Demographic inference using the SFS with `moments` and `demes`

Nicholas W. Collier[1] and Aaron P. Ragsdale[1,*]

[1]Department of Integrative Biology, University of Wisconsin–Madison

[*]apragsdale@wisc.edu

ORCID IDs: 0009-0005-0385-9798 (NWC), 0000-0003-0715-3432 (APR)

August 26, 2025

Abstract

The site frequency spectrum (SFS) describes the distribution of allele frequencies in one or more populations and sees wide use in evolutionary inference. Here we show how to use `moments`, a Python library for working with the SFS and other measures of genetic diversity, and `demes`, a standard format for parameterizing population models, to infer demographic history. Following an introduction to SFS inference, we illustrate the general usage of `moments` and its `demes` interface, and then show two specific examples, using both simulated and human genetic data. We conclude with caveats and considerations regarding the `moments` library and SFS inference more broadly.

Keywords: Demographic inference, site frequency spectrum, diffusion equation

Introduction

The genetic composition of a sample of individuals is shaped by their genome biology and evolutionary history. Variation resulting from this history can be fully represented by the ancestral relationships among samples at each locus in the genome and how those gene-genealogical relationships change along a chromosome due to recombination (that is, by information stored in the Ancestral Recombination Graph; see chapters 1, 8 and 11, and [1]). However, the ARG can be large and unwieldy, and methods for reconstructing history directly from the ARG, while showing promise (e.g., [2–4]), are in their infancy and so far limited in application and scalability. Instead, evolutionary inference using informative summaries of genetic variation remains a tractable and powerful alternative for learning parameters of population history, natural selection and genome biology.

One such summary that has seen wide use is the site frequency spectrum (SFS), which stores the counts (observed or expected) of alleles carried by a given number of genomes in a set of samples. Like any summary of the data, the SFS discards some information stored in the ARG – in this case, loci are treated independently so that haplotypic information is lost. Even so, the relative densities of allele frequencies and the overall scale of the SFS are both sensitive to demographic and non-demographic processes.

Past population processes, including changes in population size, splits, gene flow and population structure, impact the SFS in predictable ways. A population size expansion (or contraction) leads to a relative excess (or deficit) of low-frequency variants, for example, and commonly used statistics measuring population divergence, such as F_{ST}, are themselves summaries of the SFS. Thus, the development of methods to learn demographic history from the SFS came soon after the publication of whole-genome sequencing from multiple individuals [5, 6]. More recent methodological advances allow for increased sample sizes and numbers of populations, providing a rich ecosystem of software for SFS-based demographic inference [7–13]. We return to this in the final section: Considerations and Caveats.

This chapter focuses on multi-population demographic inference from the SFS using `moments` [10]. Briefly, `moments` computes the expected SFS by numerically solving a system of ordinary differential equations closely related to the diffusion approximation [7]. The flexibility of the numerical methods in `moments` allows for non-constant population size histories including up to five populations at a time, with gene flow. Selection can also be incorporated, allowing for recessive, additive or dominant selection, or over- or underdominance at a locus. This differs from most coalescent-based approaches (e.g., [8, 13, 14]), which are typically limited to neutrality.

Many evolutionary mechanisms are known to affect allele frequency dynamics and the SFS. This presents a challenge for demographic inference, as the observed SFS may be distorted by non-demographic processes. However, this also presents an opportunity to learn about different evolutionary processes, if demographic history can be controlled for. The SFS has been used to infer the distribution of fitness effects of new mutations, primarily for nonsynonymous variation in coding regions (e.g., [15–17]); to scan for historical selective sweeps (e.g., [18, 19]); and to understand how selection on quantitative traits impacts the genetic architecture underlying those traits (e.g., [20, 21]). The SFS has also been used to infer relative mutation rates and how they have changed over time [14]. Each of these analyses typically requires partitioning the data in some way – such as by functional annotation, mutation context, local recombination rate – so that comparisons can be made across SFS from different classes of mutations.

Obtaining and installing the software

moments is available through pypi as moments-popgen (using pip install moments-popgen). Source code can be downloaded from https://github.com/MomentsLD/moments, and documentation is hosted at https://momentsld.github.io/moments/. The detailed online examples accompanying this chapter can be found at https://github.com/StatisticalPopulationGenomics-2ndEd/moments/, which uses moments version 1.4 and demes version 0.2.

