequences using a semi-parametric approach, which approximates the process of mutation accumulation after the population split.

4.3 Generating *in silico* population splits

Our approximate simulation method leverages the observed genome sequences to model the effects of a population split, by simulating the accumulation of additional mutations over time. This semi-parametric approach avoids the need for computationally intensive, species-specific recombination simulations while incorporating species-specific parameters in a natural way.

Key method assumptions. First, the within-population IBS distribution observed in present-day genomes must be a good approximation of the ancestral IBS distribution prior to the population split. Second, spontaneous mutations must be the primary factor breaking apart long IBS tracts over time, with recombination playing a secondary role over the length- and time-scales of interest.

Description of approximate simulation algorithm. For a given species, we estimated the L99 length versus isolation time relationship using the following procedure:

1. For each pair of sequences from the same population, we recorded the observed locations of synonymous SNVs and the total (core) genome length L .
2. For a given isolation time T_{split} , we introduced k new SNVs randomly along the genome, where k is a Poisson random variable with mean $2\mu \cdot T_{\text{split}}\lambda \cdot L$, with μ and λ denoting the average mutation rate and growth rate from Eq. S2. We used the yearly mutation rate of $\mu\lambda \sim 10^{-6.5}$ mutations per site per year estimated Section 1.2.
3. We then computed the IBS tract lengths based on the new set of SNVs and recorded the longest tract for each pair.
4. We then repeated steps 1-3 for all within-population sequence pairs and a range of isolation times T_{split} .
5. We then downsampled the data to match the observed number of between-population pairs for the species of interest and computed the L99 length for each simulated isolation time T_{split} .

To test the accuracy of this approximation, we applied it to the FastSimBac simulation data above. We generated approximate L99- T_{split} curves using within-population pairs from a single simulation and compared it to the ground truth computed in Fig. S2. Figure S4 shows that our approximate simulation method accurately recapitulates the ground truth relationship across a range of isolation times. The estimated curve closely follows the ground truth, with minor underestimation of the L99 length likely due to unmodeled recombination effects. This agreement suggests that mutation accumulation alone provides a strong first-order approximation of IBS tract dynamics in our regime of interest, supporting the validity of this approach for estimating isolation times in empirical data.

4.4 Application to empirical data

We used this approach to calculate IBS tracts and L99 lengths for all of the species in Fig. 3. To minimize the impact of purifying selection, we focused on fourfold synonymous sites within the core genes of each species (i.e. those present in >90% of MAGs, see Methods). To ensure that our analysis is independent of the clonal pair analysis in Fig. 2 (Section 3), we excluded genome pairs with >10% identical genes. For the remaining genome pairs, we concatenated contigs of the reference MAGs and recorded the spacings between consecutive SNVs. Only genomic positions covered in both genomes for a given pairwise comparison were included when computing SNV spacings.

We recorded IBS tracts distributions for all pairs of MAGs within each population as well as between populations (Tsimane-Hadza, North America-Europe, North America-East Asia). We found that within-population L99 lengths are highly consistent across populations (Fig. S5), suggesting that this statistic is robust to sampling differences across studies. For comparison, an analysis without clonal-strain filtering (Fig. S6) shows a much weaker correlation across populations, indicating that within-population L99 estimates can be sensitive to uneven sampling of closely related strains and underscoring the importance of our filtering step. Consequently, we utilized the average within-population L99 length regardless of the specific population pair.

Finally, we assessed whether differences in MAG assembly quality could influence our results by comparing N50 values across all five populations included in this study (Tsimane, Hadza, Asia, Europe, and North America; Fig. S3). While we detected statistically significant differences in N50 across populations (Kruskal–Wallis test, $p < 0.001$), the associated effect size was negligible (eta-squared, $\eta^2 = 0.0016$), indicating that population membership explains only 0.16% of the total variance in N50. Consistent with this, median N50 values were highly similar across populations (Tsimane: 21,513 bp; Hadza: 24,275 bp; Asia: 25,921 bp; Europe: 25,149 bp; North America: 30,509 bp), suggesting that differences in assembly contiguity are unlikely to contribute meaningfully to the observed patterns in L99.

Inferring the population split time. We generated species-specific L99- T_{split} curves by performing mutation accumulation simulations using the algorithm described in Section 4.3. In these simulations, mutation accumulation was restricted to genomic sites that were covered in both genomes for each simulated pairwise comparison. We pooled all within-population genome pairs across all populations and excluded species with fewer than 200 such pairs. We simulated 20 different values of T_{split} that were evenly spaced on a log scale between 10^3 and $10^{5.5}$ years. For each population pair, we downsampled the simulated data to match the number of observed between-population genome pairs and computed the corresponding L99 length for the simulated value of T_{split} .

In practice, we found that at shorter isolation times, the simulated L99 lengths were often indistinguishable from the within-population L99 lengths due to sampling error. This observation suggests a lower bound of approximately 1,000 years on the resolvable isolation time between populations.

To obtain a continuous function relating the L99 length to T_{split} , we applied isotonic regression to enforce monotonicity in the simulated results. This step, which was implemented using the Python package scikit-learn [18], reduces noise-related fluctuations, particularly for small values of T_{split} where the L99 lengths are similar. Using this approach, we extracted the smoothed L99 lengths at the simulated values of T_{split} and performed linear interpolation between $\log(\text{L99})$ and $\log(T_{\text{split}})$. We chose to interpolate in log space to account for the power-law scaling observed in simulations (Fig. S4).

We inferred the implied isolation time between populations by solving for the point where the interpolated L99- T_{split} curve intersected the observed between-population L99 length, using the default root-finding algorithm in scipy [19]. When the observed L99 length fell outside of the simulated range, we assigned the inferred isolation time as either below 10^3 years or above $10^{5.5}$ years. In practice, no comparisons exceeded the upper range, while Europe-North America pairs frequently fell below the lower bound, suggesting no detectable isolation. This result is consistent with the near-identical within- and between-population L99 values reported in Fig. 3C of the main text.

To quantify the uncertainty in the inferred isolation times, we performed 1,000 bootstrap replicates for each population pair and calculated 95% confidence intervals. To assess the statistical significance of the observed genetic isolation, we also performed a permutation test independently for each species. Genome-pair labels (within-population versus between-population) were randomly permuted across all genome pairs for that species, preserving the number of comparisons in each category, and the difference between the

within- and between-population L99 values was recomputed for each permutation. Repeating this procedure for $N = 10,000$ times yielded an empirical null distribution of L99 differences, from which p-values were calculated as the fraction of permutations with a difference greater than or equal to the observed value. P-values across species were subsequently adjusted for multiple testing using the Benjamini–Hochberg procedure. The full list of point estimates of isolation times, their confidence intervals, and the permutation test p-values are provided in Supplementary Table 5.

4.5 Mapping-based validation of IBS tract analysis

The IBS-based analyses above rely on identifying SNVs and IBS tracts from MAG-derived core genome alignments. While this approach has the advantage of operating directly on genome sequences, it depends on a MAG reconstruction pipeline whose parameter choices and MAG quality (e.g., contiguity) could potentially impact the inferred IBS tract lengths. To assess the robustness of our results to potential biases introduced during MAG reconstruction, we developed an orthogonal validation method based on read mapping to a common reference genome. This approach generalizes IBS tracts to metagenomic samples by identifying contiguous reference-genome regions without that lack genetic variation between a pair of samples (see below), providing an independent confirmation of the patterns described above.

Read mapping and allele frequency estimation. We mapped reads from each metagenomic sample to a database of species-representative genomes using Bowtie2 [20]. For every pair of samples included in the genome-based analysis, we computed nucleotide frequencies at each position of each reference genome. These allele-frequency profiles form the basis for calling metagenomic population SNVs (pop-SNVs) and for defining generalized IBS tracts between metagenomic samples.

Metagenomic population SNVs (pop-SNVs). Following the terminology introduced in inStrain [21], we define a metagenomic population SNV (pop-SNV) as any genomic position at which two samples share no allele in common—major or minor—after applying the sequencing-error filters described below. Positions where the two samples share at least one allele are not considered pop-SNVs and therefore do not break the generalized IBS tracts defined in this mapping-based framework.

Sequencing error filtering. To distinguish genuine allelic differences from sequencing errors, we applied an error-correction procedure similar to that used in inStrain [21]. For each allele with sufficient coverage, we computed the probability that its observed count could arise from a fixed sequencing error rate of 10^{-3} under a Poisson model. Alleles whose error probability exceeded 0.05 were treated as compatible with sequencing noise and ignored, while alleles failing this criterion were designated as true pop-SNVs.

Identification of generalized IBS tracts. We generalized the concept of IBS tracts from genome–genome comparisons to metagenome–metagenome comparisons by defining contiguous regions along the reference genome that contain no pop-SNVs between the two samples. Positions with insufficient coverage in both samples (default threshold $< 5\times$) were not considered pop-SNVs but also did not break contiguity, allowing us to accommodate heterogeneous depth profiles without introducing false boundaries. Using this criterion, we identified the longest such tract for each sample pair across all reference genomes.

As in the MAG-based workflow, we retained only the longest tract per pair and constructed cumulative tract-length distributions for within- and between-population comparisons. A copy of our implementation can be found in the ZipStrain repository: <https://github.com/OlmLab/ZipStrain>.

