An exact solution for the site-frequency spectrum of a selected allele

Matthew D. Schwartz[1] **and Michael M. Desai**[2,†]

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA, [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

ABSTRACT Natural selection removes the most damaging mutations fastest, so they are the rarest variants in a population. The site-frequency spectrum, which counts the variants found on one, two, or more sampled chromosomes, therefore carries the signature of selection, and its mathematical theory, the Sawyer–Hartl Poisson random field, has underpinned fitness-effect inference for three decades. That theory could never be evaluated exactly at strong selection and large sample size: the formulas span hundreds of orders of magnitude, the standard recursion is numerically unstable, and at the scale of modern catalogs floating-point evaluation fails outright. Working over the rational numbers, we evaluate the spectrum exactly at any selection strength for samples in the classical size range, and with controlled error at biobank sample sizes and for selection with arbitrary dominance. A fit to the site-frequency spectrum of the roughly 730,000 human exomes in the gnomAD database places the majority of mutations that disable a gene in the strongest-selection window the data can resolve, a regime that samples of classical size could not reach. Comparing the standard inference packages, polyDFE, dadi-based fitting, and fastDFE, with exact values, we find all three mostly reliable at the sample sizes for which they were developed, and we identify the parameter ranges in which each loses accuracy.

KEYWORDS site-frequency spectrum; distribution of fitness effects; Poisson random field; dominance; exact computation; gnomAD

1. Introduction

Understanding the role of natural selection in shaping the frequency of genetic variants within a population is a central goal in statistical genetics. One widespread approach to this question is to infer the distribution of fitness effects (DFE) of variants from observations of the site-frequency spectrum (SFS, defined as the number of polymorphic sites at which the derived allele is observed in i out of n chromosomes sampled from the population). The basic intuition underlying this approach is that purifying selection tends to reduce the frequency of these variants (analogously, positive selection also causes variants to spend less time at low frequencies), with stronger selection suppressing the frequency of variants to ever-lower frequencies. Thus the distribution of fitness effects acting on new variants can in principle be read out from the shape of the SFS.

In recent years, multiple studies have exploited the enormous sample sizes of current biobank-scale data sets to infer the DFE using this approach. The essential foundation for these methods is the Poisson random field model of Sawyer and Hartl (1992). This model describes how the frequency distribution of a variant is expected to depend on population size, mutation rate, and selection pressures in terms of an integral over diffusion sojourn times (Sawyer and Hartl 1992; Bustamante et al. 2001). Widely used computational methods such as polyDFE (Tataru et al. 2017), fit*dadi* (Kim et al. 2017) (built on *dadi*; Gutenkunst et al. 2009), and fastDFE (Sendrowski and Bataillon 2024) all evaluate this object, either directly or through simulation, and use it to fit distributions of fitness effects to observed site-frequency spectra. The modern picture of the human distribution of fitness effects comes from this class of methods (Boyko et al. 2008).

However, there is a key numerical difficulty in evaluating the integral at the heart of the Poisson random field model. The problem is that the integrand spans hundreds of orders of magnitude across the selection strengths of interest, and the standard recursion used to evaluate this integral alternates in sign. At the sample sizes of modern catalogs, floating-point evaluation of the exact theory becomes intractable. The computational packages described above manage this with various approxima-

Manuscript compiled: Friday 25th September, 2026
†Corresponding author: [to be set by the authors].

tion schemes, but the domain of validity of these approaches is difficult to determine from within floating-point arithmetic.

In this paper, we remove this numerical difficulty entirely by evaluating the Sawyer–Hartl spectrum exactly at any sample size and any selection strength, including selection with arbitrary dominance. Our key advance is to conduct these calculations in rational arithmetic, rather than floating point. In the classical range of sample sizes ($n \lesssim 2000$) and without dominance, our computation is exact. At biobank sample sizes and in the presence of arbitrary dominance, we instead evaluate the spectrum numerically, but using rational arithmetic with rigorous or validated error bounds.

Our results have two important applications. First, the exact values serve as a benchmark for the standard software. Specifically, we compare POLYDFE, FIT*dadi*, and FASTDFE with our exact results. We find these software packages are mostly reliable at the small sample sizes where they were developed, and identify the parameter ranges in which each loses accuracy. Second, our exact computation runs at biobank scale, where previous implementations currently do not (Section 5). As an example of this, we use our method to infer the strong-selection tail of the DFE from 730,000 exomes in the gnomAD data set. We also report two further results: (i) exact arithmetic lets us determine which selected spectra can be mimicked by a neutral model with some demographic history (or vice versa), and (ii) we use the X-chromosome data, combined with the exact dominance spectrum, to yield a one-sided bound on the dominance of loss-of-function coding mutations.

The paper is organized as follows. Section 2 reviews Poisson random field inference from the site-frequency spectrum and presents the exact computation: the model, the closed form, the evaluation routes and their error control, the extension to dominance, and the separation and representation tests used in the mimicry analysis. Section 3 applies the exact computation to the strong-selection tail of gnomAD. Section 4 establishes when a neutral demographic history can mimic selection and when it provably cannot. Section 5 compares the standard software with exact values, and Section 6 bounds dominance from the X chromosome. We close with a discussion of limitations and extensions in Section 7.

2. Inferring fitness effects from the site-frequency spectrum

The site-frequency spectrum (SFS), ξ_i, is the number of polymorphic loci at which the derived allele is carried by i of the n chromosomes sampled randomly from a population (Kimura 1969). In a neutral, randomly mating, panmictic population of constant size, the SFS follows the classical Kingman expectation (Kingman 1982), $\xi_i \propto 1/i$. Distortions of the SFS from this expectation are widely used to infer the action of other evolutionary forces (e.g. demographic history). Here, we focus on inferring selection pressures from the SFS.

2.1. Poisson random field expression for the SFS

The Poisson random field (PRF) model introduced by Sawyer and Hartl (1992) and Bustamante et al. (2001) is the foundation of many widely used approaches to this problem. This PRF model assumes a panmictic population of constant size with free recombination, so the expected SFS can be written as an integral over the expected distribution of allele frequencies at each site. Specifically, for a scaled selection coefficient $S = 4N_e s$, the expected number of sites where a derived allele appears i times in a sample of n chromosomes is

$$\xi_i(S) \;=\; \theta \binom{n}{i} \int_0^1 x^{i-1}(1-x)^{n-i-1} \, \frac{1-e^{-S(1-x)}}{1-e^{-S}} \, dx \,. \quad (1)$$

Throughout this work, we assume that $S = 4N_e s$ with $S > 0$ advantageous (and $S < 0$ deleterious), the SFS is unfolded with derived-allele polarization, and $\theta = 4N_e \mu$ per site; we set $\theta = 1$ except where stated. POLYDFE and FASTDFE use exactly these conventions; *dadi*'s $\gamma = 2N_A s$ corresponds to $S = 2\gamma$ (its heterozygote fitness is $1 + 2sh$).

The SFS across a set of sites which experience a range of selection pressures drawn from some DFE, $\phi_\omega(S)$, is then given by Eq. (1) mixed over S. The parameters ω describing the DFE can thus be estimated by maximizing a Poisson likelihood over the observed spectrum. This is the computational core of POLYDFE (Tataru et al. 2017), of FIT*dadi* (Kim et al. 2017), and of FASTDFE (Sendrowski and Bataillon 2024).

2.2. Challenges in evaluating the PRF site-frequency spectrum

We can find a closed form solution for Eq. (1) by substituting $u = 1 - x$. We find

$$\xi_i(S) \;=\; \theta \, \frac{n}{i(n-i)} \, \frac{1 - {}_1F_1(n-i;\, n;\, -S)}{1-e^{-S}} \,, \quad (2)$$

where ${}_1F_1$ is the confluent hypergeometric (Kummer) function. This expression is well known: in the methods literature it appears as Eq. (24) of Živković et al. (2015), and it was derived independently in the supplementary material of POLYDFE (File S1, §2 of Tataru et al. 2017), in both cases in passing and in neither case with any discussion of numerical error.

We can evaluate Eq. (2) by writing a three-term recurrence for $M_i \equiv {}_1F_1(n-i; n; -S)$, using Gauss's contiguous relations (classical identities linking Kummer functions whose parameters differ by one):

$$\begin{gathered} i\,M_{i+1} + (n - 2i - S)\,M_i - (n-i)\,M_{i-1} = 0 \,, \\ M_0 = e^{-S}, \qquad M_n = 1 \,. \end{gathered} \quad (3)$$

Because the two end values M_0 and M_n are known exactly, we can compute the whole vector of M values $(M_0, \ldots, M_n)$ in $\mathcal{O}(n)$ arithmetic operations by evaluating the recurrence starting from either end. However, this does not work in floating-point arithmetic, because the recurrence has two solutions. The desired physical solution is minimal in a direction set by the sign of S, and the companion undesired (unphysical) solution grows (Gautschi 1967). The essential problem is that a three-term recurrence has two independent solutions, and every rounding error made while iterating injects a small admixture of the undesired solution. In other words, the unphysical solution "leaks" into the physical one, and if the unphysical solution grows relative to the physical one this error is amplified at every subsequent step. This is true regardless of which end we start from in evaluating the recurrence (though one can be better than the other). We measured this amplification directly for this specific recurrence, and find that even starting from the "better" end, the unphysical solution gains a factor of 10^{23} compared to the physical solution across the $n = 100$ steps at $|S| = 10$. Thus to prevent leakage into the desired physical solution, 23 extra digits of working precision are required.

Because both end values in Eq. (3) are known, the textbook cure is not to evaluate the recurrence one step at a time, but instead to solve for all the M_i simultaneously. The $n-1$ recurrence

relations together with the two end values form a tridiagonal system of linear equations, and solving that system (Olver's method; Olver 1967; Gautschi 1967) is numerically stable in the usual sense, meaning that this method results in a solution that satisfies the equations it was given to within rounding error.

However, the problem is that satisfying these equations does not mean that the solution is correct. The general solution of the recurrence is a combination of the modest physical solution and a companion unphysical solution that is astronomically large. For example, writing the exact solution as $M_i = \alpha_i + \beta_i e^{-S}$, the companion part has a maximum α_i of at least 10^{83} for $n = 100$ and $S = \pm 10$ (or 10^{304} for $n = 500$ and $S = \pm 100$), while the desired physical solution in both cases has M_i of order one. If we can precisely impose the two desired end conditions, $M_0 = e^{-S}$ and $M_n = 1$, the unphysical solution exactly cancels out and only the physical solution remains. But unless we carry out the calculation to 80-odd (respectively 300-odd) decimal places, the unphysical solution will not cancel to high enough precision in the numerics, and because this solution is so much larger it will dominate over the desired physical solution.

One might imagine we can simply "be careful" and carry out the calculation at high enough precision. However, the precision required is set by the size of the unphysical companion solution, which grows without bound with n (and also depends weakly on $|S|$). And the key problem is that we cannot check if we have been precise enough: any numerical check will report that the solution obeys Eq. (3) perfectly (because the unphysical solution is still a solution), despite sharing no digits with the true desired physical M_i. Thus there is no precision at which one can simply be careful. The number of digits that must be carried grows rapidly with n, and the only check available at fixed precision (the residual) does not detect the failure.

If we are only interested in a single spectrum entry $\xi_i(S)$, at one sample size and one selection coefficient, we can avoid this problem by not using the recurrence at all and summing the power series that defines the Kummer function directly; this can be done in ordinary double precision, provided $|S|$ stays below roughly 700, beyond which the individual terms of the series (of order $e^{|S|}$) overflow the double-precision range . POLYDFE and FASTDFE evaluate their likelihoods by this per-entry route (*dadi* solves the diffusion equation numerically instead), and the limitations of these two tools documented in Section 5 concern implementation and floating-point range, while the series itself is sound. However, this single-entry evaluation is not enough for the computations that follow. For example, the likelihood fit to gnomAD (with derivatives at biobank-scale sample sizes) requires the full spectrum *vector* and its gradient in a single evaluation, and any recurrence or boundary solve for the vector demands a working precision that grows with n. That fit also needs spectrum entries at $|S|$ up to 10^5 and n above 10^6: there the per-entry series overflows double-precision range, and any floating-point value that could be produced would come with no statement of how many of its digits are correct and no independent reference to check it against, whereas the fit requires values with a known error bound . The mimicry analysis of Section 4 requires the spectrum as an exact element of $\mathbb{Q}(e^{-S})$, and the accuracy measurement of Section 5 needs exact values to subtract from.

2.3. Three routes to a precise solution

Our resolution is to avoid solving Eq. (3) in floating point at all. For rational S, every M_i can be written as:

$$M_i = \alpha_i + \beta_i\, e^{-S}, \qquad \alpha_i, \beta_i \in \mathbb{Q}\,. \tag{4}$$

The critical point is that α_i and β_i are rational, and the boundary values ($M_0 = e^{-S}$, $M_n = 1$) split the tridiagonal system into two rational solves (one for α, one for β), which we perform exactly. The companion solution's enormous interior magnitude reappears as the *heights* of the rationals α_i, β_i (their bit lengths), where it does no harm. The heights, which we compute alongside the solution, *bound the cancellation in Eq. (4) a priori*, so that assembling M_i at any target precision with a height-dependent working precision is correct by construction.

We obtain derivatives from the same matrix. Differentiating Eq. (3) k times with respect to S leaves the tridiagonal operator fixed and produces the source $k\,M_i^{(k-1)}$ on the right-hand side, with boundary values $M_0^{(k)} = (-1)^k e^{-S}$, $M_n^{(k)} = 0$. Each derivative order of the *entire* vector is therefore one additional rational tridiagonal solve. Analytic score and curvature integrals for the PRF spectrum were given by Bustamante et al. (2001), but none of the three standard tools implements an analytic derivative of its likelihood. Here the full gradient and Hessian are exact, and we use them both for optimization and for observed-information confidence intervals in Section 5.2.

In practice, this exact route is affordable up to $n \approx 2000$; beyond that the growth in the height of the α and β terms makes it computationally very expensive. To reach $n = 10^5$ we switch instead to a recursion in the stable direction carried out in ball arithmetic (i.e. interval arithmetic that propagates a midpoint together with a rigorous error radius). The radii bound the accumulated rounding error described in Section 2.2, and we double the working precision until the radius meets the target accuracy. The sign-dependent stable direction needs starting values one step from the boundary, which have the closed form ${}_1F_1(1;n;z) = (n-1)!\,z^{1-n}\left(e^z - \sum_{m\le n-2} z^m/m!\right)$. We switch between this closed form and the power series at $z \approx 2n$ to avoid the Poisson-tail cancellation of the closed form; Appendix B describes this and related numerical hazards.

Finally, for some applications we take a direct route, which evaluates one entry in $\xi_i(S)$ at a time as described above (Section 2.2). Specifically, for $S > 0$ we use Kummer's identity (${}_1F_1(a;b;z) = e^z\,{}_1F_1(b-a;b;-z)$) to rewrite $M_i = {}_1F_1(n-i;n;-S)$ as $e^{-S}\,{}_1F_1(i;n;S)$, a series in which every term is positive (for $S < 0$ the original series is already positive). A sum of positive terms loses no digits to cancellation, so we can sum this using ordinary double precision, and estimate the error using proven geometric tail bounds (NIST DLMF 2026). We use the terms calculated via this direct route as an independent reference for the other two.

Each of these three approaches (the exact route, ball-arithmetic route, and direct route) carries its own error bound (rational heights, ball radii (Johansson 2017, 2019), and geometric series tails respectively). The composite Gauss–Legendre quadratures built on these routes (the kernel cache below and the dominance reference evaluator of the next subsection) are checked by the agreement between two quadrature orders and against a rigorously bounded value. The elementary and special function values that enter (e^{-S}, and the incomplete-gamma and error functions, taken from `mpmath`; mpmath development team 2023) are compared with exact-rational values in Appendix B.