Interfacing moments with demes

Specifying demographic models requires defining populations (or "demes") and their relationships via splits and gene flow. demes provides a standardized and accessible format for defining demographic models [22], and has been adopted by widely used software for population genetics simulation and inference (including msprime [23], momi [13], fwdpy11 [24], and GADMA [25]). demes encourages interoperability and reuse of code across software, ease of model implementation and reduction of errors.

Demographic models are specified in YAML format (see [22] and the associated documentation). To illustrate, the two-population split-with-migration model depicted in Figure 1A is specified as below:

```yaml
description: split-with-migration example model
generation_time: 29
time_units: years
demes:
- name: ancestral
  epochs:
  - {start_size: 15000, end_time: 300000}
  - {start_size: 25000, end_time: 75000}
- name: popA
  ancestors: [ancestral]
  epochs:
  - {start_size: 30000}
- name: popB
  ancestors: [ancestral]
  epochs:
  - {start_size: 1200, end_size: 14000}
migrations:
- demes: [popA, popB]
  rate: 5e-5
```

In moments, a demes-specified demographic model can be directly used to compute expectations for either the SFS or multi-population LD statistics [26, 27]. For the SFS, we simply need to load the demographic model using demes and then specify the number of haploid samples to draw and from which populations.

```
import demes, moments
```

```
g = demes.load('model.yaml')
samples = {'popA': 60, 'popB': 60}
fs = moments.Demes.SFS(g, samples=samples)
```

Here, we have not specified a mutation rate. The default behavior sets $u = 1$, so that $\theta = 4N_e u = 4N_e$, where N_e is the initial size of the ancestral population. With a known mutation rate, the SFS will be properly scaled by passing the mutation rate as a keyword argument: `moments.Demes.SFS(g, samples=samples, u=u)`. This is equivalent to scaling the output SFS above by multiplying by `u`, as `u*fs`.

Examples: inference in `moments` using `demes`

Below, we show two examples of multi-population demographic inference using the joint SFS. The first example simulates data with known demographic parameters, which we attempt to reinfer using both the original model and simpler misspecified models. In the second example, we infer a relatively simple Neanderthal-human history using data from two human populations and a high-coverage ancient sample from the Neanderthal lineage. For both of these examples, we briefly describe data processing and inference setup, emphasizing the high-level components of the analyses. For detailed information about each example, including managing data, specifying models, performing inference, computing confidence intervals and visualizing fits, we refer readers to the accompanying GitHub repository (`https://github.com/StatisticalPopulationGenomics-2ndEd/moments`), which is maintained with up-to-date versioning.

In order to optimize parameters in a `demes` model, we need a way to specify which parameters should be fitted. For this, we use a separate YAML-formatted file to define `parameters`, each of which points to desired value(s) in the demographic model and for which lower and upper bounds may be set. This file also specifies any relative `constraints` between pairs of parameters, to ensure the optimization routine only explores valid model space. Below, we have included a truncated parameters file – the full example is available on GitHub (URL above).

```
parameters:
- name: T0
  values:
  - demes:
      ancestral:
        epochs:
          0: end_time
  lower_bound: 0
  upper_bound: 2e5
- name: T1
  values:
```

```
  - demes:
      ancestral:
        epochs:
          1: end_time
  lower_bound: 0
  upper_bound: 1e6
- name: Ne
  values:
  - demes:
      ancestral:
        epochs:
          0: start_size
  lower_bound: 1e2
  upper_bound: 1e6
...
constraints:
- params: [T0, T1]
  constraint: greater_than
```

Taking the demographic model, parameters file, and observed SFS together, we can perform inference using the `moments.Demes.Inference.optimize()` function. The initial parameter guesses are given by the input demographic model, which may be perturbed using a keyword argument. The likelihood surface may be rugged or flat in some dimensions, and optimization routines sometimes fail to converge to the global (or sometimes even local) optimum. Thus, we often try many different initial guesses for parameter values, which can be randomly assigned using the `perturb` keyword argument. Running multiple iterations of optimization with different optimization functions is also recommended to ensure convergence.