Relation to genome-based IBS tracts. When both samples are dominated by a single strain, the generalized metagenomic IBS tracts inferred from read mappings should closely match the genome-based IBS tracts obtained from MAGs, aside from minor differences introduced during MAG reconstruction. In samples containing mixtures of strains at appreciable frequencies, however, the generalized IBS profile may differ from the genome-based profile, since the presence or absence of pop-SNVs will depend on the specific combinations of strain genotypes present in each sample. Despite this added complexity, the generalized IBS profile still reflects the recent recombination history among the strains represented in the two samples, and thus retains signatures of population isolation, even if the connection is less direct than in pure genome–genome comparisons.

Comparison with MAG-based IBS estimates. We first compared the longest generalized IBS tract identified for each sample pair to the corresponding genome-based value derived from MAGs. For a representative species, *Prevotella faecis*, the two approaches yielded highly concordant results (Pearson $r = 0.92$), demonstrating that the mapping-based procedure closely reproduces the MAG-based tract lengths when evaluated on a pairwise basis (Fig. S7A). Extending this comparison across all species, we observed a distribution of correlation coefficients with a peak around $r \approx 0.9$ (median $r = 0.73$), indicating overall good agreement between the two approaches (Fig. S7B). The minority of species with lower pairwise correlations may arise from differences in underlying strain mixtures, uneven coverage, or other technical or biological factors that influence the generalized IBS estimates.

We next examined population-level summaries, where heterogeneity in within-sample strain composition is less likely to obscure broader patterns of population isolation. For each species, we computed the L99 statistic for the generalized IBS distributions within and between populations (Fig. S8A-B). The resulting values reproduced the overall patterns reported in Fig. 3C of the main text: the L99 lengths between the Tsimane and Hadza were consistently shorter than those between Asia and North America, indicating stronger isolation between the former groups. The metagenomic L99 values were also highly consistent across within-population comparisons (Fig. S8C-D), similar to the genome-based case (Fig. S5). Together, these results demonstrate that the mapping-based approach validates our key observation of significantly reduced gene flow between the Tsimane and Hadza compared to population pairs with comparable geographical separation.

4.6 Assessing the impact of within-host strain heterogeneity

Our IBS analyses assume that each MAG represents the genotype of a single dominant strain within a sample. However, real metagenomic samples often contain mixtures of conspecific strains, which could in principle introduce spurious variation into assembled genomes or the inferred IBS tract lengths. To evaluate the potential impact of within-host strain heterogeneity on our conclusions, we leveraged allele-frequency information from the read-mapping pipeline described in Section 4.5 to quantify strain-level heterogeneity and test its influence on the L99-based isolation analysis in Fig. 3.

Defining a measure of strain heterogeneity. Following previous work [3], for each species–sample pair, we computed the fraction of high-coverage sites (coverage ≥ 20) at which the major allele has frequency $< 80\%$. This “heterogeneity score” is expected to be low when a single strain dominates the sample, and higher when mixtures of divergent conspecific strains are present at intermediate frequencies. Across species, we observed consistently higher heterogeneity scores in samples from non-industrialized populations than in industrialized populations, potentially reflecting known differences in microbial diversity across lifestyles (Fig. S9A). Among the non-industrialized samples, those from which a high-quality MAG was reconstructed tended to display lower heterogeneity scores (Fig. S9A), suggesting that MAG recovery is biased toward

samples dominated by a single strain.

Evaluating the effect of heterogeneity on L99 estimates. To test whether strain heterogeneity could influence the IBS-based population-isolation analysis, we restricted the dataset to samples with low heterogeneity scores (threshold $< 10^{-3}$) and repeated the full L99 pipeline using only MAGs derived from these samples (Fig. S9B-C). The resulting L99 values showed strong concordance with those computed using the complete MAG set (Fig. S9C), indicating that our estimates are not strongly biased by samples with higher levels of strain diversity. Importantly, species continued to exhibit the same qualitative patterns of population isolation observed in Fig. 3C (Fig. S9D), in which the Tsimane-Hadza comparisons displayed markedly shorter between-population L99 lengths than industrialized-pair comparisons. Together, these results indicate that our inferences based on IBS patterns are robust to within-host strain heterogeneity.

5 Demographic inference using Moments

To corroborate the estimates above, we also employed an existing demographic inference method (Moments; 22) which leverages information from the joint site frequency spectrum (SFS) of SNV frequencies within populations. Moments is traditionally applied to diploid organisms, so we implemented several additional filtering procedures to adapt it to the microbial populations analyzed here.

To ensure sufficient resolution in the observed SNV frequencies for demographic inference, we focused on species with the largest number of MAGs. Specifically, we retained species with more than 30 MAGs in both the Tsimane and Hadza populations. For context, Medina-Muñoz et al. used 40 haploid copies per population in their inference of human demographic history [23]. We computed the site frequency spectrum for these selected Tsimane and Hadza MAGs using four-fold degenerate synonymous sites. As part of this process, we applied the projection function in Moments that standardizes total allele counts across sites by downsampling to a specified threshold. We set this projection threshold to 0.9 times the number of MAGs in each population, rounded down to the nearest integer.

Since bacterial MAGs derived from metagenomes may have higher rates of SNV calling errors compared to high-quality haplotypes from whole-genome sequencing (WGS) of human genomes (e.g. in Ref. 23), we sought to minimize their impact by masking singleton SNVs in each population during model fitting. Specifically, this was implemented by setting `data.mask[1, :]=True` and `data.mask[:, 1]=True` in Moments. These masks were removed when evaluating the model fit using Poisson residuals (see Section 5.2 below).

5.1 Demographic model of a historical population split

We fit the observed joint SFS data to a simple demographic model of a population split with subsequent migration, implemented in the `moments.Demographics2D.split_mig` function in Moments. This model assumes that an ancestral population of constant size N_a instantaneously splits into two populations $T_{\text{split}} \cdot \lambda$ generations ago, where T_{split} is measured in years and λ is the average growth rate from Eq. S2. The sizes of the new populations are $N_a v_1$ and $N_a v_2$, respectively, allowing for the total population size to increase or decrease after the split. We also allow for continuous migration between the two populations at a rate m_0 per individual per generation. This model constitutes a multi-population generalization of the two-epoch model employed by Mah *et al* [24] to study microbial demographic histories in several industrialized human cohorts. In particular, the addition of an explicit migration rate parameter allows us to quantify the timing and extent of genetic isolation between the Tsimane and Hadza microbiomes.

Since Moments outputs the inferred split time in population-scaled units ($\tilde{T}_{\text{split}} \equiv T_{\text{split}}\lambda/2N_a$), we applied the following steps to convert this estimate into years. We note that the ancestral population size (N_a) does not affect the relative probability density of the predicted SFS; it only determines the number of sites that are polymorphic in a population, which is proportional to the total number of sites (L). Moments provides the scaling factor $\theta \equiv 4N_a\mu L$ using the function `Inference.optimal_sfs_scaling`, which compares the model with the observed number of SNVs in the data. Since these SNVs are derived from the subset of sites that passed projection, we calculated L as the number of synonymous sites sufficiently covered in the MAGs to meet the projection threshold. We then used the following formula to convert the inferred split time into years:

$$T_{\text{split}} = \frac{\tilde{T}_{\text{split}} \cdot 2N_a}{\lambda} = \tilde{T}_{\text{split}} \cdot \frac{\theta}{4\mu\lambda L} \cdot 2. \quad (\text{S13})$$

using the yearly mutation rate of $\mu\lambda \sim 10^{-6.5}$ mutations per site per year estimated in Section 1.2.

5.2 Identifying species where the model can be applied

In this section, we describe the quality control filters applied to refine the dataset for demographic inference, some of which are based on the quality of model fit (though not on the numerical value of T_{split}). Applying these filters reduced the total number of species analyzed to the 15 that are shown in Fig. 4 in the main text. The number of species retained at each filtering step is summarized in a flowchart (Fig. S10). The inferred model parameters and estimated population split times are reported in Supplementary Table 6.

Impact of clade structure. Some bacterial species harbor complex population structures, such as distinct clades or subspecies, where horizontal gene transfer between clades occurs at lower rates. This complexity violates the assumptions of our simple demographic model and could bias the inferred split times. To focus on species with sufficiently simple population structures that align with the assumptions of our model, we applied a heuristic approach to detect and filter out species with strong clade structure. We applied hierarchical clustering to the pairwise distance matrix based on ANI differences, using the `scipy.hierarchy.fcluster` function with the criterion set to 'maxclust_monocrit' to form two maximally separated clusters. To be conservative, we excluded a species from the Moments analysis if the smaller of the two clusters contained more than one MAG.

To illustrate the impact of clade structure on model fit and inference, we compared both the mean residual sizes and the inferred split times between species identified as clade-structured and those without clade structure (Fig. S11). Here, residuals refer to the difference between the model prediction and the observed SFS, calculated using `moments.Inference.linear_Poisson_residual`. The Poisson residual is defined as $(\text{model} - \text{data})/\sqrt{\text{model}}$, where "model" and "data" represent individual elements of the predicted and observed SFS.

We found that species without clade structure tend to have smaller residuals, indicating better model fit, while clade-structured species show much larger inferred split times as expected. This discrepancy is likely due to the larger and older genetic diversity maintained by the presence of two or more clades, which can inflate the inferred split time defined by geography alone. These results highlight the importance of controlling for additional population structure during demographic inference.

Filters based on uncertainty and residual size One internal check for model fit is the uncertainty in parameter estimates provided by Moments. Moments calculates this uncertainty using the inverse of the Godambe Information Matrix [25]. In practice, we observed that the uncertainty in some parameter estimates

can exceed the magnitude of the parameter itself, potentially indicating model misspecification or poorly constrained parameters. To address this, we retained only species with well-defined uncertainties, specifically those where the ratio of uncertainty to inferred split time was smaller than 0.5.