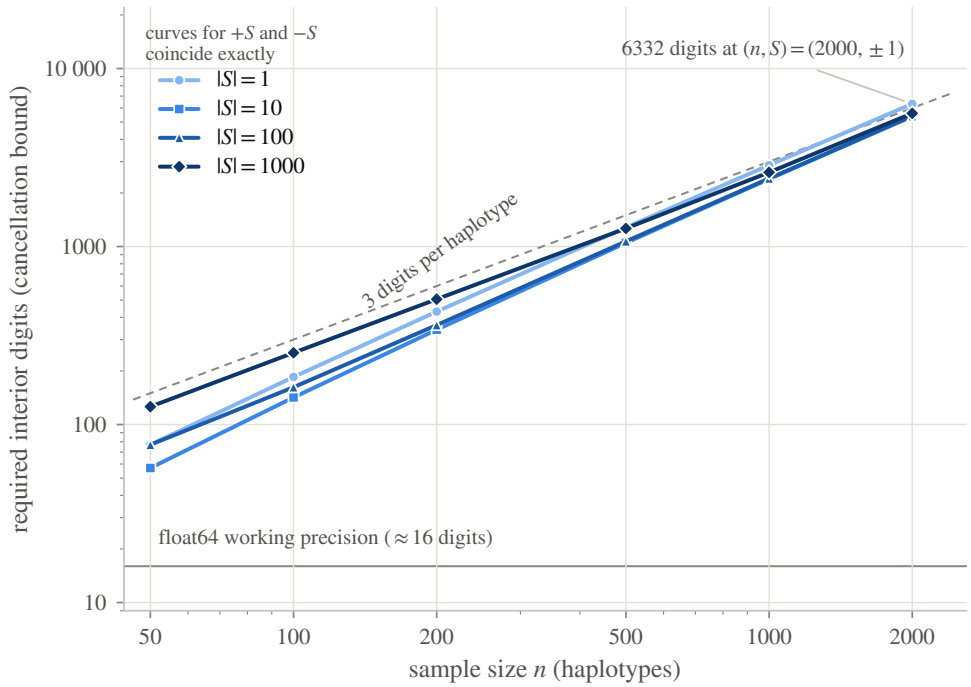

Figure 1 Working precision required to evaluate the selected SFS vector by recurrence or boundary solve. Each curve is the interior working precision, in decimal digits, required by the two-point boundary problem of Eq. (3). The requirement is symmetric under $S \to -S$, so one curve per $|S|$ is drawn. It grows at about three digits per sampled allele (dashed guide), is largest near $|S| \sim 1$ (6332 digits at $(2000, \pm 1)$), and far exceeds the ≈ 16 digits of double precision (solid rule) already at $n = 50$. Fixed-precision arithmetic therefore cannot evaluate the selected SFS vector at modern sample sizes; the exact method computes this requirement for each instance and works at that precision (Section 2.2).

The exclusion results of Section 4 depend in addition on the representation lemma (Section 2.5.1).

We tested routes E and B against a direct error-bounded quadrature of the defining integral Eq. (1) and an exact-rational reference series, at out-of-sample points in (n, S) that include the moderate-$|S|$ band where a recursion in either direction loses accuracy to the unwanted solution. Both routes agree with these references within the error bounds, and the computed gradients satisfy the parameter-shift identity $\partial_S {}_1F_1(n-i; n; -S) = -\frac{n-i}{n} {}_1F_1(n-i+1; n+1; -S)$ and the Kummer ODE (Appendix B).

For DFE likelihoods we precompute a *kernel cache*: whole-SFS vectors with gradients at fixed quadrature nodes in $\log|S|$ (composite Gauss–Legendre, with panel widths chosen to bound the range of the integrand on each panel). At small S the cache is exact: there the S-expansion of Eq. (1) has rational coefficients and the DFE moments are incomplete-gamma functions in closed form. The expected spectrum under a DFE of POLYDFE's model C form (deleterious reflected Gamma plus beneficial exponential) (Tataru et al. 2017) is then a weighted dot product against the cache, at the same cost per likelihood evaluation as the cached kernels of FIT*dadi* and FASTDFE. Unlike those kernels, ours comes with explicit tail bounds and exact parameter gradients, including the analytic s-derivative of the incomplete gamma function (tests in Appendix B).

2.4. Selection with dominance

For dominance coefficient $h \neq \frac{1}{2}$ the sojourn density no longer has the form Eq. (1). With $Q(x) = px + \tilde{q}x^2$, $p = 2hS$, $\tilde{q} = (1-2h)S$, the expected spectrum is the two-fold integral

$$\xi_k(S,h) = C(n,k)\, \frac{\displaystyle\int_0^1 x^{k-1}(1-x)^{n-k-1} e^{Q(x)} \int_x^1 e^{-Q(y)}\, dy\, dx}{\displaystyle\int_0^1 e^{-Q(y)}\, dy}, \tag{5}$$

which reduces to Eq. (1) at $h = \frac{1}{2}$ ($\tilde{q} = 0$). The sojourn density here is Wright's classical stationary formula (Wright 1938); the sampled integral Eq. (5) is the object Williamson et al. (2004) and Boyko et al. (2008) evaluate numerically for PRF inference with dominance, and spectral expansions of the underlying transition density exist (Song and Steinrücken 2012). We are not aware of a finite closed form in classical special functions or of an error-controlled evaluation at arbitrary (S, h) in the literature. Among the standard tools, *dadi* computes this spectrum with its diffusion solver (Huber et al. 2018), while FASTDFE evaluates Eq. (5) directly by an outer 1000-point power-clustered trapezoid over x and an inner 200-point linear trapezoid over y, in double precision, with no error control.

The integral Eq. (5) is an *entire deformation* of the semidominant Kummer integral, in the following sense. The building block $K(a, c; p, \tilde{q}) = \int_0^1 x^{a-1}(1-x)^{c-1} e^{px+\tilde{q}x^2} dx$ satisfies the heat relation (a diffusion equation in the parameters) $\partial_{\tilde{q}} K = \partial_p^2 K$ together with contiguity in (a, c), and the system is holonomic of rank three. To see this, write $I_m \equiv \partial_p^m K$; integrating $\partial_x \left[x^{a+m}(1-x)^c e^{px+\tilde{q}x^2} \right]$ over $[0, 1]$, where the boundary terms

vanish for $a+m, c \geq 1$, gives

$$(a+m)\,(I_m - I_{m+1}) \;-\; c\,I_{m+1} \;+\; p\,(I_{m+1} - I_{m+2}) \\ +\; 2\tilde{q}\,(I_{m+2} - I_{m+3}) \;=\; 0\,, \tag{6}$$

so for $\tilde{q} \neq 0$ the span of $\{K, \partial_p K, \partial_p^2 K\}$ closes under ∂_p, and the heat relation closes the $\tilde{q}$-direction on the same three-dimensional space (we checked the generic independence of the three generators numerically). The $\tilde{q}$-expansion of the building block,

$$K(a, c; p, \tilde{q}) \;=\; \sum_{j \geq 0} \frac{\tilde{q}^{\,j}}{j!}\, B(a+2j, c)\; {}_1F_1(a+2j;\, a+c+2j;\, p) \tag{7}$$

converges everywhere, faster than any geometric series, along the same even-a adjacent-parameter (contiguity) ladder the semidominant evaluation already uses. The inner integral of Eq. (5) expands in the same way with elementary coefficients, so the dominance spectrum becomes a convergent series of Kummer and Beta functions, each of which can be evaluated with an error bound. In words, dominance requires no new special function: the spectrum at any dominance coefficient is built from the same Kummer functions that every DFE tool already computes for the semidominant case, combined in a power series in the dominance parameter.

We evaluate Eq. (5) in two ways. The first, which we call the reference evaluator, computes the double integral directly: the inner integral has a closed form by completing the square (an erf/erfi expression valid for either sign of $\tilde{q}$), and the outer integral uses the same composite Gauss–Legendre quadrature as the semidominant kernel cache. At $h = \frac{1}{2}$ this evaluator collapses onto the exact closed form Eq. (2). The second is the series Eq. (7). The reference evaluator supplies the $h \neq \frac{1}{2}$ target spectra and error radii of Section 4 (Tables 5 and 6; assumption A7 of Appendix C10) and the reference values against which Section 5 compares FASTDFE (Figure 8); the series serves as the independent cross-check, at the ten comparison points described below and at three entries ($k = 1, 10, 19$) of each of those target spectra. Summed with elementary antiderivatives for the inner coefficients, the series is unstable at $|p| \lesssim 2|\tilde{q}|$ through a factorial-over-power blowup; a rearranged form converges uniformly at the cost of additional working precision. The two evaluators agree at the ten points across the (S, h) range where we compared them, including the corner $(500, 0.7)$ (Appendix B). At the remaining difficult corner, $(S, h) = (-1000, 0.2)$, only the reference evaluator was run; its accuracy there is judged from the agreement of its two quadrature degrees and from its exactness at $h = \frac{1}{2}$. As a check outside our parameterization, we gave *dadi*'s diffusion solver the same (S, h) points in its own dominance parameterization; it reproduces the reference spectra to within *dadi*'s own accuracy, which also rules out an h-polarization mismatch between the two conventions (Appendix B). These reference values are the standard against which we compare the tools' dominance evaluations in Section 5.

2.5. Separation and representation tests for neutral mimicry

In Section 4 we ask whether the expected spectrum of a selected site can be reproduced exactly by some neutral population-size history in the panmictic Kingman class. To show that it can, we exhibit an explicit exact convex representation of the selected spectrum over sampled neutral spectra; the spectrum then lies *inside* the sampled neutral hull and selection at that strength is mimicked. To show that it cannot, we exhibit an explicit affine separating functional, proven nonnegative on the spectrum of every neutral history in the class and negative on the selected spectrum. The spectrum then lies *outside* the neutral set and no history in the class reproduces it (an exclusion). If we find neither object at the resolution used, we report the point as *unresolved*. The expected-spectrum values that enter both constructions carry the error bounds of Section 2.3.

2.5.1. The representation lemma. The exclusions rest on the following representation of the neutral class. We measure depth in the genealogy by the cumulative coalescent intensity $g(\tau) = \int_0^\tau \mathrm{d}\tau'/2N(\tau')$ instead of in generations, and set $t = e^{-g} \in (0, 1]$, so that $t = 1$ is the present and $t \to 0$ the remote past. In this time coordinate the ancestral process of the sample is the standard Kingman coalescent whatever the size history is; the history only determines how much real time, $\mathrm{d}\tau = 2N\,\mathrm{d}g$, is spent at each depth. For $k = 2, \ldots, n$ let $P_{n,k}(t)$ be the probability that k ancestral lineages remain at depth t (an explicit polynomial in t, Eq. (C1) of Appendix C1), and define the curve

$$\gamma_i(t) \;=\; \sum_{k=2}^{n} P_{n,k}(t)\; k\, \frac{\binom{n-i-1}{k-2}}{\binom{n-1}{k-1}}, \qquad i = 1, \ldots, n-1, \\ \Gamma_n(t) \;=\; \frac{\gamma(t)}{\sum_i \gamma_i(t)}, \tag{8}$$

so that $\gamma_i(t)$ is the expected number of lineages at depth t that subtend exactly i of the n sampled genomes, and $\Gamma_n(t)$ is the normalized spectrum of a history that accrues all of its branch length at the single depth t.

Lemma 1 (Representation of the neutral class)**.** *For every size history $N(\tau) > 0$ with finite expected total branch length, the expected unnormalized spectrum is $\xi_i = \int_0^\infty \gamma_i(e^{-g})\,\mu(\mathrm{d}g)$ with the nonnegative measure $\mu(\mathrm{d}g) = 2N(\tau(g))\,\mathrm{d}g$, and the normalized spectrum $m = \xi / \sum_i \xi_i$ is an average of curve points, $m = \int \Gamma_n(e^{-g})\,\nu(\mathrm{d}g)$, against the probability measure $\nu \propto \big(\sum_i \gamma_i(e^{-g})\big)\mu(\mathrm{d}g)$. Consequently the closure $\mathcal{N}_n$ of the set of achievable normalized neutral spectra is contained in the closed convex hull of $\Gamma_n((0, 1])$, and an affine functional that is nonnegative at every point of the curve is nonnegative at the spectrum of every history in the class.*

The proof is a change of variables and is given in Appendix C1; the converse inclusion, that every point of the hull is a limit of spectra of genuine histories, is Eq. (10) of Section 4. The geometry of the set of achievable neutral spectra has been studied before. Myers, Fefferman, and Patterson exhibited distinct size histories with identical expected spectra (Myers et al. 2008), and Bhaskar and Song proved that within bounded piecewise-defined families of histories the expected spectrum of a sufficiently large sample determines the history (Bhaskar and Song 2014). Rosen, Bhaskar, Roch, and Song characterized the set of expected spectra achievable by panmictic size histories, showing that every such spectrum is also produced by a piecewise-constant history with at most $2n - 1$ epochs and tracing the degenerate fits of demographic inference to this geometry (Rosen et al. 2018). Rosen, Scholten, and Vinzant then showed that at every n the closure of this set, in coordinates that an invertible linear map carries to the spectrum, is the convex hull of a sparse moment curve, and that histories with $n - 2$ epochs already reach all of it (Rosen et al. 2022). Lemma 1 and Eq. (10) describe the same set in the intensity coordinate t, in the form the separation proofs below use: its closure is the closed convex hull of a single polynomial curve. What Appendix C4 adds is a self-contained proof in these coordinates whose inner half is an

explicit family of step histories with an ℓ_1 error bound, Eq. (C3), which is what makes a mimicry statement a statement about genuine histories and not only about the hull. Because the exponents of t in $P_{n,k}$ are the triangular numbers $\binom{j}{2}$, $2 \le j \le n$, every γ_i is a polynomial in t with a common factor t, which is why the sign of an affine functional along the whole curve reduces to a univariate polynomial problem of degree $\binom{n}{2} - 1$ (189 at $n = 20$).

2.5.2. Proving the separation inequalities. Each exclusion requires two facts: the separating functional is nonnegative along the whole neutral curve, and it is strictly negative at the selected spectrum. Both facts are proved in exact arithmetic, each in two ways (Appendix C2). Because spectra are normalized, the value of the affine functional along the curve is a polynomial in t of degree at most 189 over a positive normalizer, and we settle its sign on the whole curve by Bernstein expansion with exact dyadic bisection (B) and by exact real-root isolation (R). The sign at the selected spectrum, whose entries lie in the field $\mathbb{Q}(e^{-S})$, is determined symbolically (P) and in ball arithmetic (D) (Johansson 2019). Methods B and R prove the polynomial itself nonnegative at every $t \in [0, 1]$, and the spectrum of any neutral history is an average of curve points against a nonnegative measure (the representation lemma). The proven sign therefore passes under the average, and the inequality holds for each history separately, without recourse to the convex-hull description of the neutral set. For the same reason the inequality is preserved under nonnegative weighting and summation across loci, as the pooled per-locus results of Section 4 require; at points where we found no separating functional there is no pooled statement either. A mimicry result consists of the complementary object, an explicit set of sampled curve points with exact nonnegative rational weights summing to one that reproduces the target spectrum entry by entry.

Each computation was accompanied by a control with a known answer (Appendix B). We omit $n = 50$, where the sampling grid was provably too coarse to recover a control spectrum known to lie inside the neutral hull. We also omit three configurations: a recessive case at $S = -300$ and a strong-purifying case at $n = 35$, which did not finish within our fixed per-point computational limit, and a candidate exclusion at $n = 35$ whose control did not complete within that limit.