Any value in the demographic model that is not specified in the parameters file will remain unchanged, and is therefore treated as a fixed parameter. Here, we also assume we have an estimate for the total mutation rate, U.

```
import moments

data = moments.Spectrum.from_file('data.fs')
U = 1e-8 * 5e8  # per-base mutation rate times the total length

model_file = 'model.yaml'
params_file = 'params.yaml'
output = 'model.fit.yaml'

ret = moments.Demes.Inference.optimize(
    model_file, params_file, data,
    perturb=1, uL=U, output=output
)
```

Confidence intervals may also be computed using `moments.Demes.Inference.uncerts()`, once a (locally) optimal model has been found. The following will emit a tab-separated file containing parameter names,

best-fit values and estimated standard errors.

```
log_file = 'log.txt'

uncerts = moments.Demes.Inference.uncerts(
    output, params_file, data,
    uL=U, output=log_file
)
```

Uncertainties may be computed using either the Fisher Information Matrix, which tends to underestimate confidence intervals due to nonindependence of linked sites, or using the Godambe Information Matrix, which uses a set of bootstrapped datasets to correct for this nonindependence in the data [28]. Example 1 in the associated GitHub repository demonstrates this using a simulated dataset.

Example 1: Inferring parameters in a simulated split-with-migration model

Using the split-with-migration model defined above (illustrated in Figure 1A), we simulated data for 30 diploid individuals from both populations using `msprime` [23]. These simulations consided of 500 regions, each of length 1 Mb, with per-base recombination and mutation rates of 10^{-8}. The joint SFS was computed by using `tskit`'s `ts.allele_frequency_spectrum()` function separately on each replicate region and summing across regions to find the total SFS across 500 Mb of data (Figure 1B–D).

We fitted three demographic models to the simulated data. In addition to reinferring the parameters from the simulated model, we fitted two simpler models to the same data. These had fewer parameters, both omitting the size change deeper in time in the ancestral population. This is meant to crudely mimic the scenario that we often face, in which the true history is more complicated than the parameterized model. It also highlights biases in inferences that can arise when features of the true history are not included.

When fitting the two misspecified models that do not allow for size changes in the ancestral population, this split time is inferred to be substantially deeper in the past than the true split time. This effect, as demonstrated in [29], is due to the decreased coalescence rate within the ancestral population between the time of the expansion and divergence of the descendent populations. These two misspecified models (Figure 1H,I) also provide a worse fit to the data, as seen by the lower log-likelihoods and the large residuals between model predictions and data (Figure 1E-G).

Example 2: Inferring Neanderthal-human demographic parameters

We used publicly available data to infer a demographic model for two modern human populations (a northern European and western African population) and a Neanderthal individual. As shown below, even with just a

single sample from the Neanderthal lineage, we are able to recover key parameters such as the Neanderthal-human divergence time and the fraction of ancestry contributed to modern Europeans by Neanderthals.

We obtained modern human genome sequences from a recent resequencing of the 1000 Genomes Project cohort [30], and used the high-coverage sequence of the Vindija Neanderthal genome [31]. From the 1000 Genomes cohort we took 85 MSL (Mende from Sierre Leone) and 91 GBR (British from England and Scotland) diploid sequences as our modern human samples. We filtered out sites that lay outside the 1000 Genomes 'strict' mask, were not genotyped in four high-coverage archaic genome sequences, lacked high-confidence ancestral state assignments, or fell within 10 kb of exonic or promoter regions. This left us with approximately 960 Mb of well-characterized and putatively neutrally-evolving sequence. We estimated the genome-wide SFS using built-in `moments` parsing functions. Details concerning data processing (including our liftover of the Vindija sequence to a more recent genome build) can be found in the linked GitHub repository.

The model that we wished to fit is relatively simple but still has many parameters (11 in total), so we constructed it in stages by first fitting marginal one- and two-population models. When we want to infer a complex demographic history involving multiple populations, it is helpful to start simple by modeling subsets of the full data and sequentially adding complexity. This procedure gives us initial guesses for parameter values in more complex models and may help suggest which features are well-supported by the data. Here, we began by fitting one-population models with ancestral and recent size changes for the MSL and GBR populations. In the MSL model, we allowed three epochs of piecewise-constant effective size. We attributed to GBR two constant-size epochs followed by a sharp contraction and exponential growth beginning $\sim$60 kya to represent the out-of-Africa bottleneck and subsequent rapid population expansion. Throughout, we used $u = 1.5 \times 10^{-8}$ as an estimate of the mutation rate per generation and 29 years as the average generation time. YAML files encoding the models and parameters in this section can be found in the linked GitHub repository.