In addition, we applied a filter based on the mean absolute residual size to focus on species with the best model fits. We computed the mean absolute residual size for the two-dimensional joint SFS and the single-population (marginalized) SFS. We used a threshold of mean absolute residual size < 2 to exclude species with poor fits. Figure S12 shows the mean absolute residual sizes for the 27 species that passed the clade structure and uncertainty filters above, highlighting the 15 that were used in Fig. 4.

5.3 Visualizing and assessing model fits

Moments predicts a two-dimensional joint SFS (Figs. S13 and S14), but the high dimensionality of this distribution makes it challenging to visualize differences between the data and the inferred model, as well as between different demographic scenarios. To assess model fit, we first examined traditional diagnostics, including the marginalized SFS for each population, direct comparisons of observed and predicted joint SFS heatmaps, and the distribution of residuals across observed SNV frequencies.

These diagnostics, which are shown in Figs. S13 and S14, demonstrate overall agreement between model predictions and observed data. However, they do not directly reveal signals of population isolation, making it hard to directly distinguish the null scenario of no population split (free migration) from the inferred model and the data.

Population specific SNVs. To address this limitation, we also employed a summary statistic for the two-dimensional SFS based on the frequency of “population-specific SNVs” – variants that are exclusively observed in only one of the two populations. The rationale is that population isolation reduces the chances that a recently produced SNV will be shared between the two populations. Importantly, the probability of this event will depend on the total frequency of the SNV across populations, making this statistic particularly sensitive to the effects of isolation.

The probability that a SNV is observed in only one population depends on three main factors (Fig. 4A). First, SNVs at high frequencies tend to have persisted in the population for longer, increasing the likelihood of migration between populations. Second, high-frequency SNVs are more likely to have been present in the ancestral population prior to the split, making them more likely to appear in both populations. Lastly, low-frequency SNVs are less likely to be sampled and may appear exclusive to one population by chance. Together, these factors contribute to a characteristic decay in the fraction of population-specific SNVs as total population frequency increases.

We tested this idea by plotting the fraction of population-specific SNVs as a function of total population frequency, using simulated SFS data for the same demographic model applied in our Moments inference. Figure S15 illustrates how this relationship varies with population split time and migration rate between populations. As expected, the sharpest decay in population-specific SNVs was observed in the no-separation scenario ($\tilde{T}_{\text{split}} = 0$). In simulations without migration following the split (population-scaled migration rate $m \equiv 2N_a m_0 = 0$), the decay was highly sensitive to split time, effectively distinguishing different separation times below $\tilde{T}_{\text{split}} = 2$. However, with moderate migration ($m = 1$), the curve still distinguished shorter split times but lost sensitivity for splits older than $\tilde{T}_{\text{split}} = 1$. At higher migration rates ($m = 5$), the ability to distinguish split times diminished further, with the curves approaching the no-split (or “free migration”) limit.

Having demonstrated the utility of this statistic with simulations, we next applied it to the 15 species with the best Moments fits. In the majority of species, this statistic revealed a clear distinction between the

observed data and the no-split scenario (Figs. 4B and S16), highlighting the robust effect of the population split. Additionally, we observed strong agreement between the predicted curves from the inferred model and the data, validating the model’s ability to capture key patterns in the site frequency spectrum.

5.4 Generalizability of the Moments analysis and comparisons to related work

A key limitation of the Moments-based demographic inference is the reliance on a specified parametric model. Although the model-fit metrics described above provide an internal check against model misspecification, our analysis is restricted to the subset of species whose SNV-frequency spectra are simple enough to be captured by our minimal demographic framework. In particular, many microbial species exhibit clade structure or more complex demographic histories, which would require more elaborate models that are beyond the scope of the present study.

To assess the generalizability of the findings beyond the 15 species included in this analysis, we compared the divergence times inferred from Moments to those obtained from the L99 analysis (Extended Data Fig 8). Despite differences in the exact numerical values, we observed broad agreement between these two independent estimates. Given that the L99 analysis reveals signatures of genetic isolation across a much larger set of taxa, this agreement suggests that the SNV-frequency patterns in more complex species are also likely to contain comparable signals of population isolation. Extending the Moments framework to recover such signals is an interesting direction for future work.

Interestingly, for the handful of species that were also analyzed in a recent study by Mah *et al* [24] — *Ruminococcus bromii*, *Ruminococcus intestinalis*, and *Faecalibacterium prausnitzii* — we found that the Tsimane-Hadza split times inferred by Moments approximately coincided with the population expansions inferred by Mah *et al* using strains from industrialized human cohorts. The reasons for this correspondence remain unclear, but would be interesting to explore in future analyses combining industrialized and non-industrialized populations.

- [1] Zhiru Liu and Benjamin H. Good. Dynamics of bacterial recombination in the human gut microbiome. *PLOS Biology*, 22(2):e3002472, February 2024.
- [2] Eduardo P. C. Rocha, John Maynard Smith, Laurence D. Hurst, Matthew T. G. Holden, Jessica E. Cooper, Noel H. Smith, and Edward J. Feil. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of Theoretical Biology*, 239(2):226–235, March 2006.
- [3] Nandita R. Garud, Benjamin H. Good, Oskar Hallatschek, and Katherine S. Pollard. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLOS Biology*, 17(1):e3000102, January 2019.
- [4] Way Sung, Matthew S. Ackerman, Samuel F. Miller, Thomas G. Doak, and Michael Lynch. Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences*, 109(45):18488–18492, November 2012.
- [5] Michael A. Savageau. *Escherichia coli* Habitats, Cell Types, and Molecular Mechanisms of Gene Control. *The American Naturalist*, 122(6):732–744, December 1983.
- [6] Olivia M. Ghosh and Benjamin H. Good. Emergent evolutionary forces in spatial models of luminal growth and their application to the human gut microbiota. *Proceedings of the National Academy of Sciences*, 119(28):e2114931119, July 2022.
- [7] Mohamed Ghalayini, Adrien Launay, Antoine Bridier-Nahmias, Olivier Clermont, Erick Denamur, Mathilde Lescat, and Olivier Tenaillon. Evolution of a Dominant Natural Isolate of *Escherichia coli* in the Human Gut over the Course of a Year Suggests a Neutral Evolution with Reduced Effective Population Size. *Applied and Environmental Microbiology*, 84(6):e02377–17, March 2018.
- [8] Shijie Zhao, Tami D. Lieberman, Mathilde Poyet, Kathryn M. Kauffman, Sean M. Gibbons, Mathieu Groussin, Ramnik J. Xavier, and Eric J. Alm. Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell Host & Microbe*, 25(5):656–667.e8, May 2019.
- [9] Marsha C. Wibowo, Zhen Yang, Maxime Borry, Alexander Hübner, Kun D. Huang, Braden T. Tierney, Samuel Zimmerman, Francisco Barajas-Olmos, Cecilia Contreras-Cubas, Humberto García-Ortiz, Angélica Martínez-Hernández, Jacob M. Luber, Philipp Kirstahler, Tre Blohm, Francis E. Smiley, Richard Arnold, Sonia A. Ballal, Sünje Johanna Pamp, Julia Russ, Frank Maixner, Omar Rota-Stabelli, Nicola Segata, Karl Reinhard, Lorena Orozco, Christina Warinner, Meradeth Snow, Steven LeBlanc, and Aleksandar D. Kostic. Reconstruction of ancient microbial genomes from the human gut. *Nature*, 594(7862):234–239, June 2021.
- [10] Susan C. Antón, Richard Potts, and Leslie C. Aiello. Evolution of early Homo: An integrated biological perspective. *Science*, 345(6192):1236828, July 2014.
- [11] Anders Bergström, Shane A. McCarthy, Ruoyun Hui, Mohamed A. Almarri, Qasim Ayub, Petr Danecek, Yuan Chen, Sabine Felkel, Pille Hallast, Jack Kamm, Hélène Blanché, Jean-François Deleuze, Howard Cann, Swapnil Mallick, David Reich, Manjinder S. Sandhu, Pontus Skoglund, Aylwyn Scally, Yali Xue, Richard Durbin, and Chris Tyler-Smith. Insights into human genetic variation and population history from 929 diverse genomes. *Science*, 367(6484):eaay5012, March 2020.