3. The strong-selection tail of gnomAD

We find that in gnomAD v4.1, 73.6% [57.4%, 83.6%] of the fitted DFE mass for high-confidence loss-of-function mutations lies in the deepest selection window the spectrum resolves, $|S| \in [10^4, 10^5]$ (quasi-Poisson 95% profile interval, with the measured model-error floor included). If $S = 4N_e s$ is converted to s with the textbook long-term N_e, this suggests that most mutations that disable a gene are close to lethal. With the larger effective sizes appropriate to recent growth the same window edge corresponds to $hs \approx 0.02$, and we compare both conversions with prior estimates below. The missense DFE holds 28.5% [7.7%, 42.7%] of its mass in the same window (Figure 2).

3.1. The regime and the data

A deleterious variant at scaled selection coefficient S segregates at frequencies of order $1/|S|$, so the information a sample of n chromosomes carries about a selection class peaks near the allele-count cell $i^* \approx n/|S|$. At classical sample sizes the entire strong-selection range $|S| \in [10^2, 10^5]$ is compressed into the first one or two cells of the spectrum. At gnomAD v4.1's $n = 1{,}461{,}894$ exome alleles (the release's uniform allele-number cap, AN) (Chen et al. 2024), i^* lies inside the observable spectrum for the whole window. This is also the regime in which the ill-conditioning described in Section 2.2 is most severe, since the companion-solution magnitudes of 10^{83}–10^{304} found there arise already at $n \le 500$. We have found no error-controlled likelihood fit at this sample size in the literature. In this section we fit the strong-selection tail of the deleterious DFE (window masses over $W_1 = [10^2, 10^3]$, $W_2 = [10^3, 10^4]$, $W_3 = [10^4, 10^5]$ in $|S|$, for missense and high-confidence loss-of-function classes) and the dominance relation $h(s) = 1/(1/h_0 + k_h|s|)$ (Agrawal and Whitlock 2011; Huber et al. 2018) jointly from autosomes plus X. We compute every expected-spectrum cell in the fit at the full sample size, with no projection from a smaller n, by the direct route of Section 2.3. For $i \le 5000 \ll n$ the per-entry Kummer series in the arrangement ${}_1F_1(i; n; \cdot)$ converges in a number of terms that does not grow with n; we sum it in arbitrary-precision arithmetic to a target of 30 significant digits, raising the working precision to absorb the partial cancellation that this arrangement has at $S < 0$, so that neither floating-point range nor the conditioning of Section 2.2 enters. Because alleles this rare are almost always heterozygous, dominance enters through the linear selection exponent $p = 2hS$ of Section 2.4 with a first correction in $\tilde{q}$; we tabulate the cell sums once on a 120-node grid in p, interpolate between nodes, and carry for each cell an error bound (the remainder of this reduction plus the interpolation error) through to the likelihood. We checked the direct route against route B at $n \approx 10^5$ and against an error-bounded quadrature of Eq. (1) at $n = 2 \times 10^6$, which bracket the gnomAD sample size, and the tabulated cells against the direct route at the gnomAD n itself (Appendix B). The X-chromosome spectrum needs no new evaluator: with dominance it stays in the same confluent family under the parameter remap $p_X = S_X(4h/3 + 2/3)$, $\tilde{q}_X = S_X\,(2/3)(1 - 2h)$, and we fix its normalization by comparison with an exact Wright–Fisher X chain.

The data are the gnomAD v4.1 public exome release (Chen et al. 2024). The release comprises 730,947 individuals, all 22 autosomes at the uniform allele-number cap AN given above, plus X non-pseudoautosomal ($n = 1{,}098{,}270$). It pools all ancestry groups, and we handle the pooled spectrum's structure- and growth-driven distortions by the synonymous calibration below, which absorbs whatever acts equally on synonymous and selected sites within a chromosome set. We use PASS single-nucleotide variants, canonical-transcript consequence classes, an allele-number completeness cut at 0.8 of the cap, and hg38 triplet mutation contexts. We assign each variant its most severe consequence on the canonical transcript, loss-of-function > missense > synonymous, with the high-confidence LOFTEE (Karczewski et al. 2020) subset as the primary LoF class. We treat alternate alleles as derived; the fits use the rare end (AC ≤ 5000, allele frequency $\le 3.4 \times 10^{-3}$), where reference/ancestral misassignment affects a negligible fraction of sites and, like other class-shared distortions, is absorbed by the synonymous calibration. After filtering, the autosomal segregating counts are 5.29M synonymous, 11.78M missense, and 653K loss-of-function (603K high-confidence). We exclude CpG transitions from the primary fits because at this n they are mutation-saturated and the infinite-sites assumption of the PRF fails for them. On chromosome 21, for example, the synonymous singleton fraction is 0.042 for CpG transitions, against 0.45/0.53 for other transitions/transversions (approaching saturation was already reported at gnomAD v2's

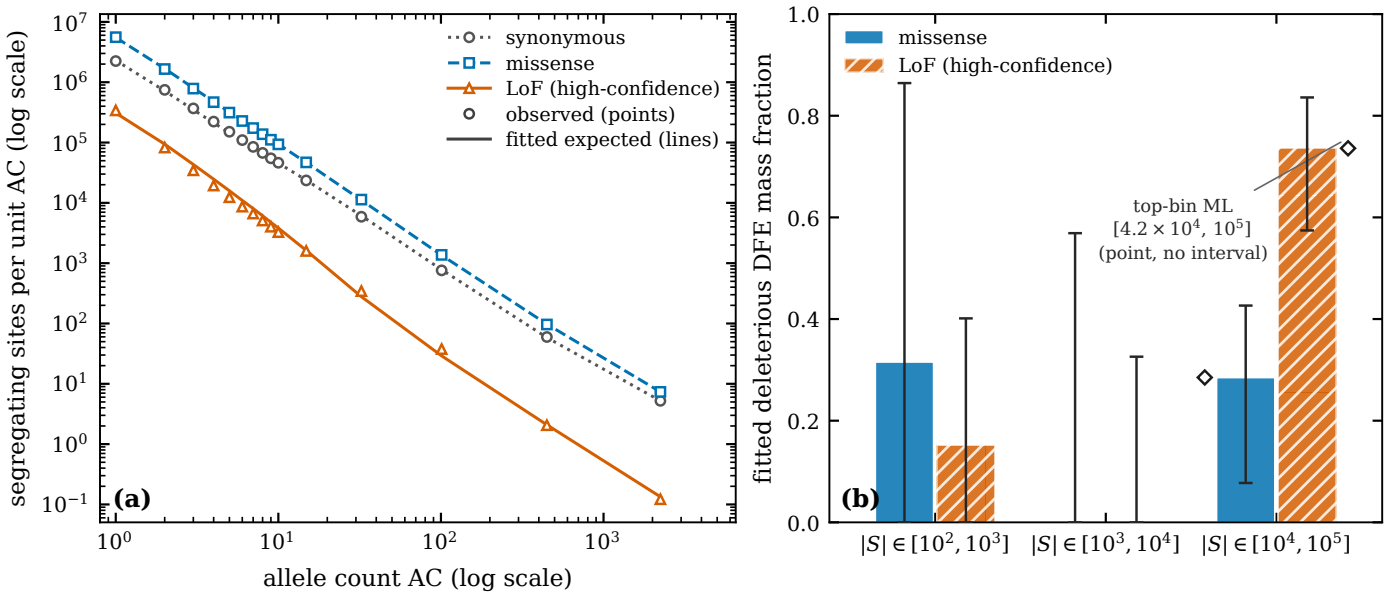

Figure 2 The strong-selection tail of the deleterious DFE in gnomAD v4.1. (a) Observed rare-end site-frequency spectra by consequence class, with the fitted expected spectra overlaid; the fits use the rare end (AC $\leq$ 5000) only. (b) Fitted deleterious DFE mass by $|S|$ window with quasi-Poisson 95% profile intervals (values in Table 1). The deepest window $|S| \in [10^4, 10^5]$ holds about three quarters of the high-confidence LoF mass and somewhat more than a quarter of the missense mass, and the maximum-likelihood configuration concentrates that mass in the top bin $[4.2 \times 10^4, 10^5]$, with an 11% LoF bulk below $|S| = 10^2$ consistent with annotation escapees. Intervals include the measured model-error floor ($\hat{\phi} = 1788$ missense, 624 LoF-HC; scan height $1.92\,\hat{\phi}$). Converted to s, the window corresponds to near-lethal heterozygous effects at the textbook long-term N_e and to substantially weaker effects at the larger effective sizes of recent growth (Section 3.3).

125,748 exomes (Karczewski et al. 2020)), and including CpG transitions degrades the fit substantially (Appendix B). The strong-selection signal lies mostly outside the singleton cells, where ascertainment is most in doubt: the $i = 1$ column carries $\leq 8\%$ of the information about the $|S| \geq 10^3$ classes, whose information peaks at allele counts near 1464 for $|S| = 10^3$ and 147 for $|S| = 10^4$. Demography and context structure enter as per-cell distortions calibrated on synonymous sites and shared within each chromosome set, with a free mutation intensity per context stratum. This is the r_i-style nuisance-multiplier pattern introduced for DFE inference by Eyre-Walker et al. (2006) and used by the standard tools, which absorbs equilibrium-model misspecification at the rare end, where the strong-selection signal is concentrated. For the strongly deleterious classes fitted here, that rare end sits near mutation–selection balance, the regime in which equilibrium-based inference of heterozygous selection has been examined and defended at large exome-cohort sample sizes (Weghorn et al. 2019). The error bounds apply to the likelihood conditional on this treatment of demography (Section 7); we fit no demographic model and use no restricted annotation fields.

3.2. Validation on simulated data

Before fitting the real spectra we simulated spectra from a known DFE at the same n and with the same strata as the gnomAD data, and fitted them with the same optimizer, to determine which functionals of the DFE are identifiable. We recovered the model parameters themselves within their uncertainties (Appendix B), but two functionals were not identifiable. (i) A $+0.03$ spike placed in one interior bin of the eight-bin tail model (described in the next subsection) was undetectable ($p = 0.46$, likelihood-ratio test): per-bin tail masses lie on a likelihood ridge along which neighboring bins absorb any one bin. (ii) Unconditional profile intervals for (h_0, k_h) do not close even with the X chromosome included. In the simulations (h_0, k_h) trade off against the X-to-autosome population-size ratio N_X/N_A along a second likelihood ridge. The X data bound the position of this ridge, so that the interval for N_X/N_A is finite, without bounding h by itself. They tighten the one-sided h_0 and k_h profile bounds by factors of 2.45 and 1.70 relative to autosomes alone. For the real data we therefore report only window functionals, whole-tail likelihood ratios, the N_X/N_A interval, and one-sided bounds on h_0 at fixed N_X/N_A, and no per-bin masses or unconditional dominance intervals. We optimize on the deviance scale and use profile-based intervals throughout (Appendix B).

3.3. The fitted tail and its robustness

The deleterious model is a Gamma bulk plus eight free mass bins log-spaced over $|S| \in [10^2, 10^5]$. The window boundaries 10^3 and 10^4 do not coincide with bin edges, so the window masses below apportion a straddling bin's mass log-uniformly within the bin, and we compute the profile intervals on the same fixed weights. We measure the model-error floor by the quasi-Poisson dispersion $\hat{\phi}$, the full deviance per degree of freedom (values in the caption of Figure 2). In both classes it corresponds to percent-scale systematic residuals from structure (demography, context, biology) that the DFE-plus-distortions model does not carry. Because the expected spectra carry error bounds far below this level, we can attribute the misfit to the model. All intervals quoted below are quasi-likelihood profile intervals at 95%, computed at scan height $1.92\,\hat{\phi}$ on the half-deviance scale, so that the measured model-error floor is built into the interval widths;

their nominal coverage holds only insofar as a single dispersion factor $\hat{\phi}$ describes the misfit.

The simulation study of Section 3.2, together with intervals that include the model-error floor $\hat{\phi}$, supports four statements about the fitted tail (Table 1).

1) Strong-tail mass. The maximum-likelihood configuration concentrates the $W_3 = [10^4, 10^5]$ mass (Table 1) in the top bin $[4.2 \times 10^4, 10^5]$ (a lethal shelf) and empties the middle of the tail (missense: 32% near $[10^2, 2.4 \times 10^2]$, nothing in between; LoF additionally keeps an 11% bulk below $|S| = 10^2$, consistent with annotation escapees). The W_3 mass is the most robust of the fitted quantities: it is stable to fitting autosomes only versus autosomes plus X (the autosome-only shift is quoted below) and to LOFTEE stringency.

2) Comparison with a pure Gamma DFE. Under the raw Poisson likelihood, freeing the tail improves decisively on a pure Gamma DFE in both classes (likelihood-ratio statistics in Table 1). Once the model-error floor is included, the Gamma is disfavored for LoF but not for missense (quasi-likelihood ratio tests, Table 1), so an extreme Gamma can mimic the missense shelf within model error. Statement 1) does not depend on this shape test.

3) Shape of the tail. A bimodal tail is preferred by the likelihood but is not established: per-bin masses are unidentifiable (Section 3.2), and the quasi-95% intervals allow up to 57% of the missense mass in the middle window. The fit therefore determines the window functional W_3 and leaves the shape of the tail within it open.

4) The $|s|$ scale. The fit measures $S = 4N_e s$, so quoting s requires a choice of N_e. At the textbook long-term $N_e = 1.5 \times 10^4$ the shelf corresponds to $|s| \sim 0.5$–1 (as strong as the diffusion parameterization can express), while recent-growth epochs (larger N_e) map the same S to smaller s. Because the rare end measures $2hS$ (Section 6), the shelf's S-location is also conditional on the fitted $h \approx 1$: at $h = 0.5$, allowed for missense at quasi-95%, the S axis slides by a factor of two.

The fit behind these four statements uses one DFE for the whole exome, with sites treated as independent and the genome as a single environment, and linkage is the natural objection to these idealizations. We therefore made two checks against linked selection, both on the W_3 functional. We first asked whether the tail moves with recombination environment. We split the analysis genes into recombination terciles on the deCODE 2019 sex-averaged genetic map, which is a crossover map and therefore does not stratify gene conversion, a short-range component of recombination whose effect on local linkage is comparable to crossing-over (Frisse et al. 2001; Gay et al. 2007). We refit each stratum with the identical exact likelihood and same-stratum synonymous calibrators. Under the raw Poisson likelihood the tail moves: the fitted W_3 is largest in the lowest-recombination tercile ($W_3 = 0.7731, 0.7110, 0.7219$ in the lowest, middle, and highest terciles). The spread across terciles (largest minus smallest fitted W_3, 0.0622 at full precision) exceeds every draw of a shuffled-gene permutation null ($p_{\text{perm}} = 0.0204$, 48 draws, permutation test), although the dependence on recombination rate is not monotone. Under the quasi-likelihood intervals, which carry the dispersion measured in each stratum, the same split cannot be resolved: the per-stratum intervals ([0.49, 0.92] against [0.42, 0.91]) overlap almost completely. At this dispersion, however, quasi-likelihood intervals computed from a third of the data are too wide to resolve a 0.06 split, so the overlap should not be read as evidence that the tail does not move. The stratification supports the robustness of statement 1): on both the recombination axis and a secondary gene-density axis (described below), every stratum optimum lies within ± 0.04 of the pooled optimum and inside the pooled quasi interval. Because the per-stratum quasi lower bounds (0.42–0.49) individually dip below $\frac{1}{2}$, the statement that a *majority* of the mass lies in the strongest window rests on the pooled fit, with the strata as consistency checks at the maximum-likelihood point. The direction of the split (more apparent strong-tail mass at low recombination) is consistent both with linkage inflation of the apparent DFE tail and with gene-composition confounding, since low-recombination and gene-sparse regions do not carry a random draw of genes. The shuffled-gene null breaks both associations at once, so exceeding it shows only that the tail is associated with recombination environment, and we draw no causal conclusion in either direction. The strata are autosomal by design, whereas the pooled primary fit is autosomes plus X (the autosome-only shift of the pooled W_3 is -0.0009, inside the interval). All stratified quantities are window functionals, for the identifiability reasons of Section 3.2, with each one-sided bound a 97.5% bound. Across terciles of the secondary axis, gene density, the fitted W_3 varies monotonically (0.767, 0.724, 0.715), but we ran no permutation null for that axis, so the significance of that trend is untested.