We next stitched these one-population models together and refitted the resulting demographic model to get a two-population model that included continuous symmetric migration between MSL and GBR and an ancestral size expansion. Following this, we added a Neanderthal branch and fitted the Neanderthal-human divergence time, the Neanderthal effective size, and the Neanderthal-GBR pulse proportion, leaving all other parameters fixed. We also fitted the divergence time between the sampled Vindija Neanderthal lineage and the Neanderthal lineage which admixed with non-African humans. Finding that this parameter tended to converge to its lower bound (imposed by the estimated time of death of the Vindija Neanderthal, 55 kya), we fixed it to a value supported by literature (90 kya, [31]) in a second round of optimization. We also fixed the time of the Neanderthal-GBR pulse to 50 kya (at the upper range of values inferrred by [32]). Fixing

parameters for which good estimates are available in the literature is often reasonable in the first stages of model optimization, and we expected this value to be especially poorly constrained due to the small pulse proportion and Neanderthal sample size. In the second round of optimization of the three-population model, we refitted all parameters aside from the two times mentioned above to refine the estimates.

After this refitting step, we were left with parameters in good agreement with those inferred by prior authors, with a Neanderthal-human divergence time of 547 kya, MSL-GBR split time of 77 kya, and Neanderthal-GBR pulse proportion of 2.7 percent. We observe some large, systematic residuals in our best-fit model (Figure 2G-I), suggesting model misspecification that could be improved through further exploration of model space.

We also fitted a nested model without the Neanderthal-GBR pulse, which had a substantially lower likelihood than the pulse model (log-likelihood [LL] $\approx -127,552$, against LL $\approx -25,103$ for the pulse model). We did not perform a formal hypothesis test; the null distribution of the likelihood-ratio test (LRT) statistic is overdispersed under a composite likelihood, inflating the rate of false positives and preventing its use in this setting. `moments` does incorporate a function (`moments.Godambe.LRT_adjust()`) which computes a multiplicative adjustment to the LRT, intended to restore the expected null distribution, using first-order moment matching [28]. This method requires bootstrapping over the data. Otherwise, we could conduct a hypothesis test by fitting the nested and pulse models to a large number of bootstrap samples and calculating the rate of rejection across them. This has the benefit of producing a distribution of (nonindependent) LRT statistics, but is time-consuming and computationally expensive.

Considerations and caveats

The site frequency spectrum is just one of many options for inferring evolutionary parameters from genetic data. It is straightforward to compute from data, and there are many methods for predicting the SFS, making it an appealing framework for demographic inference. As shown here in two examples, we have incorporated the `demes` format for specifying demographic models into `moments` to easily compute the SFS and perform demographic inference, including computing uncertainties.

`moments` is limited in the size of the dataset that it can consider. Sample sizes can be large (in the hundreds), especially when considering models with one or few populations, but memory issues will develop when sample sizes are extreme or as the number of populations grows. We are limited to a maximum of five populations existing at any one time, including ancient samples. Other SFS-based methods, such as `momi` [12, 13], can include additional populations, typically with smaller sample sizes.

When inferring models with multiple populations, especially four or more, both the numbers of parameters

and the space of plausible model topologies can grow rapidly. It can therefore become challenging for the optimization routines to find optima in parameter space, and overparameterization can become a problem. One hint that models are overparameterized (or models misspecified) is runaway behavior of parameter values, in which they converge to unreasonably small or large values (or approach the imposed bounds). Starting with simple parameterizations and sequentially adding complexity helps to alleviate this issue. The challenge of large model space is more difficult to address, and must be handled through building and testing a wide range of plausible model topologies. While time-consuming, a broad exploration of model space, especially when considering many populations, is crucial for evaluating robustness of particular features of inferred history.

Finally, natural selection distorts allele frequencies not only at sites directly under selection, but also in linked regions of the genome through background selection. This can have a sizeable impact on the SFS and lead to biases in demographic inference [33, 34]. In gene-sparse species, such as primates, considering data from regions away from functionally constrained elements is feasible, as we did in example 2 here. Alternatively, the genome may be partitioned by B-value (a measure of the local strength of background selection, [35]) and regions experiencing low levels of background selection may be used (for an example, see [36]). In some species with gene-dense genomes, like *Drosophila* species, this may be infeasible, and interspersed putatively neutral variation must be used, such as synonymous variation or short introns, which will still experience background selection. The appropriate strategy will depend on the genome architecture of the species and genomic resources available.