- [12] T. Pico, J. X. Mitrovica, and A. C. Mix. Sea level fingerprinting of the bering strait flooding history detects the source of the younger dryas climate event. *SCIENCE ADVANCES*, 6(9), FEB 2020.
- [13] Kelley Harris and Rasmus Nielsen. Inferring Demographic History from a Spectrum of Shared Haplotype Lengths. *PLoS Genetics*, 9(6):e1003521, June 2013.
- [14] Daniel B Weissman and Oskar Hallatschek. Minimal-assumption inference from population-genomic data. *eLife*, 6:e24836, July 2017.
- [15] Shiping Liu, Eline D. Lorenzen, Matteo Fumagalli, Bo Li, Kelley Harris, Zijun Xiong, Long Zhou, Thorfinn Sand Korneliussen, Mehmet Somel, Courtney Babbitt, Greg Wray, Jianwen Li, Weiming He, Zhuo Wang, Wenjing Fu, Xueyan Xiang, Claire C. Morgan, Aoife Doherty, Mary J. O’Connell, James O. McInerney, Erik W. Born, Love Dalén, Rune Dietz, Ludovic Orlando, Christian Sonne, Guojie Zhang, Rasmus Nielsen, Eske Willerslev, and Jun Wang. Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, 157(4):785–794, May 2014.
- [16] Gilean A.T McVean and Niall J Cardin. Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1387–1393, July 2005.
- [17] Nicola De Maio and Daniel J. Wilson. The Bacterial Sequential Markov Coalescent. *Genetics*, 206(1):333–343, May 2017.
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [19] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [20] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4):357–359, 2012.
- [21] Matthew R. Olm, Alexander Crits-Christoph, Keith Bouma-Gregson, Brian A. Firek, Michael J. Morowitz, and Jillian F. Banfield. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nature Biotechnology*, 39(6):727–736, June 2021.
- [22] Julien Jouganous, Will Long, Aaron P Ragsdale, and Simon Gravel. Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. *Genetics*, 206(3):1549–1567, July 2017.
- [23] Santiago G. Medina-Muñoz, Diego Ortega-Del Vecchyo, Luis Pablo Cruz-Hervet, Leticia Ferreyra-Reyes, Lourdes García-García, Andrés Moreno-Estrada, and Aaron P. Ragsdale. Demographic modeling of admixed Latin American populations from whole genomes. *The American Journal of Human Genetics*, 110(10):1804–1816, October 2023.

- [24] Jonathan C Mah, Kirk E Lohmueller, and Nandita R Garud. Inference of the Demographic Histories and Selective Effects of Human Gut Commensal Microbiota Over the Course of Human History. *Molecular Biology and Evolution*, 42(2):msaf010, February 2025.
- [25] Alec J. Coffman, Ping Hsun Hsieh, Simon Gravel, and Ryan N. Gutenkunst. Computationally Efficient Composite Likelihood Statistics for Demographic Inference. *Molecular Biology and Evolution*, 33(2):591–593, February 2016.

Supplementary Figures

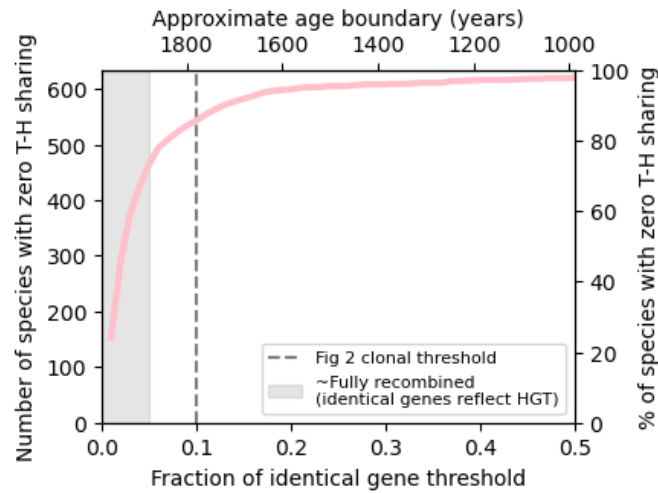


Figure S1: Lack of clonal strain sharing between Tsimane and Hadza is robust to the definition of clonality. The clonal strain-sharing rates shown in Fig. 2 were recalculated using a range of thresholds for the fraction of identical genes between genomes. The solid line indicates the number of species with zero strain sharing at each threshold. The secondary x-axis shows the estimated age boundary corresponding to each threshold for defining clonal relationships, as inferred from recombination rate analysis. The shaded region marks thresholds below 5% identical genes, which no longer reflect clonality and are thus not informative for this analysis. This sensitivity analysis demonstrates that the vast majority of species show no recent strain sharing between Tsimane and Hadza, regardless of the specific threshold used to define clonal strains.

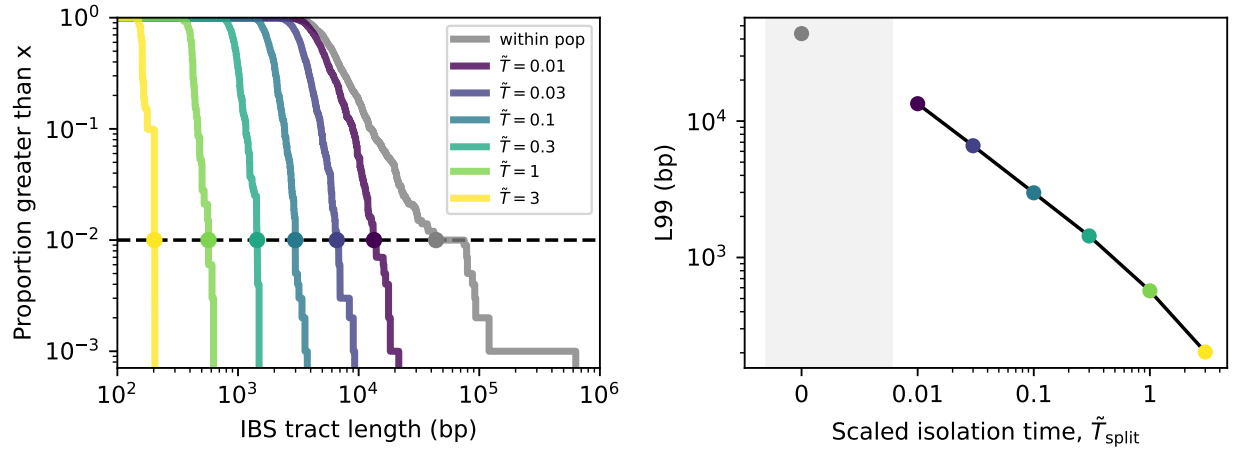


Figure S2: Left: IBS tract length distributions of simulated microbial genomes using FastSimBac. Curves show the reverse cumulative distributions of the longest IBS tract aggregated over 1000 pairs of sequences. All distributions are computed for between-population genome pairs, except for the “within pop” curve. Dashed line and circles indicate the tract lengths at the 99th percentile (L99). Right: Same data as the left panel, showing the L99 lengths as a function of the scaled isolation time between the populations in the simulation. Symbol at zero isolation time indicates the value observed for within-population pairs.

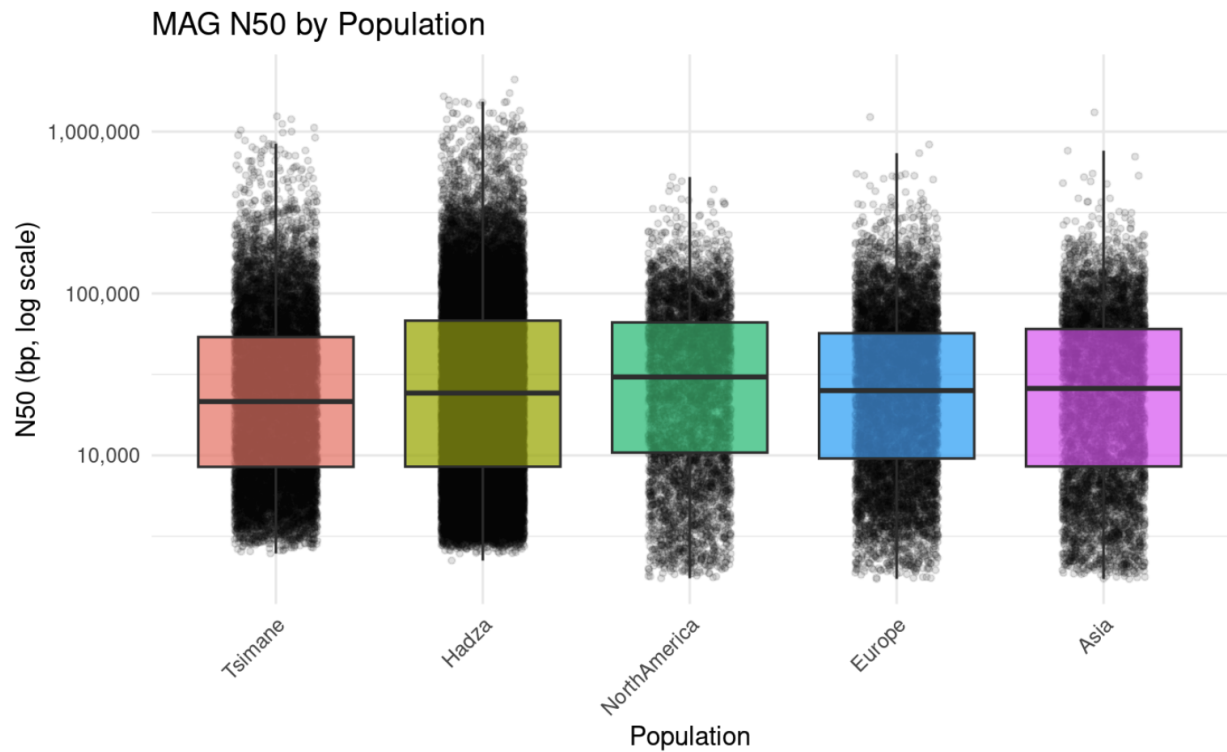


Figure S3: Comparison of MAG N50 values across populations. Boxplots showing the distribution of N50 values for metagenome-assembled genomes (MAGs) from five human populations: Tsimane ($n = 12,746$), Hadza ($n = 32,034$), Asia ($n = 6,426$), Europe ($n = 8,647$), and North America ($n = 5,463$). Individual data points are shown as jittered dots behind boxplots. Box centers represent medians, box limits represent the 25th and 75th percentiles (interquartile range), and whiskers extend to $1.5\times$ the interquartile range. N50 values are displayed on a log scale.