We then asked whether linkage could have produced the tail outright, that is, whether a fully linked genome carrying no strong-selection mass at all could be fit as $|S| \geq 10^4$. We ran full forward simulations of nine deliberately severe linked-selection regimes (extreme background selection, deep Hill–Robertson interference, saturated sweeps, and draft through the Bolthausen–Sznitman class) and fit neutral windows of $n = 20$ sampled chromosomes from each with the identical exact likelihood. These simulations have two limitations: the regime parameters follow our reading of the source studies and were not re-derived from them, and the simulated selection strengths were half the nominal literature values (a property of the simulator's fitness encoding); simulation-side strengths below are therefore reported at the values actually simulated, which leaves the fitted values and output spectra unchanged.

Linkage did not produce an apparent strong tail in any regime. Harsh linked selection is fit as *weak purifying* selection, more than three orders of magnitude below the window, although the fit grid ran to $\pm 10^5$ and the fits were free to reach the window edge. The resolved regimes are fit at $|S|$ of order 2–6. Across all nine regimes the point estimates lie in $[-6.09, -0.36]$, on the purifying side of zero and opposite in sign to the excluded range of Section 4, and no 2-unit likelihood interval endpoint reaches that range (Appendix B). The units in these fits are nominal: each regime is reported at its simulated parameters, and the population-scaled mapping to human units is itself the diffusion-limit approximation these regimes are chosen to break. At $n = 20$, moreover, the conditional spectrum saturates beyond $|S| \sim 20$ (a near-lethal site and an $S = -100$ site are indistinguishable there), so the value 10^4 specifically cannot be recovered by any fitter at this n. The proven exclusion range of Section 4, $S \approx 1.375$–100, begins well inside the resolved regime. Two controls indicate that this null result is a property of the simulated spectra and is unlikely to have arisen from an insensitive fitter. A simulated neutral window is fit near zero with all intervals inside $(-10^4, +10^4)$. The identical fitter, run on exactly computed spectra of known S, recovers S through the resolved regime and returns, for every strongly selected input ($S \leq -10$), an estimate more negative than the entire range the linked-selection regimes produced (Appendix B).

quantity	missense	LoF (high-confidence)
W_1 mass, $\lvert S\rvert \in [10^2, 10^3]$	0.32 [0, 0.86] (not robust)	0.15 [0, 0.40]
W_2 mass, $\lvert S\rvert \in [10^3, 10^4]$	0.00 [0, 0.57]	0.00 [0, 0.33]
W_3 mass, $\lvert S\rvert \in [10^4, 10^5]$	**0.285** [0.077, 0.427]	**0.736** [0.574, 0.836]
whole-tail LRT vs. Gamma	$D = 4944$; $D/\hat{\phi} = 2.8$; $p_{\text{quasi}} = 0.95$	$D = 8991$; $D/\hat{\phi} = 14.4$; $p_{\text{quasi}} = 0.072$
N_X/N_A	1.07 [0.68, 1.22]	1.07 [0.68, 1.22]
h_0 at fixed $N_X/N_A = 0.75$	bound crosses in (0.15, 0.5] (coarse)	bound crosses in (0.5, 0.934]

Table 1 Identifiable quantities from the gnomAD v4.1 fits (quasi-Poisson 95% profile intervals; model-error floor included via $\hat{\phi}$). Window masses are fractions of the deleterious DFE. The h_0 rows are *one-sided bounds* at fixed N_X/N_A, bracketed by a coarse profile scan and conditional on that population-size ratio; they are not confidence intervals (Section 6). W_1 is the most demography-entangled window and is labeled accordingly.

Relation to prior estimates. We are not aware of a prior measurement of the W_3 functional; the nearest published quantities differ from it by construction, and the comparison depends on the choice of N_e discussed under 4). On the per-variant axis, Agarwal, Fuller, Myers, and Przeworski weight per-gene posteriors by mutational opportunity and report that "more than 56% of all possible autosomal LOF mutations have an estimated $hs > 1\%$, while 20% have an hs of 10% or greater" (Agarwal et al. 2023). They note that their sampler "rarely sample[s] hs values very close to 1, potentially underestimating the number of dominant lethal genes." The published s_{het} distributions are per-gene quantities, whereas W_3 is a fraction of mutation events. Per gene, the mean is $s_{\text{het}} \approx 0.059$, with 2,984 of 15,998 scored genes above 0.1 (Cassa et al. 2017); with a correction for drift, two-thirds of genes lie above 0.01 (Weghorn et al. 2019); and with a feature-informed prior at gnomAD v2 scale, $s_{\text{het}} > 0.1$ marks roughly the top 15% of genes (Zeng et al. 2024). The pre-biobank site-frequency DFEs (Boyko et al. 2008; Kim et al. 2017) are missense-only and take their strong fraction from a fitted Gamma family. At the textbook long-term $N_e = 1.5 \times 10^4$ the window edge $S = 10^4$ corresponds to $hs \approx 0.17$ at the fitted $h \approx 1$. The LoF W_3 mass above that edge would then be a substantially heavier tail than the 20% at $hs \geq 0.1$ that Agarwal and colleagues report, in the direction that their remark about dominant lethals anticipates. Under recent-growth effective sizes an order of magnitude larger, the same edge corresponds to $hs \approx 0.02$, and the LoF W_3 mass then falls inside their per-variant range. The population-scaled window is the measured quantity; which value of s it corresponds to depends on the effective size relevant to these variants, which we do not attempt to determine here. The rare end by itself also leaves dominance unresolved, and we return to it with the X chromosome in Section 6.

4. Mimicry of selection by panmictic demographic history

Demographic history moves the site-frequency spectrum in the same directions selection does: expansions inflate rare variants and bottlenecks deplete them, so anyone inferring selection from frequency data must ask whether a history of population sizes could have produced the signal on its own. The usual practice is to fit a small number of demographic models and show that the signal survives. With the exact methods of Section 2.3 and Section 2.5 we address the question that those fits approximate, whether *any* size history in the panmictic Kingman class reproduces the expected spectrum of a selected site. Because the comparison is carried out in exact arithmetic, the answer can take the form of a proof. The computation at each sampled point is given in Appendix C and the assumptions in Appendix C10. All results are at sample size $n = 20$ unless stated and are relative to the panmictic Kingman class only, so mimicry by structured populations or by multiple-merger genealogies is not excluded. The geometry of the set of achievable neutral spectra on which the proofs rest was characterized by Rosen et al. (2018) (Section 2.5.1). The nearest prior work on the mimicry question itself is statistical: the robustness of sweep tests to demography at single loci (Jensen et al. 2005), and the failure of parametrized families of histories to fit particular datasets (Árnason et al. 2023).

Mimicry and exclusion. Whether a neutral history can reproduce a selected spectrum depends on the strength of selection. At each weak strength we sampled, $S \in \{-10, -1, -0.1, +\frac{1}{2}, +1\}$ with $h = \frac{1}{2}$, the selected spectrum admits an explicit exact convex representation over sampled neutral curve points. We also constructed explicit eighteen-epoch neutral histories whose spectra reproduce the weakly selected ones; the ℓ_1 distances, all orders of magnitude below the resolution of the analysis, are listed in Appendix C6. In this sampled-hull sense, demography can reproduce weak positive selection and weak-to-moderate purifying selection. At moderate and strong positive selection, additive or with dominance, and at the two recessive deleterious points $(S, h) = (-30, 0.1)$ and $(-100, 0.1)$, we find the opposite: twenty-five distinct (S, h) points in all are excluded (Appendix C3 and Figure 3). Each of the twenty-five is excluded by an explicit affine separating functional, proven nonnegative on the whole neutral curve and negative at the selected spectrum (Section 2.5.2), so that by Lemma 1 no history in the class reproduces that spectrum. One such functional, which we call the shared separating functional, with nineteen rational coefficients and one dyadic constant printed in full in Appendix A, proves six of these exclusions ($S \in \{2, 5, 10, 20, 50, 100\}$) at once and can be checked with rational arithmetic alone. Between the two regimes the onset of distinguishability, $S^* := \inf\{S > 0 : \text{excluded}\}$, is bracketed on the grid points we analyzed by

$$S^* \in (1.125,\ 1.375], \tag{9}$$

meaning that $S = 1.125$ is the largest analyzed strength that is mimicked (Appendix C6) and $S = 1.375$ the smallest that is excluded. The two interior grid points ($S = 1.25$ and 1.3125) are unresolved (Appendix C5). Strong purifying additive selection ($S = -100$ and -1000 at $h = \frac{1}{2}$) and the point $(S, h) = (-100, 0.3)$ also remain unresolved. These spectra lie exponentially close to the sampled hull, closer than the separating functionals we found can resolve, and whether they are

mimicked or excluded by an exponentially small margin is open (Appendix C7 and Table 8).

The closure theorem and quantitative standoffs. The representation lemma (Section 2.5.1) gives an inclusion, namely that every achievable neutral spectrum lies in the closed convex hull of the curve Γ_n. We prove that the inclusion is an equality at every sample size,

$$\mathcal{N}_n \;=\; \mathrm{conv}\big(\Gamma_n([0,1])\big) \qquad \text{for every } n \geq 3, \tag{10}$$

where $\mathcal{N}_n$ is the closure of the set of achievable neutral spectra. The reverse inclusion is proved constructively: for any point of the hull we build an explicit sequence of genuine step histories whose spectra converge to it at a computable rate (Appendix C4, Eq. (C3)). The closure is needed because the $t = 0$ endpoint of the curve is a limit that no genuine history attains. A first consequence of the equality is that pooled spectra over arbitrary finite families of per-locus histories and weights are convex combinations of single-history spectra, so allowing each locus its own history does not enlarge the set of achievable expected pooled spectra. A second is that, if a functional is proven nonnegative on the whole curve, with violation margin $\mu > 0$ at the selected spectrum, then one further finite computation on its coefficients shows that every history in the class keeps ℓ_1 distance at least $\mu/\|a\|_\infty$ from that spectrum. We have carried out that computation at thirteen points: the twelve additive strengths from $S = 1.375$ to $S = 100$ at $h = \frac{1}{2}$, with proven lower bounds on the ℓ_1 distance (standoffs) rising from 8.9×10^{-13} at $S = 1.375$ to 1.3×10^{-2} at $S = 100$, and the dominance point $(S, h) = (10, 0.1)$ (Appendix C4 and Figure 6). Because a bound that holds at every point of the curve is preserved by any nonnegative weighted average over the curve, the separating functionals also exclude every assignment of per-locus neutral histories combined with data-independent weights. This covers background selection in its standard first-order (drift-rescaling) description at any strength and genomic granularity, and we refer to the statement as the transfer theorem. The theorem applies wherever a curve-wide separating functional has been proven. The quantitative standoffs have so far been computed only at the thirteen points above, none of them at $S < 0$ or among the remaining dominance exclusions; no exclusion has been established at any $n \neq 20$, and pooling conditioned on observed counts lies outside the theorem (Appendix C3).

The boundary of the class. Linked selection could still produce mimicry where its exact description leaves the Kingman class. The exact structured coalescent for background selection, hitchhiking during the sample's genealogy, and the multiple-merger genealogies (Λ-coalescents; Pitman 1999; Sagitov 1999) produced by strong draft all lie outside that class. For multiple mergers the boundary can be located exactly: on the grids we swept at constant population size, every $\mathrm{Beta}(2-\alpha, \alpha)$ coalescent with $\alpha \leq 19/10$ and every $\mathrm{Dirac}(\psi)$ coalescent (Eldon and Wakeley 2006) with $\psi \geq 1/100$ has its neutral spectrum on the negative side of the shared separating functional, and the Beta family crosses to its nonnegative side at $\alpha^* \in (1.950757, 1.950801)$. The Kingman-class exclusions therefore cannot be widened to Λ-coalescents with the existing functionals. Lying on the negative side of the existing functionals is necessary but not sufficient for mimicry, however. A numerical search over the Beta and Dirac–Kingman families and a full forward simulation of a draft regime both suggest that multiple mergers move the spectrum away from the excluded selected spectra, since they inflate the singleton class whereas additive selection at these strengths depletes it (Appendix C7). In one simulated regime, exact background selection at zero recombination, the pooled neutral-window spectrum falls on the excluded side of the shared functional's hyperplane. That regime lies outside the rescaling class and at finite θ, which is the case the transfer theorem leaves open, and we discuss it in Appendix C7.

Two identifiability theorems. With the methods of this section we also answer two identifiability questions. Dominance and selection are often said to be confounded in frequency data (Balick et al. 2015; Sendrowski et al. 2026). At each of twelve points we analyzed at $n = 20$, we prove that the joint Jacobian of the composite demography–selection–dominance map has full rank, so no exact confounding ridge passes through those points. The practical confounding appears instead as poor conditioning, with enclosures of the smallest singular value as low as $[1.13 \times 10^{-20}, 4.047 \times 10^{-9}]$ (Appendix C8, which states the restrictions on this local result). The fraction of strictly neutral sites, by contrast, is exactly unidentifiable from the spectrum. For $n = 6$ through 10, a model placing mass $\pi_0 \in \{\frac{1}{10}, \frac{3}{10}\}$ on exactly neutral sites, with the rest Gamma-distributed, has the same normalized expected spectrum as a mixture of at most $n-1$ Gamma distributions with no atom. No maximum-likelihood fit, likelihood-ratio test, or other scale-invariant procedure can therefore distinguish the two, however many sites are observed (Appendix C9). Whether the rigid single-Gamma family keeps a positive distance from the atom model remains open; the best lower bound we have been able to prove on that distance is zero. The poor conditioning of dominance against selection has an empirical counterpart in the gnomAD fits of Section 3, where dominance is constrained only through the X chromosome and the quantities estimated are those that the simulations of Section 3.2 showed to be identifiable.

5. Comparison of the standard software with exact values

We compared three widely used implementations of PRF-based DFE inference with the exact values of Section 2.3: polyDFE v2.0 (Tataru and Bataillon 2019), built with the GNU Scientific Library (GSL 1.16; Galassi et al. 2009), fit*dadi* (Kim et al. 2017) with the *dadi* diffusion solver beneath it (release 2.4.4), and fastDFE 1.3.3 (Sendrowski and Bataillon 2024). We first compared each numerical layer in isolation (Table 2) and then refit identical data with each tool's numerics and with the exact likelihood, expressing the difference in fitted parameters in units of the exact-likelihood confidence interval (Section 5.2). We took the parameter conventions of each tool (Section 2.1) from its source code, since convention mismatches are the largest source of spurious discrepancies in comparisons of this kind. All statements below refer to the versions named here, and per-configuration values are collected in Appendix D.