References

[1] Rasmus Nielsen, Andrew H Vaughn, and Yun Deng. Inference and applications of ancestral recombination graphs. *Nature Reviews Genetics*, 26(1):47–58, 2025.

[2] Débora YC Brandt, Xinzhu Wei, Yun Deng, Andrew H Vaughn, and Rasmus Nielsen. Evaluation of methods for estimating coalescence times using ancestral recombination graphs. *Genetics*, 221(1):iyac044, 2022.

[3] Caoqi Fan, Jordan L Cahoon, Bryan L Dinh, Diego Ortega-Del Vecchyo, Christian Huber, Michael D Edge, Nicholas Mancuso, and Charleston WK Chiang. A likelihood-based framework for demographic inference from genealogical trees. *bioRxiv*, 2023.

[4] Débora YC Brandt, Christian D Huber, Charleston WK Chiang, and Diego Ortega-Del Vecchyo. The

promise of inferring the past using the ancestral recombination graph. *Genome biology and evolution*, 16(2):evae005, 2024.

[5] Gabor T Marth, Eva Czabarka, Janos Murvai, and Stephen T Sherry. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics*, 166(1):351–372, 2004.

[6] Scott H Williamson, Ryan Hernandez, Adi Fledel-Alon, Lan Zhu, Rasmus Nielsen, and Carlos D Bustamante. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences*, 102(22):7882–7887, 2005.

[7] Ryan N Gutenkunst, Ryan D Hernandez, Scott H Williamson, and Carlos D Bustamante. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS genetics*, 5(10):e1000695, 2009.

[8] Laurent Excoffier and Matthieu Foll. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27(9):1332–1334, 2011.

[9] Simon Gravel, Brenna M Henn, Ryan N Gutenkunst, Amit R Indap, Gabor T Marth, Andrew G Clark, Fuli Yu, Richard A Gibbs, 1000 Genomes Project, Carlos D Bustamante, et al. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*, 108(29):11983–11988, 2011.

[10] 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, 2017.

[11] Aaron P Ragsdale, Claudia Moreau, and Simon Gravel. Genomic inference using diffusion models and the allele frequency spectrum. *Current Opinion in Genetics & Development*, 53:140–147, 2018.

[12] Jack Kamm, Jonathan Terhorst, Richard Durbin, and Yun S Song. Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, 115(531):1472–1487, 2020.

[13] Enes Dilber and Jonathan Terhorst. Faster inference of complex demographic models from large allele frequency spectra. *bioRxiv*, 2024.

[14] William S DeWitt, Kameron Decker Harris, Aaron P Ragsdale, and Kelley Harris. Nonparametric coalescent inference of mutation spectrum history and demography. *Proceedings of the National Academy of Sciences*, 118(21):e2013798118, 2021.

[15] Adam Eyre-Walker, Megan Woolfit, and Ted Phelps. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, 173(2):891–900, 2006.

[16] Adam R Boyko, Scott H Williamson, Amit R Indap, Jeremiah D Degenhardt, Ryan D Hernandez, Kirk E Lohmueller, Mark D Adams, Steffen Schmidt, John J Sninsky, Shamil R Sunyaev, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS genetics*, 4(5):e1000083, 2008.

[17] Bernard Y Kim, Christian D Huber, and Kirk E Lohmueller. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics*, 206(1):345–361, 2017.

[18] Yuseob Kim and Wolfgang Stephan. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160(2):765–777, 2002.

[19] Rasmus Nielsen, Scott Williamson, Yuseob Kim, Melissa J Hubisz, Andrew G Clark, and Carlos Bustamante. Genomic scans for selective sweeps using snp data. *Genome research*, 15(11):1566–1575, 2005.

[20] Roshni A Patel, Clemens L Weiß, Huisheng Zhu, Hakhamanesh Mostafavi, Yuval B Simons, Jeffrey P Spence, and Jonathan K Pritchard. Conditional frequency spectra as a tool for studying selection on complex traits in biobanks. *bioRxiv*, 2024.

[21] Aaron P Ragsdale. Archaic introgression and the distribution of shared variation under stabilizing selection. *bioRxiv*, pages 2024–08, 2024.

[22] Graham Gower, Aaron P Ragsdale, Gertjan Bisschop, Ryan N Gutenkunst, Matthew Hartfield, Ekaterina Noskova, Stephan Schiffels, Travis J Struck, Jerome Kelleher, and Kevin R Thornton. Demes: a standard format for demographic models. *Genetics*, 222(3):iyac131, 2022.