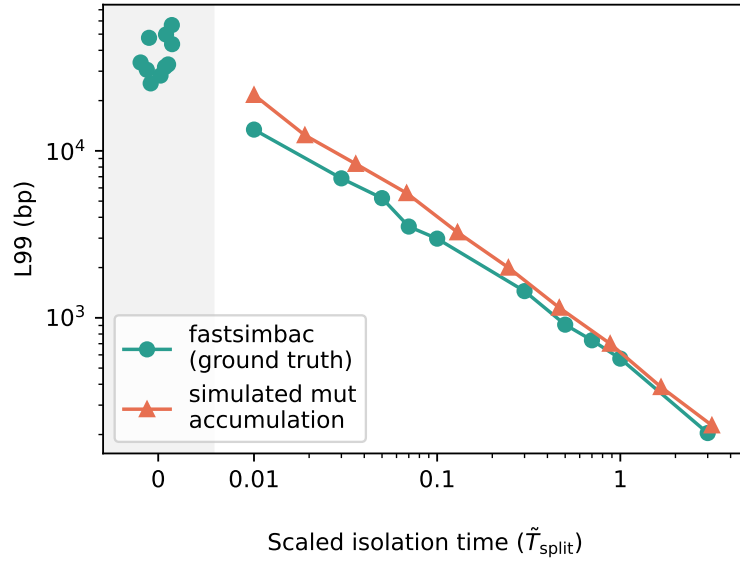


Figure S4: Comparison between the ground truth $L99-T_{\text{split}}$ relationship observed in FastSimBac simulations, and the approximate version obtained from the mutation accumulation approach in Section 4.3. The x-axis represents the scaled isolation time, and the y-axis shows L99 length. The ground truth, derived from full recombination-based FastSimBac simulations, is shown in cyan, while the approximate mutation accumulation simulation is shown in orange. The close agreement between the two curves suggests that mutation accumulation alone provides a reasonable first-order approximation of IBS tract dynamics over time, particularly at longer isolation times.

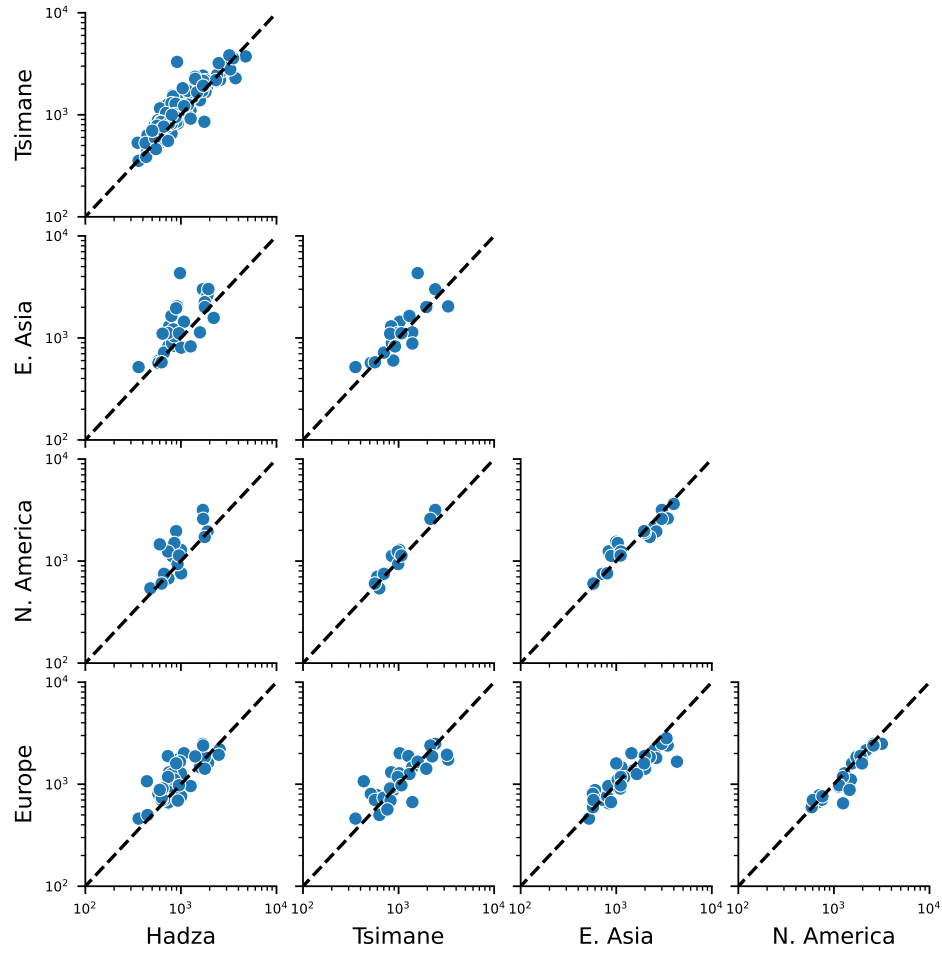


Figure S5: Scatterplots comparing the observed within-population L99 lengths between different pairs of populations. Each point indicates a single species. Dashed lines indicate the equality line ($y = x$). For each panel, only species with more than 200 within-population pairs in both populations are included.

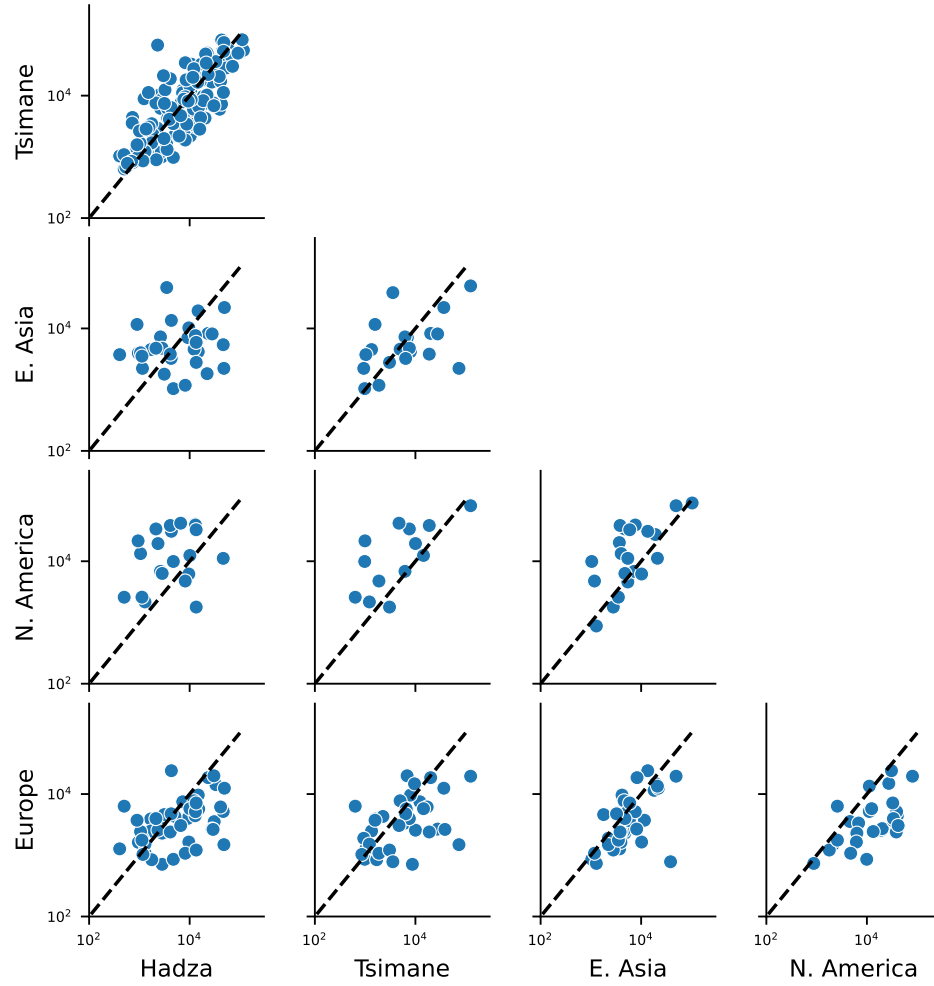


Figure S6: Analogous to Fig. S5, but without filtering clonal strain pairs sharing more than 10% identical genes. The reduced correlation across populations indicates that within-population L99 estimates are sensitive to uneven sampling of closely related strains, highlighting the importance of applying this filtering step.

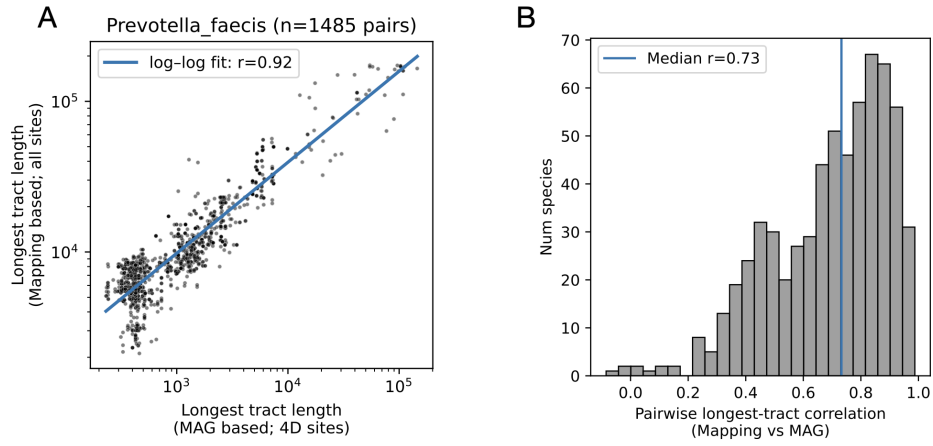


Figure S7: **(A)** Comparison of the longest IBS tract lengths inferred from MAG-based and mapping-based approaches for one representative species. Each point corresponds to a pair of samples for which both methods produced an estimate. Pearson's r denotes the correlation between the log-transformed tract lengths. **(B)** Distribution of correlation coefficients (Pearson's r) across all species, computed analogously to panel A. The distribution peaks near $r \approx 0.9$, indicating generally strong agreement between the two approaches.