The main finding is that, at the sample sizes and deleterious-DFE settings for which these tools were developed and at which most published estimates were obtained, their default pipelines agree with the exact likelihood to a small fraction of a confidence interval. The differences we do measure are confined to three settings. These are polyDFE model C at sample sizes $n \gtrsim 90$, when the DFE has beneficial mass above $S \approx 38$; the treatment of DFE mass beyond the cached γ grid in fit*dadi*, where the 2017 and current releases differ from each other; and sample sizes of

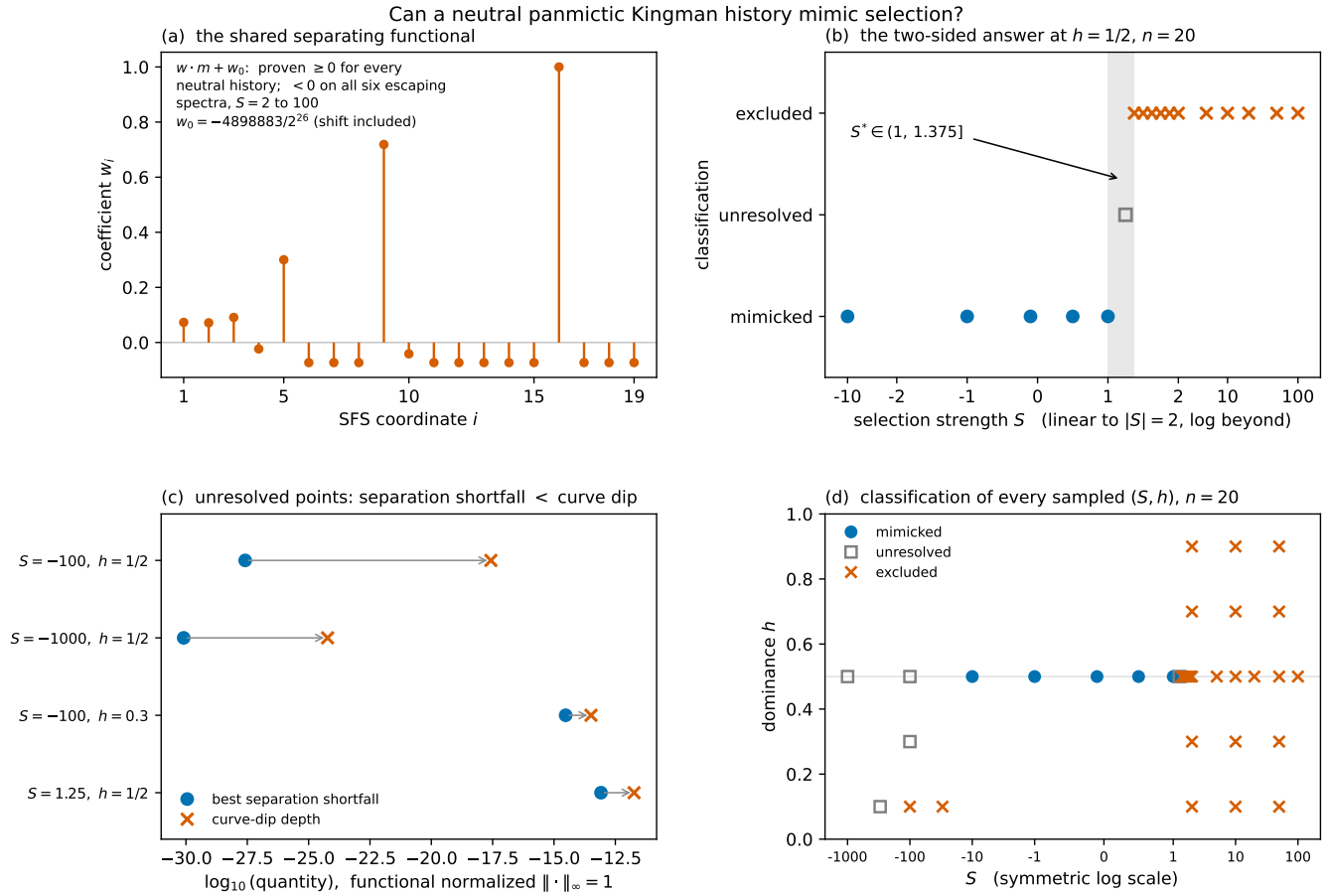

Figure 3 Mimicry and exclusion across the (S, h) plane at $n = 20$. (a) The shared separating functional: nineteen rational coefficients and one dyadic constant w_0 (the slack shift is already included in w_0); the affine functional is proven nonnegative on the entire neutral curve and negative on all six excluded spectra $S = 2$–100. (b) Results at $h = \frac{1}{2}$: mimicry (inside the sampled neutral hull) up to $S = 1$, exclusion (outside) from $S = 1.375$ on at each sampled strength, and the sampled bracket $S^* \in (1, 1.375]$ with the intermediate point unresolved. Appendix C6 adds grid points at $S = 1.125$, 1.3125, and 1.4375, tightening the sampled bracket to $(1.125, 1.375]$; the panels show the original grid. (c) The four unresolved points shown: in each row the best proven separation shortfall is smaller than the proven curve-dip depth (Appendix C5), so none of the functionals found can be shifted into one that separates the selected spectrum from the whole neutral curve; these spectra lie exponentially close to the neutral set. (d) All sampled (S, h) points at $n = 20$. Results are relative to the panmictic Kingman class $\mathcal{N}_{20}$ only; mimicry by structured populations or by Λ-coalescent genealogies is not excluded. "Inside" denotes membership in the sampled neutral hull and "outside" exclusion for every history in the class (Lemma 1); the assumptions, including the error bounds of the dominance evaluator on which the $h \neq \frac{1}{2}$ points depend, are listed in Appendix C10. Demography mimics weak selection and provably does not mimic moderate or strong positive additive selection; the single functional in panel (a) proves six of the exclusions.

order 10^5, which none of the three implementations currently handles. We also find that the error in an individual spectrum entry is a poor predictor of the error in a fitted parameter, in either direction, so we base all statements about parameters on the refits.

5.1. Spectrum entries and DFE mixtures

polyDFE. polyDFE evaluates Eq. (2) through `gsl_sf_hyperg_1F1_int` in double precision, with a Taylor switch at $|S| \leq 10^{-5}$ and overflow-avoiding branches. It integrates the DFE by adaptive Gauss–Kronrod quadrature with a relative tolerance of 10^{-5} that is relaxed stepwise to 10^{-1} when convergence fails, after which an error is returned; derivatives are finite differences ("first and second derivatives calculated numerically"). For deleterious DFEs at the sample sizes of its published applications the per-entry error is 10^{-7}–10^{-6}, and we found no quadrature failure of consequence at cited deleterious parameters. Two features of the GSL routine limit the accessible range. First, in the window $|S| \in [1.0, 5.0] \times 10^{-5}$ the routine loses all significant figures to cancellation. The DFE quadrature reaches this window at the entry $i = n/2$, which exists only for even n, and at every even n we tested from 90 to 100 the distributed binary then exits with an error. We reproduced the report of a user in 2019 (polyDFE GitHub issue #3, https://github.com/paula-tataru/polyDFE/issues/3) with the executable distributed by the authors, and the mechanism applies at any larger even n. At positive S in the same window the routine returns mixing integrands that differ from the

exact value by factors of 10^2–10^3 while the quadrature reports convergence. POLYDFE's README requires GSL 1.16, so this behavior is tied to that library version. Second, for beneficial selection at $S \gtrsim 38$, inside the $S_b \in [0, 100]$ range of the distributed configuration, the expected entries differ from the exact values by amounts that grow with sample size, reaching a relative error of 22.8 at $n = 1000$. Whether this affects a fit depends on how much beneficial mass the DFE places above $S \approx 38$ (Section 5.2).

FIT*dadi* and *dadi*. FIT*dadi* precomputes per-γ spectra with *dadi*'s implicit finite-difference solution of the forward Kolmogorov equation on a 500-node logarithmic γ grid and integrates the DFE by the trapezoid rule over the grid. DFE mass beyond the grid is dropped in the 2017 release and assigned to the most-deleterious cached spectrum in the current *dadi* release (2.4.4). On the example distributed with FIT*dadi* ($n = 250$, with data and model held fixed) the two rules give DFE scale estimates that differ by 8.6% and $E[s]$ by 7.2%, with the true value between them, and a log-likelihood difference of the order of a one-degree-of-freedom likelihood-ratio criterion. The grid trapezoid contributes a common-mode relative error of about 1.7×10^{-4} for human-like DFEs. The diffusion solver itself, compared at demographic equilibrium where the exact answer is known, has a median relative error of 2–3% and up to 57% at strong selection under its default grids at $n = 100$, and the alternative grids suggested in the manual increase this (Table 2). Runs with demography solve the same discretized operator; we did not test the demographic layer itself.

FASTDFE. FASTDFE tabulates Eq. (2) at 2002 bin edges with `mpmath` at its default of 15 significant figures, casts to double precision, and evaluates mixtures as matrix products against a trapezoid kernel; $h \neq \frac{1}{2}$ spectra come from nested trapezoids (Section 2.4) on a 21-point linear h grid. At its defaults the largest kernel-layer error we measured is 2.4×10^{-5}. With 100 bins per sign, a setting its documentation describes as "nearly identical" to the default, the kernel error is 2.4×10^{-3}; the refits below indicate that this difference is immaterial at pre-biobank data sizes.

5.2. Fitted parameters

An error common to all spectrum entries is absorbed by the mutation-rate nuisance parameters, and only the component that varies across frequency classes, projected on the score, moves fitted DFE parameters. We therefore refit synthetic data generated from exactly computed expectations, and the examples distributed with the tools, with each tool's numerics and with the exact likelihood. We report the difference in fitted parameters in units of the exact-likelihood 95% confidence half-width as a function of data size L, the expected total polymorphic count (10^4–10^5 is exome scale, 10^6 genome scale, 10^8 biobank scale), and plot it in Figure 4. These comparisons concern the *computation* of the standard PRF likelihood. Its modeling assumptions (independent Poisson counts across sites, no linkage, polarization handled through the tools' own nuisance parameters) are inherited unchanged, and at biobank scale they may well dominate the total error of any DFE fit, ours included. Where the infinite-sites assumption fails, for example, the PRF underestimates negative selection and can infer spurious positive selection (Desai and Plotkin 2008). Each refit configuration includes a control in a regime where the tool is accurate (for POLYDFE, $n = 19$ with moderate deleterious selection), and there the two fits coincide at all data sizes (Appendix B).

POLYDFE: positive-selection parameters at large sample sizes. For model C at $n = 101$ with truth $(b, S_d, p_b, S_b) = (0.4, -400, 0.05, 40)$, the parameters fitted with the distributed numerics differ from the exact refit of identical data by 2.7 (S_d) to 11.8 (S_b) confidence half-widths at genome scale, and α_{dfe} by 15.6; the exact refit recovers the truth. The difference equals one standard error already at $L \approx 2$–4×10^3 counts, below a single exome (Appendix D). The size of the effect is governed by the beneficial mass above the threshold, $p_b\, e^{-38/S_b}$, which we mapped by sweeping S_b at fixed (b, S_d, p_b) (Figure 7). At $S_b = 2.84$ and $p_b = 0.031$, the values fitted on the example distributed with POLYDFE, all differences are below 10^{-3} of a confidence interval at genome scale, whereas at $S_b = 10$ the largest is 2.3 intervals. Estimates with modest beneficial strength ($S_b \lesssim 5$) are therefore unaffected by this mechanism. Estimates of p_b, S_b, or α from model C at $n \gtrsim 90$ with $S_b \gtrsim 10$ fall in the affected range, where the quadrature reports convergence and the error is not visible in the tool's output.

FIT*dadi*: treatment of mass beyond the grid. On the DFE of the distributed example (Gamma shape 0.186, scale 687 in γ units, $n = 250$), the 2017 rule shifts the fitted scale by +74% and $E|s|$ by +72% asymptotically when the mutation rate is profiled, which is the estimand used throughout this section. With the mutation rate held at truth, as in the distributed example, the shifts are −7% (2017 rule) and +5% (current rule). In either convention the two rules are about one standard error apart at the example's data size and 3.4–4.0 confidence intervals apart at genome scale. The current rule is milder, −11% in the profiled scale, inside the confidence interval at the example's size and exceeding it from $L \approx 5 \times 10^5$; on steeper DFEs both rules shift the scale by larger amounts (Appendix D). By contrast, the diffusion-solver error of up to 57% in individual entries moves no fitted parameter by more than 0.5% once grid and tail rule are held fixed, because it is common-mode and the score projection removes it.

FASTDFE: kernel resolution. At the default 2×1000-bin kernel the fitted-parameter difference is at most 0.03 standard errors even at $L = 10^8$. At the 100-bin setting $|S_d|$ is inflated by 1.6–1.8%, which crosses one standard error at $L^* = 1.7 \times 10^7$ and so matters only for biobank-scale counts (Appendix D).

Sample sizes of order 10^5. At $n = 50{,}000$ and $n = 100{,}000$ (with $L = 10^8$) we tabulated the exact kernel and carried out full maximum-likelihood refits (Appendix B), which recover all four DFE parameters within 1.2 standard errors of truth. In the versions tested, POLYDFE does not run at even n and its bundled simulator does not generate data at this n; *dadi* requires diffusion grids finer than n, with entry error already at the percent level two orders of magnitude below this n; and FASTDFE's double-precision cast returns exact zeros in the deep tail of the spectrum. For the equilibrium PRF/DFE-mixture likelihood we are not aware of another implementation with a computed error bound at this n. One other method reaches biobank n with provable error control: the compressed discrete-time Wright–Fisher transition matrices of Spence et al. (2023), which Zeng et al. (2024) used to fit s_{het} at exome scale, carry total-variation bounds on the error of compressing the transition matrix, which is a different problem from evaluating the equilibrium selection-mixture likelihood considered here.

tool	layer	typical rel. error	largest measured	affected regime
POLYDFE	fixed-S, float64	$\leq 10^{-16}$–10^{-12}	1.7×10^{5} (no warning) / exits	$\lvert S\rvert \in [1.08, 5.0] \times 10^{-5}$ (fixed-S grid; end-to-end exit nodes from 1.0×10^{-5}), even $n \geq 90$; $S \geq 38$ high-i
POLYDFE	DFE mixing (qag)	10^{-7}–10^{-6}	3.2×10^{3} (no warning in a build that passes the window; the distributed binary exits)	carries the window error into mixtures
FIT*dadi*	γ-grid trapezoid	$+1.7 \times 10^{-4}$	1.2×10^{-2}	singletons under steep DFEs
FIT*dadi*	tail rule (2017 vs current)	10^{-4}	1.8×10^{-2} disagreement	DFE mass beyond the grid
dadi	PDE + extrapolation	10^{-6}–10^{-4} ($\lvert S\rvert \leq 100$); median 2–3% at $S \leq -10^{3}$, $n = 100$	0.57; 1.5×10^{5} (manual's grids)	strong selection, high i
FASTDFE	fixed-S + kernel	10^{-9}–10^{-5}	2.4×10^{-3} (100-bin setting)	none found at defaults

Table 2 Relative error of each numerical layer of the three implementations against the exact values of Section 2.3, over the tools' own documented parameter ranges (versions as in Section 5). "No warning" marks entries for which the tool reports success.

5.3. Dominance

Compared with the reference values of Eq. (5) (Section 2.4), FASTDFE's nested-trapezoid dominance evaluation has relative error of about 2×10^{-5} over most of the (S, h) grid we tested, rising to 1.6×10^{-3} at $(S, h) = (-100, 0.2)$ and 2.3×10^{-2} at $(-1000, 0.2)$. These largest errors occur at strong purifying selection on partially recessive mutations, which is the regime of most interest for inferring dominance (Figure 8, Appendix D). They are errors in individual spectrum entries; we have not yet measured their effect on fitted (h, DFE) parameters by a full dominance refit, and because the affected entries carry little likelihood weight at strong $\lvert S\rvert$ the parameter-level consequence may be small.

5.4. Implications for published estimates

The default deleterious-DFE pipelines of FIT*dadi* (current release) and FASTDFE agree with the exact likelihood at the parameter level at current data sizes, and our results give no reason to revisit published deleterious-DFE point estimates obtained with those defaults within their documented ranges. The same holds for POLYDFE estimates whose beneficial component is weak (Section 5.2). Two classes of estimate fall in ranges where we measured parameter-level differences: positive-selection quantities (p_b, S_b, α) from POLYDFE model C at $n \gtrsim 90$ chromosomes with $S_b \gtrsim 10$, and DFE scale or tail-mass estimates obtained with FIT*dadi*'s 2017 treatment of mass beyond the grid, particularly for steep DFEs. Estimates obtained in these settings may warrant re-examination, and the exact kernels released with this paper can be substituted for the corresponding layer of either tool.