[23] Franz Baumdicker, Gertjan Bisschop, Daniel Goldstein, Graham Gower, Aaron P Ragsdale, Georgia Tsambos, Sha Zhu, Bjarki Eldon, E Castedo Ellerman, Jared G Galloway, et al. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3):iyab229, 2022.

[24] Kevin R Thornton. Polygenic adaptation to an environmental shift: temporal dynamics of variation under gaussian stabilizing selection and additive effects on a single trait. *Genetics*, 213(4):1513–1530, 2019.

[25] Ekaterina Noskova, Nikita Abramov, Stanislav Iliutkin, Anton Sidorin, Pavel Dobrynin, and Vladimir I Ulyantsev. Gadma2: more efficient and flexible demographic inference from genetic data. *GigaScience*, 12:giad059, 2023.

[26] Aaron P Ragsdale and Simon Gravel. Models of archaic admixture and recent history from two-locus statistics. *PLoS genetics*, 15(6):e1008204, 2019.

[27] Aaron P Ragsdale and Simon Gravel. Unbiased estimation of linkage disequilibrium from unphased data. *Molecular Biology and Evolution*, 37(3):923–932, 2020.

[28] 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, 2016.

[29] Paolo Momigliano, Ann-Britt Florin, and Juha Merilä. Biases in demographic modeling affect our understanding of recent divergence. *Molecular biology and evolution*, 38(7):2967–2985, 2021.

[30] Marta Byrska-Bishop, Uday S Evani, Xuefang Zhao, Anna O Basile, Haley J Abel, Allison A Regier, André Corvelo, Wayne E Clarke, Rajeeva Musunuri, Kshithija Nagulapalli, et al. High-coverage whole-genome sequencing of the expanded 1000 genomes project cohort including 602 trios. *Cell*, 185(18):3426–3440, 2022.

[31] Kay Prüfer, Cesare De Filippo, Steffi Grote, Fabrizio Mafessoni, Petra Korlević, Mateja Hajdinjak, Benjamin Vernot, Laurits Skov, Pinghsun Hsieh, Stéphane Peyrégne, et al. A high-coverage neandertal genome from vindija cave in croatia. *Science*, 358(6363):655–658, 2017.

[32] Arev P Sümer, Hélène Rougier, Vanessa Villalba-Mouco, Yilei Huang, Leonardo NM Iasi, Elena Essel, Alba Bossoms Mesa, Anja Furtwaengler, Stéphane Peyrégne, Cesare de Filippo, et al. Earliest modern human genomes constrain timing of neanderthal admixture. *Nature*, 638(8051):711–717, 2025.

[33] Gregory B Ewing and Jeffrey D Jensen. The consequences of not accounting for background selection in demographic inference. *Molecular ecology*, 25(1):135–141, 2016.

[34] Parul Johri, Kellen Riall, Hannes Becher, Laurent Excoffier, Brian Charlesworth, and Jeffrey D Jensen. The impact of purifying and background selection on the inference of population history: problems and prospects. *Molecular biology and evolution*, 38(7):2986–3003, 2021.

[35] Graham McVicker, David Gordon, Colleen Davis, and Phil Green. Widespread genomic signatures of natural selection in hominid evolution. *PLoS genetics*, 5(5):e1000471, 2009.

[36] Santiago G Medina-Muñoz, Diego Ortega-Del Vecchyo, Luis Pablo Cruz-Hervert, 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, 2023.

Figure 1: **Inference using simulated data.** (A) The true demographic model, including an expansion in the ancestral population and symmetric migration between the descendent populations. (B–D) The simulated data, showing both marginal and the joint SFS. (E–G) Residuals between the data SFS (in D) and the predicted SFS from the inferred demographic models in (H–J), resp. The two misspecified models (H and I) provide worse fits to the data and both overestimate the split time due to disallowing any size change in the ancestral population.

Figure 2: **Inference using human and Neanderthal data.** (A) The best-fit model including the Vindija Neanderthal and two human populations. (B, C) Marginal SFS fits between the model and empirical data are shown for MSL and GBR. (D–I) Marginal two-way spectra predicted by the model and their residuals against the data are plotted for the three pairs of populations MSL, GBR (D and G), MSL, Vindija (E and H), and GBR, Vindija (F, I).