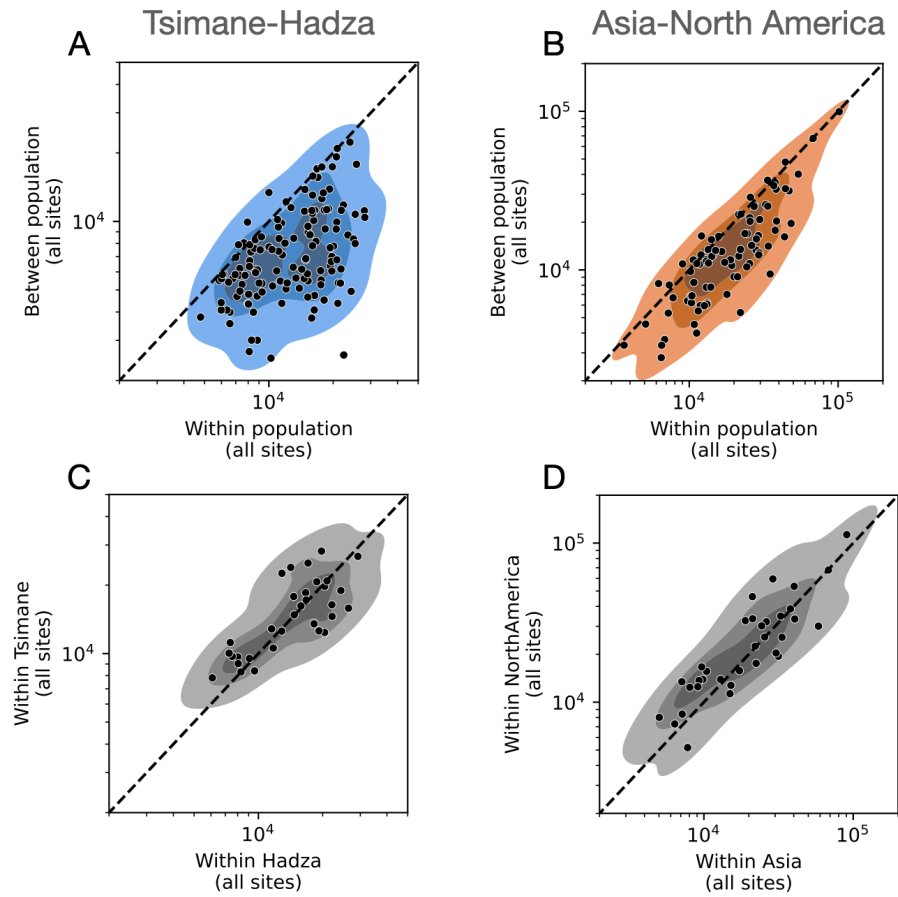


Figure S8: **(A–B)** Scatter plots comparing between-population and within-population L99 values across species, analogous to Fig. 3C of the main text but computed using mapping-based IBS tracts. The mapping-based analysis reproduces the qualitative pattern of stronger microbial isolation between Asia and North America than between the Tsimane and Hadza. **(C–D)** Within-population comparisons of mapping-based L99 values, analogous to Fig. S5.

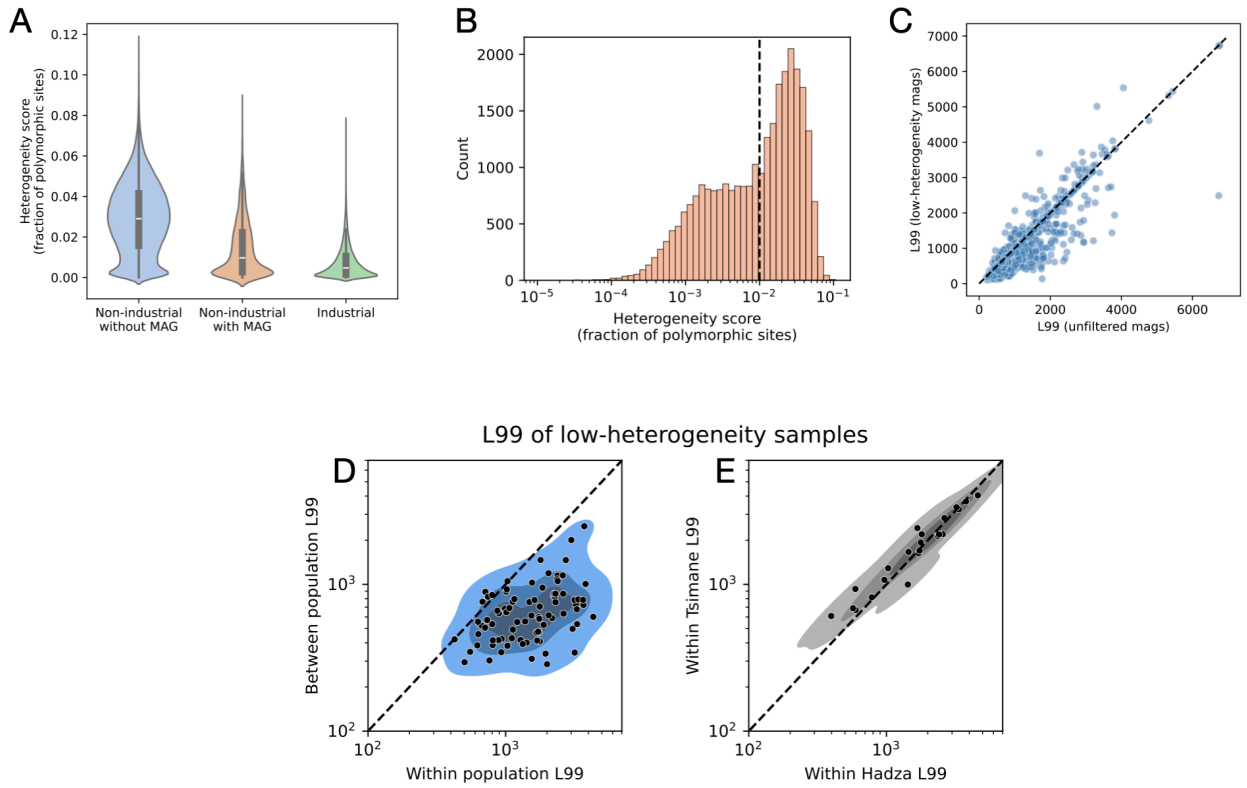


Figure S9: **(A)** Violin plots showing the distribution of strain-heterogeneity scores across species and samples (Section 4.6). The non-industrial set includes all Tsimane and Hadza samples from our main analysis, and the industrial set includes the Asia and North America samples. **(B)** Distribution of heterogeneity scores for species–sample pairs from non-industrial populations that yielded a MAG. The dashed line indicates the threshold used to define low-heterogeneity samples. **(C)** Scatter plot comparing L99 values computed from all MAGs versus those computed using only MAGs derived from low-heterogeneity samples. **(D–E)** Between-population and within-population comparisons of L99 values for low-heterogeneity samples, analogous to Fig. 3C in the main text. Restricting the analysis to low-heterogeneity samples does not alter the qualitative pattern of strong microbial isolation between the Tsimane and Hadza.

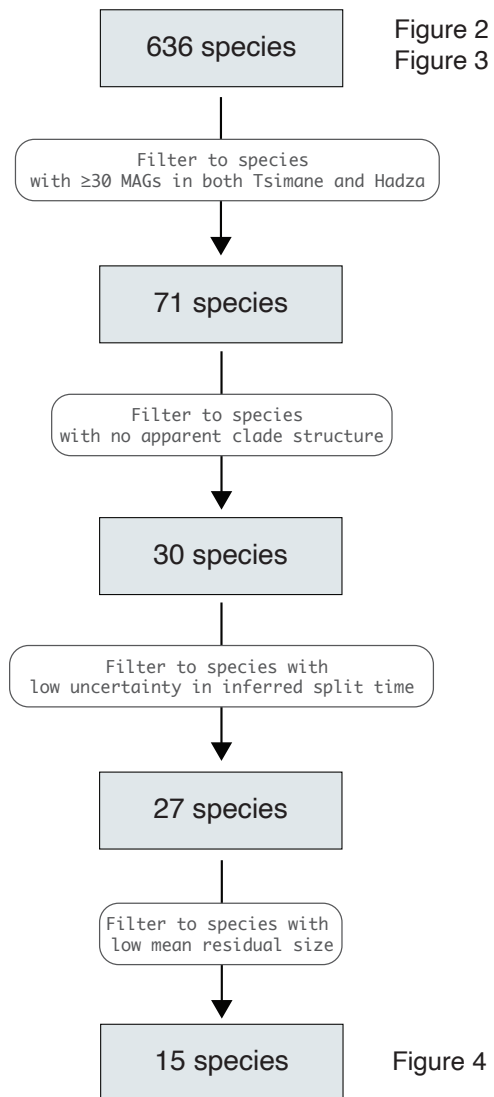


Figure S10: Flowchart summarizing the data filtering steps used for the Moments analysis (Section 5.2).