6. Dominance from the X chromosome

A deleterious mutation may reduce fitness already in a single copy or only when both copies carry it, and the dominance coefficient h, which measures where a mutation falls between these extremes, determines how efficiently selection removes it while it is rare. The autosomal rare end, on which the fits of Section 3 rest, measures only the product $2hS$. Homozygotes enter at order x^2 in the sojourn density, and at allele frequencies below 3.4×10^{-3} that term is invisible, so h and S do not separate. This is the s_{het} regime of the constraint literature (Cassa et al. 2017; Weghorn et al. 2019; Agarwal et al. 2023; Zeng et al. 2024), which here follows from the structure of the likelihood. Agarwal, Fuller, Myers, and Przeworski make the same point for their loss-of-function estimates: on autosomes "only the compound *hs* parameter can be estimated" (Agarwal et al. 2023).

Only on the X chromosome does h enter separately from S: hemizygous males carry a single copy and experience S_X undiluted by h. This X-versus-autosome contrast is the classical source of information on dominance (Charlesworth et al. 1987) and was applied recently to the dominance of *adaptive* substitutions by McDonough et al. (2024). The X contrast does not determine h by itself, however. In the simulation study of Section 3.2, (h_0, k_h) lies on a one-dimensional ridge with the X-to-autosome population-size ratio N_X/N_A. The real data bound the position of this ridge ($N_X/N_A = 1.07$ [0.68, 1.22]) without bounding h alone, and the raw maximum is at the ridge corner ($h_0 \to 1$) in both classes. Whether the ridge reflects an exact degeneracy of the model can be checked only locally: at twelve parameter points at $n = 20$ the composite demography–selection–dominance map is provably locally identifiable (Section 4 and Appendix C8), which leaves the global question, and other n, open.

The X-chromosome data therefore give only a *one-sided bound at fixed* N_X/N_A. We adopt $N_X/N_A = 0.75$, the neutral equal-sex-ratio expectation, as a reference value, noting that empirical human X-to-autosome ratios vary with sex-biased demography (Hammer et al. 2008). At that value the profile optimum stays at the $h \to 1$ corner and the quasi-95% lower crossing is bracketed by a coarse scan (Figure 5). For high-confidence LoF it lies in $(0.5, 0.934]$, so LoF mutations appear to be at least semidominant at the rare end, consistent with the constraint literature's interpretation of LoF intolerance as selection on heterozygotes (Fuller et al. 2019). For missense it lies in $(0.15, 0.5]$, so strong recessivity appears to be disfavored there, in agreement with the $h \gtrsim 0.15$ floor Kyriazis and Lohmueller obtain for nonsynonymous mutations by fitting dominance models to autosomal spectra (Kyriazis and Lohmueller 2024). We do not report unconditional dominance intervals, since the simulation study showed that they do not close. The X-side LoF data are thin ($\approx$ 3.1K transition plus 3.1K transversion sites), and they constrain LoF dominance correspondingly weakly; most of the X-chromosome information on dominance comes from the missense class.

Conditional on the assumed N_X/N_A, then, a mutation that disables a gene appears to do more than half its harm already in the heterozygote; the bound is coarse, compatible with anything from h just above $\frac{1}{2}$ to full dominance. The assumed value $N_X/N_A = 0.75$ lies near the lower edge of the fitted interval, and we give no bound at the fitted ratio itself.

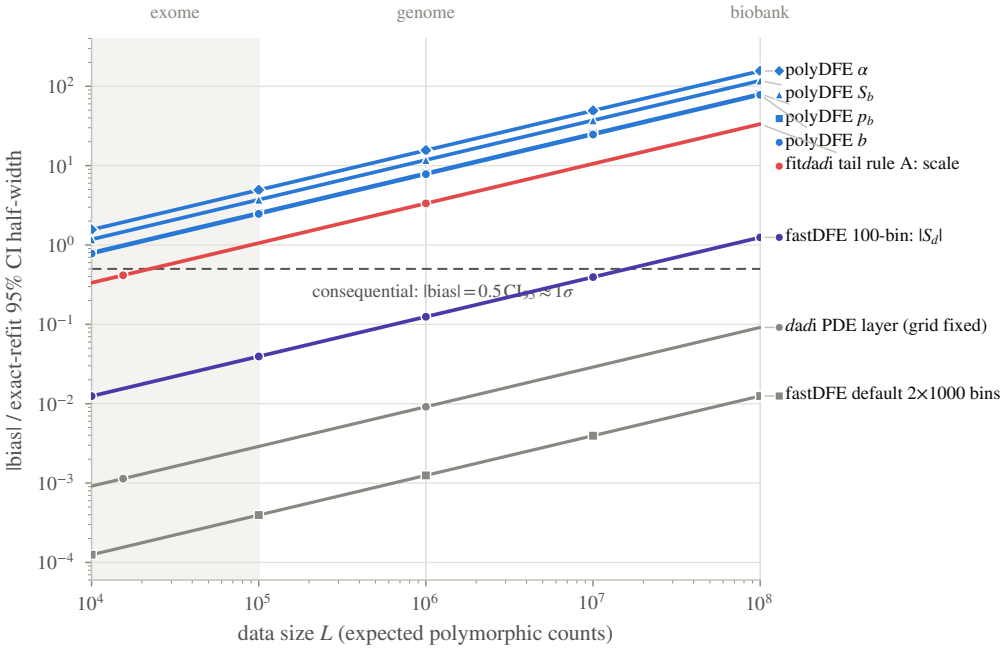

Figure 4 Difference between parameters fitted with each tool's numerics and with the exact likelihood, in units of the exact-likelihood 95% confidence half-width, versus data size L (expected total polymorphic count). Configurations: POLYDFE model C with beneficial selection at $n = 101$, truth $S_b = 40$ (its α_{dfe}, S_b, p_b, b curves); FIT*dadi*'s 2017 rule for mass beyond the grid on the distributed example's DFE; FASTDFE at the 100-bin setting; FASTDFE defaults; and the *dadi* diffusion layer. The difference is a fixed offset in each configuration, so curves scale as $\sqrt{L}$ between measured points (markers); the first three configurations cross the one-standard-error line (dotted) within realistic data sizes and the last two remain well below it.

7. Discussion

We have applied exact evaluation of the selected site-frequency spectrum to four problems. In the largest human variant catalog, gnomAD, roughly 74% of the fitted DFE mass for mutations that disable a gene lies in the strongest-selection window the spectrum reaches ($|S| \geq 10^4$), one that classical sample sizes could not resolve. If S is converted to s with the textbook long-term N_e these mutations would be close to lethal in heterozygotes, while the larger effective sizes appropriate to recent growth imply substantially weaker effects, and we do not attempt to decide between the two conversions here (Section 3.3). At each selection strength we sampled below the onset of distinguishability, some explicit neutral demographic history imitates the selected spectrum, whereas at each sampled additive strength $S \geq 1.375$ ($h = \frac{1}{2}$) no history in the panmictic Kingman class matches it at the sample size we analyzed ($n = 20$). Two strong purifying strengths and one partially recessive point remain unresolved (Table 8), and Appendix C10 states the assumptions on which these results rest. The inference programs POLYDFE, FIT*dadi*, and FASTDFE agree with the exact likelihood to a small fraction of a confidence interval at the sample sizes where they are routinely used, and differ from it in specific parameter ranges that we identify. At sample sizes of order 10^5 none of the three currently returns a usable likelihood in the versions we tested. On dominance, autosomal rare-end data measure selection only through the product $2hS$, and the X chromosome partly breaks that degeneracy. The resulting coarse one-sided bounds, conditional on an assumed X-to-autosome population-size ratio, suggest that a mutation that disables a gene does more than half its harm in a single copy. All four results rest on one method, the exact solution of the expected spectrum and its derivative stack over $\mathbb{Q}(e^{-S})$, evaluated with rigorous or validated error bounds (Appendix B) small enough that numerical error does not contribute to the reported uncertainties.

The closed form Eq. (2) itself is not new; it appears in print at least twice and is the expression that POLYDFE and FASTDFE implement, and the contiguity relations are Gauss's. The new methodological elements are the conditioning analysis that explains why double precision fails on this class of problems; the exact $\mathbb{Q}(e^{-S})$ solution of the whole SFS vector and its derivative stack, with error-bounded evaluation to biobank sample sizes and an error-bounded DFE-mixing layer; and the entire-deformation structure with an error-controlled evaluation of the dominance spectrum. The new applications are the parameter-level comparison of POLYDFE, FIT*dadi*, and FASTDFE with exact values, and a DFE fit to gnomAD in the strong-selection regime that biobank sample sizes make accessible, reported together with its identifiability limits. Parts of the analytic route have precedents: Vogl and Bergman (2015), developing maximum-likelihood machinery for the adjacent boundary-mutation spectrum, give an analytic score equation in γ for that spectrum (their Eq. (42) and Appendix A), which they solve numerically in double precision. Their discussion states that for the full Wright–Fisher diffusion with directional selection "no analytically tractable results are available." Analytic score and curvature integrals for the PRF spectrum itself go back to Bustamante et al. (2001). Beyond them, we supply the exact derivative stack to all orders for the PRF selection spectrum and its DFE mixtures, with error-bounded evaluation.

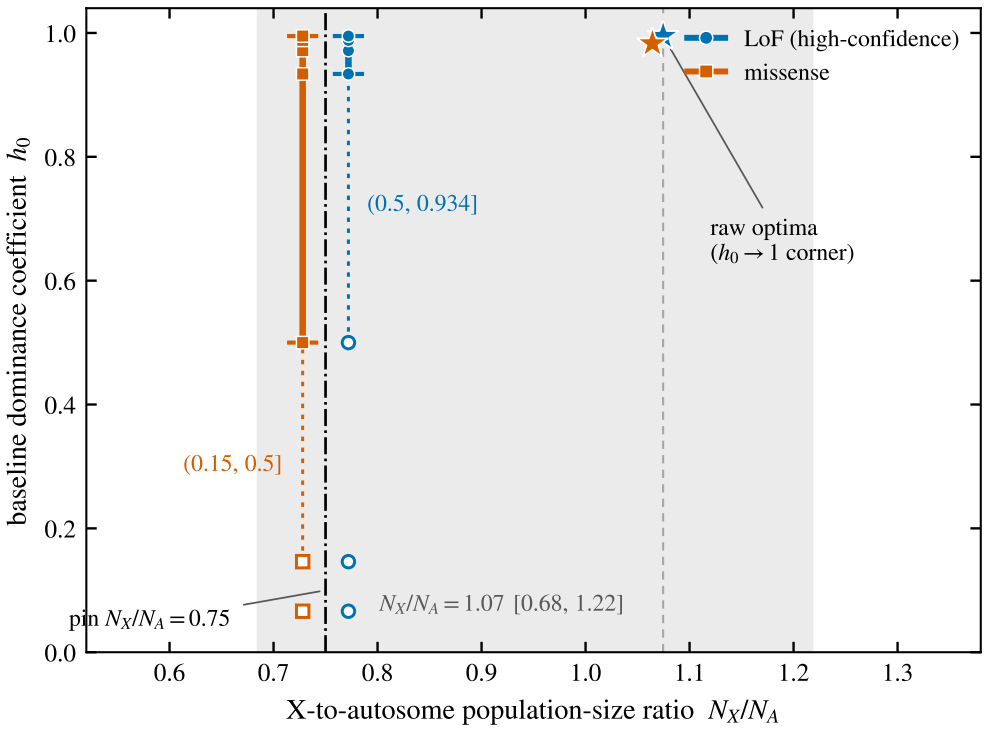

Figure 5 Profile scan of the baseline dominance coefficient h_0 against the X-to-autosome population-size ratio N_X/N_A. The likelihood has a ridge along which h_0 trades off against N_X/N_A; the data bound the position of this ridge ($N_X/N_A = 1.07$ [0.68, 1.22]) but not h_0 alone, and the raw optimum is at the $h_0 \to 1$ corner in both classes. At $N_X/N_A = 0.75$ the quasi-95% lower crossing falls in (0.5, 0.934] for high-confidence LoF and in (0.15, 0.5] for missense. The marked h_0 values are *one-sided bounds* at fixed N_X/N_A, bracketed by a coarse profile scan and conditional on that ratio; they are not confidence intervals (Section 3.2).

These claims are subject to several limitations. The analysis is panmictic: population structure and linked selection enter through the synonymous calibration and the robustness checks and do not enter the likelihood, so all fitted quantities are conditional on that demography treatment. Rigorous error control covers the equilibrium PRF selection integral and the DFE mixture over it, whereas the inference programs combine these with a nonequilibrium demographic correction (FIT*dadi* through the diffusion solver, POLYDFE and FASTDFE through nuisance multipliers). We test the selection component and, where the tools use nuisance-based corrections, the full likelihood conditional on them, but not a demographic model. Extending rigorous error control to the nonequilibrium demographic layer is therefore an important direction for future work. The time-dependent PRF density solves a forward equation whose moment systems again close on hypergeometric-type towers (Evans et al. 2007; Živković et al. 2015), and an error-controlled treatment there would likely remove the last uncontrolled layer of these inference methods. The dominance treatment, moreover, is at present for the stationary spectrum only (Section 6). The evaluation of the dominance spectrum can also be improved: the heat relation makes Eq. (5) a two-variable holonomic system of low rank, and transporting in $\tilde{q}$ from the Kummer line is likely the better evaluation method for recessive lethals. These limitations apply to the evaluation method, the software comparison, and the gnomAD results; the mimicry and identifiability results rest instead on the assumptions collected in Appendix C10. Finally, the linked-selection results reported alongside the main results are of different kinds (theorems, exact computations, data stratifications, and simulations), each with its own assumptions, and we think it would be premature to combine them into a causal claim about linkage. In particular, deciding whether the dependence of the fitted tail on recombination environment (Section 3.3) reflects linkage or gene composition would probably need nulls that break one association at a time.

Outlook. Given exact evaluation, the singleton and doubleton counts of the coming million-exome catalogs can be compared with expected values whose numerical uncertainty is stated. Exactly computed single-site spectra are also the foundation for multi-site analyses of the same catalogs, in the two-site line of Fenton et al. (2025), and such an analysis is the subject of a companion paper. Further extensions suggested by the present results include quantitative exclusion distances (the standoffs of Section 4) beyond (n=20, h=$\frac{1}{2}$, $S \in [1.375, 100]$), error bounds on the replicate uncertainty of the linked-selection simulations, and nulls and stratifications that separate linkage from gene composition. On the theory side they include separating functionals for folded or truncated spectra, time-inhomogeneous towers, and a representation lemma for Λ-coalescents.

The mimicry results also bear more broadly on how the catalogs are interpreted. They suggest that at the rare end a good fit under an estimated demographic model is not by itself evidence that the demography is right, and that weak selection cannot be separated from history by fit quality alone. For moderate and strong positive additive selection, within the panmictic Kingman class and on the assumptions of Appendix C10, no history matches, so that a misfit is itself informative about selection. Both conclusions presuppose that the likelihood itself is evalu-

ated reliably, which at biobank sample sizes likely requires the kind of error control developed here. The catalogs are now large enough that the uncertainty in a fitted quantity may often be limited by its identifiability more than by the number of genomes, which suggests that identifiability limits could usefully be reported as results alongside the confidence intervals.

Data and code availability

All numerical results in this paper, including the linked-selection simulations and stratified fits, can be reproduced with the code and data package that will be deposited in a public repository [to be specified] on acceptance. The package includes the exact computational core as an importable library (`popcorn`) and tables of error-bounded expected spectra, among them spectra evaluated on *dadi*'s own cache grids (Section 5.2), which an existing tool can adopt directly. It also includes the code for the software comparison of Section 5 and for the dominance analysis of Section 2.4. The gnomAD v4.1 data used are described in Section 3; the primary data are CC0 1.0.