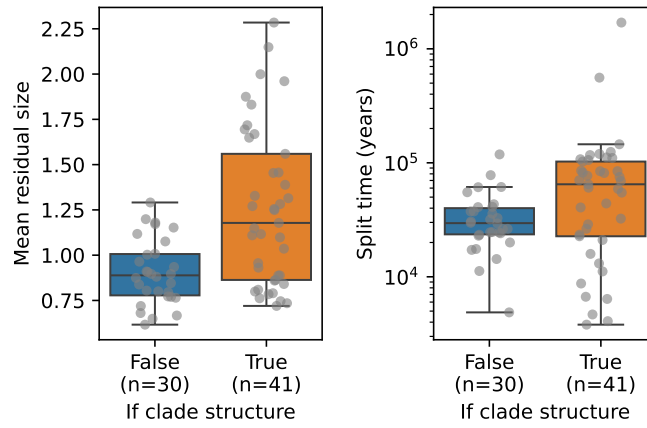


Figure S11: Box plots comparing the mean residual sizes (left) and inferred split times (right) of demographic model fits for species with and without clade structure, as identified by our clade detection method. Gray dots indicate individual species. The sample size n denotes the number of species in each bin (note: these species have not yet been filtered for parameter uncertainty or residual size, as described in Section 5.2). These data show that species without clade structure exhibit smaller residuals, indicating better alignment with the assumptions of the simple demographic model in Section 5.1. Species with clade structure also show higher inferred split times, potentially reflecting increased genetic diversity within species.

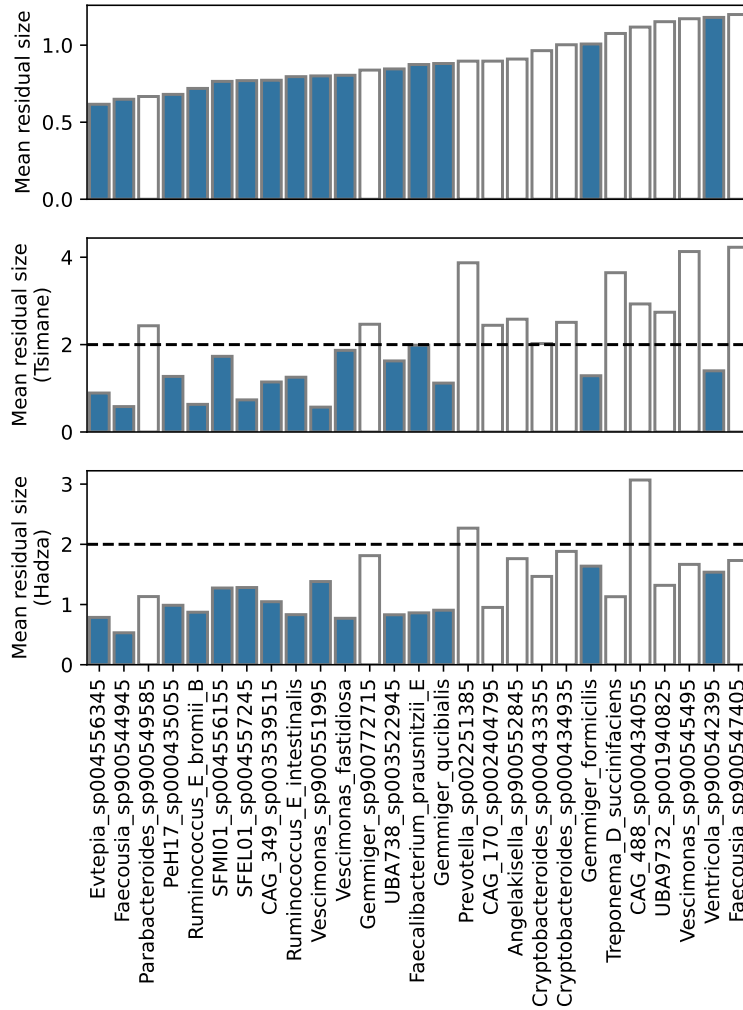


Figure S12: Bar plot showing the mean absolute residual sizes for the 27 species that passed the clade structure and uncertainty filters. The top panel shows the mean residual for the joint SFS, while the bottom two panels show the mean residual of the single-population (marginalized) site frequency spectrum. The dashed line represents the residual size filter threshold of 2. Solid bars denote species that passed the residual size filter.

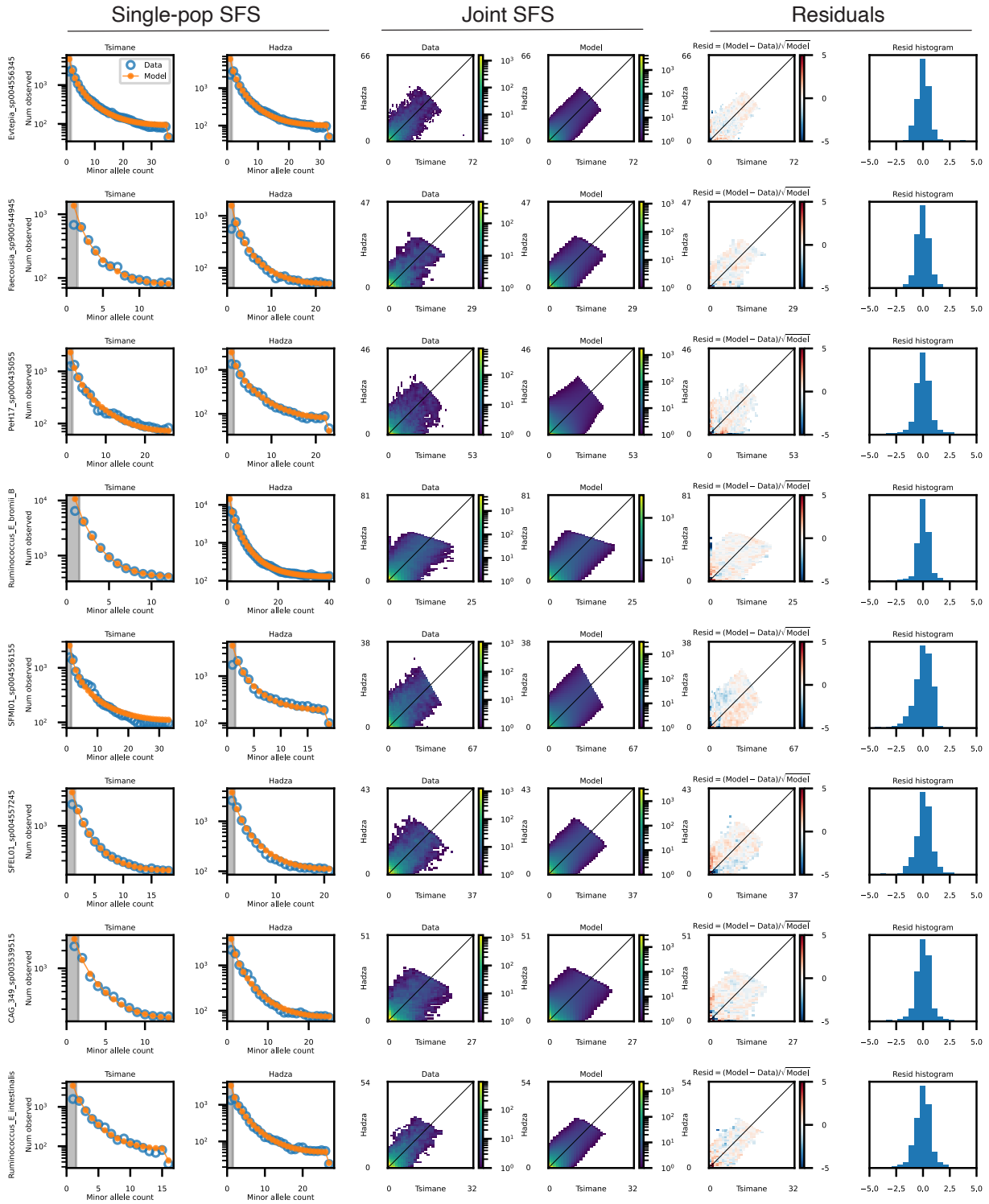


Figure S13: Summaries of demographic model fits inferred by Moments for 8 of the 15 analyzed species. The remaining species are shown in Fig. S14. Each row represents one species. The first two columns show the single-population (marginalized) site frequency spectrum (SFS) for the Tsimane and Hadza microbial populations, respectively. The shaded region indicates singleton SNVs (minor allele count = 1), which are masked during model fitting. Columns 3 and 4 show heatmaps of the joint SFS between Tsimane and Hadza microbial populations for observed data and model predictions. Column 5 shows heatmap of residuals between the model predictions and observed data. Column 6 shows a histogram of the residual values from column 5, with the y-axis representing probability density on a linear scale (ticks omitted for compactness).