Literature Cited

I. Agarwal, Z. L. Fuller, S. R. Myers, and M. Przeworski, "Relating pathogenic loss-of-function mutations in humans to their evolutionary fitness costs," eLife **12**, e83172 (2023).

A. F. Agrawal and M. C. Whitlock, "Inferences about the distribution of dominance drawn from yeast gene knockout data," Genetics **187**, 553–566 (2011).

H. Akashi, N. Osada, and T. Ohta, "Weak selection and protein evolution," Genetics **192**, 15–31 (2012).

E. Árnason, J. Koskela, K. Halldórsdóttir, and B. Eldon, "Sweepstakes reproductive success via pervasive and recurrent selective sweeps," eLife **12**, e80781 (2023).

D. J. Balick, R. Do, C. A. Cassa, D. Reich, and S. R. Sunyaev, "Dominance of deleterious alleles controls the response to a population bottleneck," PLoS Genetics **11**, e1005436 (2015).

A. Bhaskar and Y. S. Song, "Descartes' rule of signs and the identifiability of population demographic models from genomic variation data," Ann. Statist. **42**, 2469–2493 (2014).

A. R. Boyko, S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez, K. E. Lohmueller, M. D. Adams, S. Schmidt, J. J. Sninsky, S. R. Sunyaev, T. J. White, R. Nielsen, A. G. Clark, and C. D. Bustamante, "Assessing the evolutionary impact of amino acid mutations in the human genome," PLoS Genetics **4**, e1000083 (2008).

C. D. Bustamante, J. Wakeley, S. Sawyer, and D. L. Hartl, "Directional selection and the site-frequency spectrum," Genetics **159**, 1779–1788 (2001).

C. A. Cassa *et al.*, "Estimating the selective effects of heterozygous protein-truncating variants from human exome data," Nature Genetics **49**, 806–810 (2017).

B. Charlesworth, J. A. Coyne, and N. H. Barton, "The relative rates of evolution of sex chromosomes and autosomes," Am. Nat. **130**, 113–146 (1987).

S. Chen *et al.* (Genome Aggregation Database Consortium), "A genomic mutational constraint map using variation in 76,156 human genomes," Nature **625**, 92–100 (2024).

M. M. Desai and J. B. Plotkin, "The polymorphism frequency spectrum of finitely many sites under selection," Genetics **180**, 2175–2191 (2008).

B. Eldon and J. Wakeley, "Coalescent processes when the distribution of offspring number among individuals is highly skewed," Genetics **172**, 2621–2633 (2006).

S. N. Evans, Y. Shvets, and M. Slatkin, "Non-equilibrium theory of the allele frequency spectrum," Theor. Popul. Biol. **71**, 109–119 (2007).

A. Eyre-Walker, M. Woolfit, and T. Phelps, "The distribution of fitness effects of new deleterious amino acid mutations in humans," Genetics **173**, 891–900 (2006).

E. F. Fenton, D. P. Rice, J. Novembre, and M. M. Desai, "Detecting deviations from Kingman coalescence using 2-site frequency spectra," Genetics **229**, iyaf023 (2025).

L. Frisse, R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack, and A. Di Rienzo, "Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels," Am. J. Hum. Genet. **69**, 831–843 (2001).

Z. L. Fuller, J. J. Berg, H. Mostafavi, G. Sella, and M. Przeworski, "Measuring intolerance to mutation in human genetics," Nature Genetics **51**, 772–776 (2019).

M. Galassi *et al.*, *GNU Scientific Library Reference Manual*, 3rd ed. (2009), ISBN 0954612078; https://www.gnu.org/software/gsl/.

W. Gautschi, "Computational aspects of three-term recurrence relations," SIAM Rev. **9**, 24–82 (1967).

J. Gay, S. Myers, and G. McVean, "Estimating meiotic gene conversion rates from population genetic data," Genetics **177**, 881–894 (2007).

R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, "Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data," PLoS Genetics **5**, e1000695 (2009).

M. F. Hammer, F. L. Mendez, M. P. Cox, A. E. Woerner, and J. D. Wall, "Sex-biased evolutionary forces shape genomic patterns of human diversity," PLoS Genetics **4**, e1000202 (2008).

C. D. Huber, A. Durvasula, A. M. Hancock, and K. E. Lohmueller, "Gene expression drives the evolution of dominance," Nat. Commun. **9**, 2750 (2018).

J. D. Jensen, Y. Kim, V. Bauer DuMont, C. F. Aquadro, and C. D. Bustamante, "Distinguishing between selective sweeps and demography using DNA polymorphism data," Genetics **170**, 1401–1410 (2005).

F. Johansson, "Arb: efficient arbitrary-precision midpoint-radius interval arithmetic," IEEE Trans. Comput. **66**, 1281–1292 (2017).

F. Johansson, "Computing hypergeometric functions rigorously," ACM Trans. Math. Softw. **45**, 30 (2019).

K. J. Karczewski *et al.*, "The mutational constraint spectrum quantified from variation in 141,456 humans," Nature **581**, 434–443 (2020).

B. Y. Kim, C. D. Huber, and K. E. Lohmueller, "Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples," Genetics **206**, 345–361 (2017).

M. Kimura, "The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations," Genetics **61**, 893–903 (1969).

J. F. C. Kingman, "The coalescent," Stochastic Process. Appl. **13**, 235–248 (1982).

A. Kousathanas and P. D. Keightley, "A comparison of models to infer the distribution of fitness effects of new mutations," Genetics **193**, 1197–1208 (2013).

C. C. Kyriazis and K. E. Lohmueller, "Constraining models of dominance for nonsynonymous mutations in the human genome," PLoS Genetics **20**, e1011198 (2024).

Y. McDonough, F. Ruzicka, and T. Connallon, "Reconciling theories of dominance with the relative rates of adaptive substitu-

tion on sex chromosomes and autosomes," Proc. Natl. Acad. Sci. USA **121**, e2406335121 (2024).
The mpmath development team, *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)*, 2023; https://mpmath.org/.
S. Myers, C. Fefferman, and N. Patterson, "Can one learn history from the allelic spectrum?," Theor. Popul. Biol. **73**, 342–348 (2008).
NIST Digital Library of Mathematical Functions, Release 1.2.7 of 2026-06-15, F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.; https://dlmf.nist.gov/.
F. W. J. Olver, "Numerical solution of second-order linear difference equations," J. Res. Nat. Bur. Standards **71B**, 111–129 (1967).
J. Pitman, "Coalescents with multiple collisions," Ann. Probab. **27**, 1870–1902 (1999).
R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970).
Z. Rosen, A. Bhaskar, S. Roch, and Y. S. Song, "Geometry of the sample frequency spectrum and the perils of demographic inference," Genetics **210**, 665–682 (2018).
Z. Rosen, G. Scholten, and C. Vinzant, "Sparse moments of univariate step functions and allele frequency spectra," Vietnam J. Math. **50**, 523–544 (2022).
S. Sagitov, "The general coalescent with asynchronous mergers of ancestral lines," J. Appl. Probab. **36**, 1116–1125 (1999).
S. A. Sawyer and D. L. Hartl, "Population genetics of polymorphism and divergence," Genetics **132**, 1161–1176 (1992).
J. Sendrowski and T. Bataillon, "fastDFE: fast and flexible inference of the distribution of fitness effects," Mol. Biol. Evol. **41**, msae070 (2024).
J. Sendrowski, B. M. Pedersen, J. Bergman, V. Pankratov, and T. Bataillon, "Comparison of the distribution of fitness effects across primates," Genetics, iyag252 (2026).
Y. S. Song and M. Steinrücken, "A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection," Genetics **190**, 1117–1129 (2012).
J. P. Spence, T. Zeng, H. Mostafavi, and J. K. Pritchard, "Scaling the discrete-time Wright–Fisher model to biobank-scale datasets," Genetics **225**, iyad168 (2023).
P. Tataru and T. Bataillon, "polyDFEv2.0: testing for invariance of the distribution of fitness effects within and across species," Bioinformatics **35**, 2868–2869 (2019).
P. Tataru, M. Mollion, S. Glémin, and T. Bataillon, "Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data," Genetics **207**, 1103–1119 (2017).
S. Tavaré, "Line-of-descent and genealogical processes, and their applications in population genetics models," Theor. Popul. Biol. **26**, 119–164 (1984).
C. Vogl and J. Bergman, "Inference of directional selection and mutation parameters assuming equilibrium," Theor. Popul. Biol. **106**, 71–82 (2015).
C. Vogl and F. Clemente, "The allele-frequency spectrum in a decoupled Moran model with mutation, drift, and directional selection, assuming small mutation rates," Theor. Popul. Biol. **81**, 197–209 (2012).
D. Weghorn, D. J. Balick, C. Cassa, J. A. Kosmicki, M. J. Daly, D. R. Beier, and S. R. Sunyaev, "Applicability of the mutation-selection balance model to population genetics of heterozygous protein-truncating variants in humans," Mol. Biol. Evol. **36**, 1701–1710 (2019).
S. Williamson, A. Fledel-Alon, and C. D. Bustamante, "Population genetics of polymorphism and divergence for diploid selection models with arbitrary dominance," Genetics **168**, 463–475 (2004).
S. Wright, "The distribution of gene frequencies under irreversible mutation," Proc. Natl. Acad. Sci. USA **24**, 253–259 (1938).
T. Zeng, J. P. Spence, H. Mostafavi, and J. K. Pritchard, "Bayesian estimation of gene constraint from an evolutionary model with gene features," Nature Genetics **56**, 1632–1643 (2024).
D. Živković, M. Steinrücken, Y. S. Song, and W. Stephan, "Transition densities and sample frequency spectra of diffusion processes with selection and variable population size," Genetics **200**, 601–617 (2015).

i	w_i
1	281653700641825/3855675953189414
2	137816638489010/1927837976594707
3	175559744490794/1927837976594707
4	−90085544044831/3855675953189414
5	578889745057952/1927837976594707
6	−140737488355328/1927837976594707
7	−140737488355328/1927837976594707
8	−140737488355328/1927837976594707
9	1385636716982648/1927837976594707
10	−158800498507535/3855675953189414
11	−140737488355328/1927837976594707
12	−140737488355328/1927837976594707
13	−140737488355328/1927837976594707
14	−140737488355328/1927837976594707
15	−140737488355328/1927837976594707
16	1
17	−140737488355328/1927837976594707
18	−140737488355328/1927837976594707
19	−140737488355328/1927837976594707
w_0	$-4898883/2^{26}$

Table 3 The shared separating functional. The nineteen rational coefficients and the dyadic constant w_0, printed in full and in lowest terms, with the slack shift already included in w_0. Together with the margins of Table 4, this table is the complete proof data for the six exclusions $S = 2$–100.

A. Coefficients of the shared separating functional

Table 3 gives the separating functional of Section 4 in exact form: the nineteen rational coefficients $w_1, \dots, w_{19}$ and the dyadic constant $w_0 = -4898883/2^{26}$. The slack shift is already included in w_0, so the constant should be used as printed. The affine functional $F(m) = \sum_{i=1}^{19} w_i m_i + w_0$ is proven nonnegative on the entire neutral curve and proven negative on all six excluded spectra $S \in \{2, 5, 10, 20, 50, 100\}$, with the margins of Table 4; checking both facts needs only rational arithmetic and an exact univariate nonnegativity check on the polynomial induced along the curve.

B. Numerical validation and reproducibility

Levels of error control. Each computed quantity in this paper carries one of three levels of error control. The first is a rigorous bound: arb ball radii (route B), exact rational heights (route E, where the cancellation bound is computed explicitly), and proven geometric series tails (route D and the reference series). The second is validated evaluation without a rigorous bound: the composite Gauss–Legendre layers (kernel cache, dominance reference evaluator) rely on agreement between two quadrature orders and with independent routes, and because a shared systematic error could pass both orders, each quadrature layer is also compared with an object of the first level. The third is library trust: the transcendental leaves are taken as correctly rounded by their implementations, and each is compared with at least one exact-rational spot value.

Spectrum evaluation. We compared the closed form Eq. (2) with a direct error-bounded quadrature of Eq. (1) and with an exact-rational series with proven tails at twenty reference points ($S = \pm 1000$, $n \le 1000$), and then tested routes E and B against these reference values and gradients at 22 out-of-sample points ($n \in \{23, 137, 1024, 10007\}$, $|S| \in [10^{-4}, 2000]$). The DFE-mixing layer was compared with an independent implementation of the mixing integral and its gradient with high-order finite differences. The two dominance evaluators of Section 2.4 were compared at ten points spanning $S \in [-100, 500]$, $h \in [0.05, 0.95]$, and the reference evaluator was in addition checked through its $h = \frac{1}{2}$ limit against Eq. (2) and through the agreement of its two quadrature degrees over the whole grid, including $(S, h) = (-1000, 0.2)$. Wherever two independent implementations were compared, they agreed to at least 40.7 significant figures, close to the precision ceiling of the comparison. *dadi*'s diffusion solver, given the same (S, h) points through its own dominance parameterization, reproduces the reference dominance spectra to 2.3×10^{-4} relative error or better, and to 2×10^{-5} on a tail entry of order 10^{-26}. A mismatch in the polarization of h between our convention and *dadi*'s would have appeared as a discrepancy of up to a factor of 23; none was seen.

gnomAD fit. We compared each layer of the fit of Section 3 (rare-entry spectra at sample sizes bracketing the gnomAD sample size, the rare-end dominance reduction, the kernel caches over 144 out-of-sample cells, and the double-precision optimization stage) with the error-bounded route. An error-bounded re-evaluation at the optimum of each primary fit reproduces the fast log-likelihood with a mass-weighted kernel error bound of $(3\text{–}4) \times 10^{-4}$, two orders of magnitude below the model-error floor $\hat{\phi}$; flagged cache cells carry a separate bias bound of $\le 2.5 \times 10^{-6}$ of a class total per S-node. In the simulation study of Section 3.2 all 13 model parameters were recovered within two standard errors or within their profile intervals (fitted deviance 65 over ~ 105 cells at Poisson truth), and the interior-bin spike test quoted there has likelihood-ratio statistic $D = 0.0095$ ($p = 0.46$). The raw log-likelihood has magnitude $\sim 1.9 \times 10^8$, at which double-precision optimization scatters by ~ 2 likelihood units between restarts; we therefore optimize on the deviance scale. The primary fit has deviance 63,464; the robustness fit that includes CpG transitions (Section 3) has deviance 1,285,483.

Linked-selection simulations. In the nine forward-simulated regimes of Section 3.3, the fitted selection strengths of the neutral windows all lie in $[-6.09, -0.36]$, far short of both the strongest selection the simulations delivered ($\sim -5 \times 10^3$) and the -10^4 window edge. All 2-unit likelihood intervals lie inside $[-11.91, +1.07]$, and the largest upper endpoint, $+1.064$ in the zero-recombination background-selection regime, is 0.31 below the $+1.375$ bracket edge. The simulated neutral control window is fit at $\hat{S} = -2.0$ ($R = 128$ replicates) with all intervals inside $(-10^4, +10^4)$. The same fitting procedure, run on exactly computed spectra of known S, fits a neutral spectrum at $\hat{S} = -0.5$, recovers $S = -2$ and $S = -50$ exactly, and tracks the truth monotonically through the resolved regime. A strongly selected site inserted into the simulations themselves would not serve as a control: the two-kilobase locus around it carries 7.71 projected sites from the regime's own flanking variants against about 0.02 expected copies from the inserted site ($S = -10^4$ lies 54.8 log-likelihood units below the optimum), so such a comparison would measure flank contamination and would say little about sensitivity.

Mimicry computations. Controls with known answers (spectra inside the neutral hull, and a constructed spectrum outside it with an exactly known violation) were run at each stage of the computation and at each hull resolution. The control spectrum outside the hull covers, in particular, the case in which a separating ray grazes the curve at an irrational near-tangency, where a slack step is required. The configurations in which a control was not recovered are listed at the end of Section 2.5.2, and no result from those configurations is used.