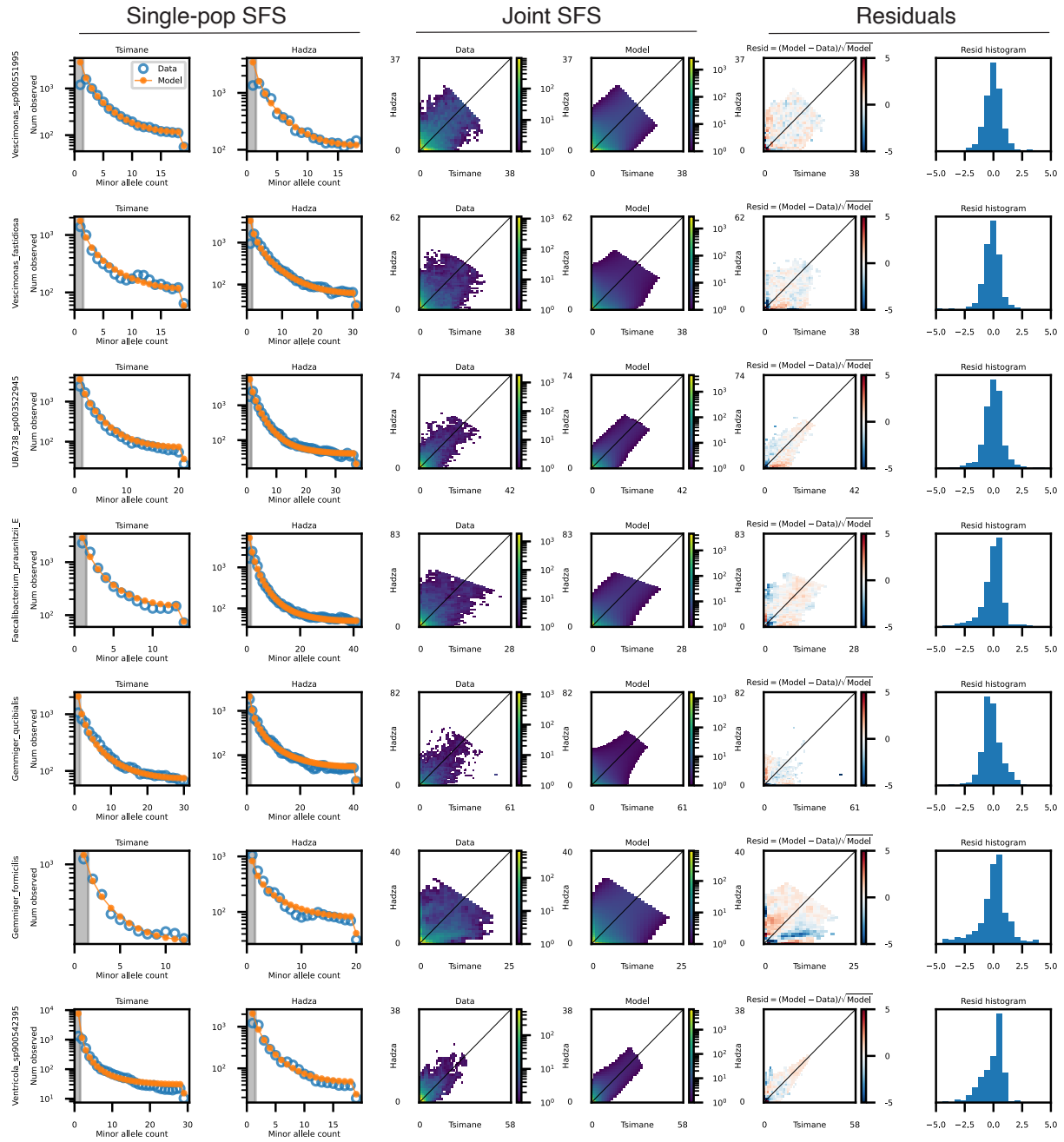


Figure S14: Same as in Fig. S13, but for the remaining seven species.

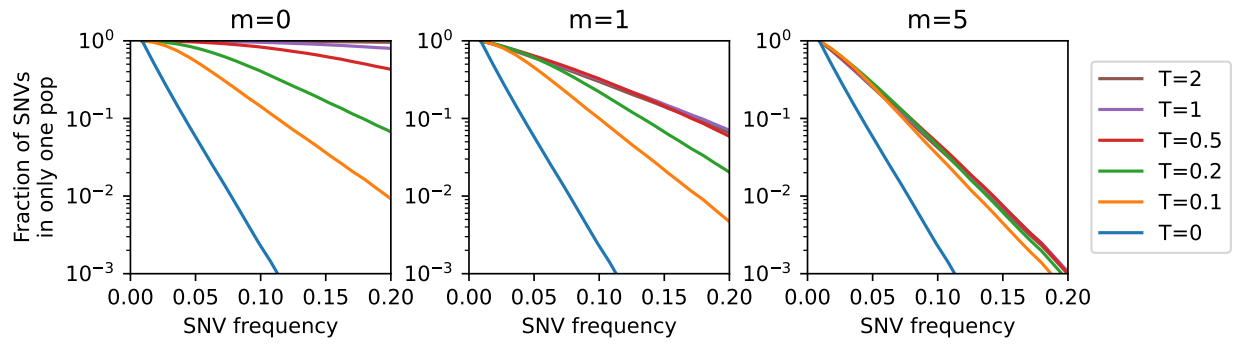


Figure S15: Plots showing the predicted fraction of population-specific SNVs for simulated data, analogous to Figure 4B in the main text. Curves are computed using the SFS simulated by the `split.mig` demographic model, with parameters $\nu_1 = \nu_2 = 2$ and varying split times and migration rates. Sample sizes were chosen to match those of the species *UBA738 sp003522945*. The parameter m represents the scaled migration rate, defined as $2N_a m_0$.

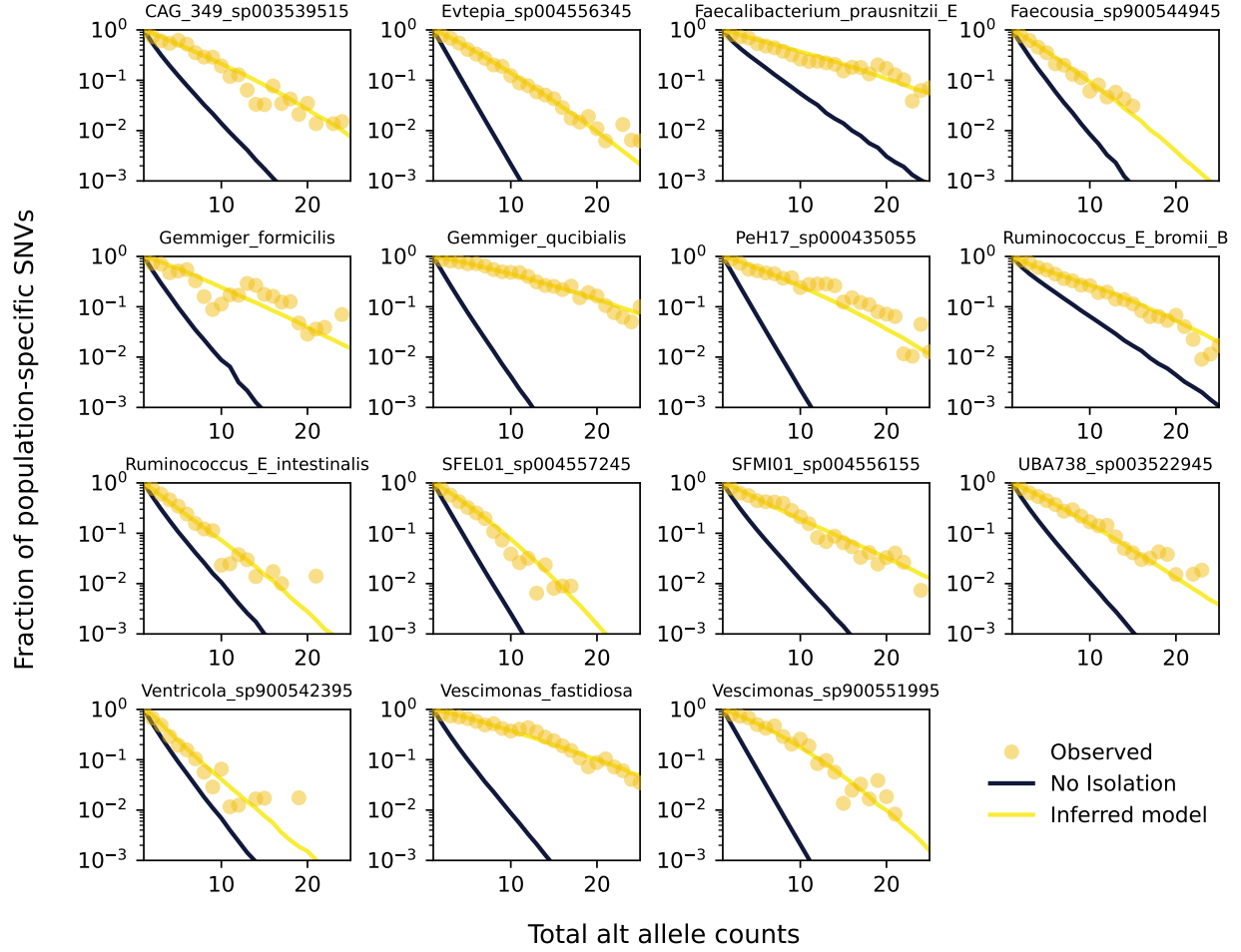


Figure S16: Analogous to Figure 4B in the main text. The yellow curve represents the prediction from the demographic model inferred using the full SFS, while the black curve corresponds to the null scenario of no population split. The null curve is computed by modeling the distribution of alternative alleles between populations using a binomial distribution, weighted by their relative population sizes. This is equivalent to randomly reshuffling alternative alleles between populations, simulating a scenario of free migration.