Software tests. In the positive-control refits of Section 5.2 each tool's fit and the exact fit agree to 10^{-4} of a confidence interval at all data sizes, the exact side recovers its own truth to 5×10^{-6} relative or better, and goodness-of-fit z-scores lie within $[-0.5, 1.4]$. At biobank sample sizes ($n = 50{,}000$ and $n = 100{,}000$) two simplifications are used: a reduced quadrature degree pair, identical on the truth and fit sides so that its error cancels in the bias comparison, and a double-precision copy of the exact kernel in the optimizer, checked against the exact kernel at the optimum.

Numerical pitfalls. Four pitfalls we met may also affect users of the released code. (1) General-purpose adaptive quadrature (mpmath's Gauss–Legendre and tanh-sinh rules) can underconverge on the smooth exponential-ramp panels of this problem, returning panel integrals wrong by 1% with optimistic error estimates that do not improve as working precision is raised. (2) Two Kummer forms of Eq. (2) evaluated through the same library share its internal reflection, so their agreement is not an independent check; one of the two evaluations should be an exact-rational series. (3) A quantity evaluated at two representations of the same S (a decimal and its binary floating-point image) differs by the full cancellation factor of Section 2.2, so a single exact rational S must be threaded through all routes. The elementary seed ${}_1F_1(1; n; z)$ likewise loses all working precision to Poisson-tail cancellation if used below its crossover $z \approx 2n$. (4) Endpoint-singular DFE densities (Gamma shape $b < 1$) place an algebraic singularity under a Gauss rule unless the neutral end of the mixing integral is handled analytically.

C. Mimicry analysis: details

This appendix gives the proofs and computations summarized in Section 4. Sample size is $n = 20$ unless noted, and every statement is relative to the panmictic Kingman class (Appendix C10).

C1. Proof of the representation lemma

Let $N_i(\tau)$ be the number of ancestral lineages of the sample that subtend exactly i of the n sampled genomes at real time τ before the present, so that the branch length carrying mutations of sample count i is $L_i = \int_0^\infty N_i(\tau)\,\mathrm{d}\tau$ and the expected unnormalized spectrum is $\xi_i = \theta\,\mathbb{E}[L_i]$ up to the common factor θ. Under a size history $N(\tau)$ the pairwise coalescence rate at time τ is $1/2N(\tau)$, so in the intensity time $g(\tau) = \int_0^\tau \mathrm{d}\tau'/2N(\tau')$ the ancestral process is the standard Kingman n-coalescent, independent of N (Kingman 1982). Hence the law of the lineage configuration at intensity g does not depend on the history, and

$$\mathbb{E}[L_i] = \int_0^\infty \mathbb{E}[N_i](g)\,\frac{\mathrm{d}\tau}{\mathrm{d}g}\,\mathrm{d}g = \int_0^\infty \gamma_i(e^{-g})\;2N(\tau(g))\,\mathrm{d}g,$$

which is the first claim with $\mu(\mathrm{d}g) = 2N(\tau(g))\,\mathrm{d}g \geq 0$; finiteness of the total branch length is finiteness of $\int \sum_i \gamma_i\,\mathrm{d}\mu$. It remains to identify γ_i. Conditional on k lineages remaining, exchangeability of the Kingman coalescent makes the sizes of the k families a uniformly random composition of n into k positive parts, so a given lineage subtends exactly i genomes with probability $\binom{n-i-1}{k-2}/\binom{n-1}{k-1}$ $(1 \leq i \leq n-k+1)$, and the expected number of such lineages is k times this. Averaging over k with the distribution of the number of ancestral lineages at intensity g (Tavaré 1984),

$$P_{n,k}(t) \;=\; \sum_{j=k}^{n} \frac{(2j-1)\,(-1)^{j-k}\,k_{(j-1)}\,n_{[j]}}{k!\,(j-k)!\,n_{(j)}}\;t^{\binom{j}{2}}, \qquad t = e^{-g}, \tag{C1}$$

where $a_{(j)} = a(a+1)\cdots(a+j-1)$ and $a_{[j]} = a(a-1)\cdots(a-j+1)$, gives Eq. (8). For the second claim write $s(g) = \sum_i \gamma_i(e^{-g}) > 0$; then $m = \xi/\sum_i \xi_i = \int \Gamma_n(e^{-g})\,s(g)\mu(\mathrm{d}g)/\int s\,\mathrm{d}\mu$, an average of points of the curve against the probability measure $\nu = s\mu/\int s\,\mathrm{d}\mu$. An average of points of a set lies in its closed convex hull, and taking closures gives $\mathcal{N}_n \subseteq \overline{\mathrm{conv}}\,\Gamma_n((0,1])$. Finally, if $w \cdot \Gamma_n(t) + w_0 \geq 0$ for all $t \in (0,1]$ then integrating against ν gives $w \cdot m + w_0 \geq 0$ for every history, and the inequality is inherited by any nonnegative combination of histories across loci. The endpoints of the curve are the star genealogy ($t = 1$: Γ_n is the pure-singleton spectrum) and the two-lineage limit ($t \to 0$: Γ_n tends to the uniform spectrum $1/(n-1)$), neither of which is attained by a genuine history, which is why closures appear in Eq. (10). □

C2. Proof methods for the separation inequalities

For the nonnegativity of a separating functional along the neutral curve (Section 2.5.2), the first method (B) expands the induced polynomial in the Bernstein basis, where nonnegativity of all coefficients proves nonnegativity of the polynomial; mixed signs trigger exact bisection over dyadic rationals until the coefficients on each subinterval are nonnegative. The second (R) isolates the real roots of the squarefree part in exact arithmetic: an empty root count in the open interval, together with nonnegative values at the endpoints and a positive sign at one interior rational point, proves the same statement independently. For the sign at the selected spectrum, whose entries lie in $\mathbb{Q}(e^{-S})$, the first method (P) evaluates the functional symbolically, by an exact zero-test on the underlying rational function followed by a rigorous enclosure of e^{-S} refined until zero is excluded; the refinement terminates by Lindemann's theorem. The second (D) re-evaluates the same quantity in 512-bit ball arithmetic (Johansson 2019) in an independent implementation and accepts a sign only when the entire ball excludes zero. Each convex representation reported as a mimicry result is re-checked in arithmetic independent of the search that proposed it. When no such combination exists over the samples, the same computation returns an exact separating vector, and these vectors seed the search for separating functionals.

C3. The individual exclusions and their margins

Twenty-five distinct (S, h) points at $n = 20$ are excluded: no panmictic Kingman size history, of any form, reproduces these selected spectra, under the assumptions of Appendix C10 (assumption A7 applies to all $h \neq \frac{1}{2}$ rows). Table 4 lists the semidominant ($h = \frac{1}{2}$) exclusions with their proven violation margins and Table 5 the dominance grid, and Figure 3 displays all sampled points.

Exclusion of every single history is not yet exclusion of linked selection, which in its standard first-order description gives each locus its own history: the pooled spectrum of many loci is then a weighted mixture of many neutral histories at once, and an

S	$\log_{10}$ violation margin
1.375	−350.7
1.5	−7.827
1.625	−7.086
1.75	−6.928
1.875	−5.632
2	−5.46
5	−3.34
10	−2.72
20	−2.28
50	−1.97
100	−1.87

Table 4 Proven exclusions at each sampled additive strength from $S = 1.375$ to $S = 100$. The margin is the proven size of the violation of the separating inequality ($\log_{10}$ of a rigorous enclosure of an exact rational). Results are relative to the panmictic Kingman class only (Appendix C10); no monotonicity in S is assumed, and each row is proved separately. The six rows $S = 2$–100 share the single separating functional of Section 4.

exclusion must apply to that mixture as a whole. The exclusions extend to such mixtures (the transfer theorem of Section 4), because each separating functional is proven nonnegative on the entire neutral curve, so that the inequality holds pointwise at every history and passes through any nonnegative pooling unchanged. The same functionals therefore exclude all finite or random per-locus assignments of neutral histories and all unconditional poolings of them, that is, rescaling-class background selection of arbitrary strength, genomic granularity and parameter values. Wherever a proven lower bound on the ℓ_1 distance to the selected spectrum has been computed (a standoff, which we use in place of the raw margin of the functional; Appendix C4 and Figure 6), the same bound holds against these mixtures. No additive ($h = \frac{1}{2}$) point with $S < 0$ is excluded, and nothing is excluded at any $n \neq 20$. Pooling here means unconditional linear pooling only (pooled expected counts, or per-locus spectra combined with data-independent weights); pooling conditioned on observed counts size-biases the genealogy ensemble and lies outside the transfer theorem. Where the exact linked models leave the neutral class itself, nothing here applies; that boundary is examined case by case in Appendix C7.

C4. The closure theorem and the quantitative standoffs

The closure theorem, Eq. (10), states that the closure $\mathcal{N}_n$ of the set of achievable neutral spectra (Section 2.5.1) equals the closed convex hull of the curve Γ_n for every $n \geq 3$. The outer half is the representation lemma. The inner half is constructive: for any point of the hull one builds an explicit sequence of genuine step histories whose spectra converge to it at a computable rate. The closure operator is essential: the achievable set itself is *not* closed (the curve's $t = 0$ endpoint is a limit that no genuine history attains; an exact history attaining it would force a nonnegative measure to have zero total mass).

The construction is this. Write $\tilde{\gamma}_i(t) = \gamma_i(t)/t$, a polynomial because every γ_i carries the factor t, let G_{ij} be the coefficient of $t^{\binom{j}{2}-1}$ in $\tilde{\gamma}_i$ ($j = 2, \ldots, n$), and let $\tilde{s}(t) = \sum_i \tilde{\gamma}_i(t)$, positive on $[0, 1]$. Writing $N(t)$ for the population size at depth t, the change of variables $t = e^{-g}$ in Lemma 1 gives $\xi = \int_0^1 \tilde{\gamma}(t)\, 2N(t)\, \mathrm{d}t$; conversely, every positive step function of depth is the size profile of a genuine history, piecewise constant in generations with the same number of epochs, because real time $\tau = \int_0^g 2N\, \mathrm{d}g'$ is then increasing and unbounded. A point of the hull is a finite mixture $m^* = \sum_{k=1}^{r} \lambda_k \Gamma_n(t_k)$ with $\lambda_k > 0$, $\sum_k \lambda_k = 1$, $t_k \in [0, 1]$ and $r \leq n - 1$ (Carathéodory). For $\delta, \epsilon \in (0, 1]$ we take the step history

$$2N_{\delta,\epsilon}(t) \;=\; \epsilon + \frac{1}{\delta} \sum_{k=1}^{r} \frac{\lambda_k}{\tilde{s}(t_k)}\, \mathbf{1}\{\, \bar{t}_k - \delta < t \leq \bar{t}_k \,\}, \qquad \bar{t}_k = \max(t_k, \delta), \tag{C2}$$

with at most $2r + 1 \leq 2n - 1$ epochs: a population of negligible size ϵ except in r short windows of depth at the t_k, where it is large and where the genealogy accrues almost all of its branch length. The window average of t^p differs from t_k^p by at most $p\delta$ for each exponent $p = \binom{j}{2} - 1 \geq 2$ and not at all for $p = 0$, and the floor adds at most ϵ to each moment $\int_0^1 t^p\, 2N\, \mathrm{d}t$, so with $\|G\| = \max_j \sum_i |G_{ij}|$ the normalized spectrum $m_{\delta,\epsilon}$ of this history satisfies

$$\left\| m_{\delta,\epsilon} - m^* \right\|_1 \;\leq\; \frac{2E}{1 - E}, \qquad E = (n - 1)\, \|G\| \left(\left[\tbinom{n}{2} - 1 \right] \delta \sum_{k=1}^{r} \frac{\lambda_k}{\tilde{s}(t_k)} + \epsilon \right), \tag{C3}$$

whenever $E < 1$ (the unnormalized error is at most E; the factor $2/(1 - E)$ pays for renormalizing). Taking $\delta = \epsilon = 2^{-j}$ gives a sequence of genuine histories in the class of Lemma 1 whose spectra converge to m^* linearly in 2^{-j}. The floor $\epsilon > 0$ keeps each member a genuine history; the endpoint $\Gamma_n(0)$ is reached only in the limit.

The equality also settles whether letting each locus have its own history enlarges the achievable set. Pooled spectra over arbitrary finite families of histories and weights are convex combinations of single-history spectra, so the closure of what per-locus-history models can reach is exactly the closure of what single-history models can reach, namely $\mathrm{conv}(\Gamma_{20}([0, 1]))$ again. Allowing each locus its own, least favorable, demography therefore does not enlarge the set of achievable expected pooled spectra.

In the outward (exclusion) direction of the closure theorem, a proven curve-wide bound together with a proven violation margin $\mu > 0$, the form taken by each exclusion proof in this paper (the post-shift affine separating functional), implies that *every* history in the class keeps ℓ_1-distance at least $\mu / \|a\|_\infty$ from the selected spectrum. The exclusion itself already holds for the whole class by Lemma 1; the quantitative standoff over the full demography space requires one further finite computation on the functional. In the inward direction, any target with an exact convex representation, the form of every mimicry result, is a limit of *genuine* neutral spectra along an explicit family with a computable rate, so a proven mimicry result means that genuine histories approach the target spectrum arbitrarily closely, a stronger statement than membership in the sampled hull. We have carried out this further computation at thirteen points. One is the exclusion at $(S, h) = (10, 0.1)$ from the independent re-analysis (the first excluded row of Table 6 below, not the Table 5 row at the same point), with a proven ℓ_1 standoff

(S, h)	$\log_{10}$ margin	note
$(2, 0.1)$	-3.804	
$(2, 0.3)$	-4.059	
$(2, 0.7)$	-7.61	
$(2, 0.9)$	-4.918	
$(10, 0.1)$	-2.269	recomputed independently from a finer sampling of the curve; same margin
$(10, 0.3)$	-2.214	recomputed independently from a finer sampling of the curve; same margin
$(10, 0.7)$	-2.028	
$(10, 0.9)$	-1.845	
$(50, 0.1)$	-2.094	
$(50, 0.3)$	-1.926	
$(50, 0.7)$	-1.458	
$(50, 0.9)$	-1.271	
$(-30, 0.1)$	-9.471	
$(-100, 0.1)$	-9.37	box-supremum bound; second, independent separating functional

Table 5 Exclusions with dominance: the full sampled h-grid at $S \in \{2, 10, 50\}$, and recessive-deleterious selection at $S = -30$ and $S = -100$. Same conventions as Table 4. All rows of this table are additionally subject to assumption A7 of Appendix C10 (the error radii from the dominance reference evaluator are consistency estimates, and each exclusion is conditional on their validity).

of 5.3771×10^{-3} from every member of the class, subject to assumption A7 of Appendix C10 as everywhere at $h \neq \frac{1}{2}$. We ran the same computation, varying only the strength, at all twelve fixed-S additive strengths from $S = 1.375$ to $S = 100$ (the rows of Table 4 together with $S = 1.4375$ of Table 7), which gives each of the twelve exclusions a quantitative standoff from every neutral history. Proven ℓ_1 standoffs range from 8.9×10^{-13} at $S = 1.375$ to 1.3×10^{-2} at $S = 100$ (Figure 6); these rows are all at $h = \frac{1}{2}$, so assumption A7 does not apply to them. For the other proven exclusions (the fourteen dominance points at $h \neq \frac{1}{2}$ of Appendix C3 and four further points from the re-analysis and bracket refinement below) the exclusion holds for every history by Lemma 1, but no standoff has been computed. In the inward direction only the theorem is available; we have not computed the constructive approximation for any specific point. We also searched numerically for counterexamples to Eq. (10) in both directions and found none. No step density outside the realizable class (the class of densities that arise from a genuine size history) was found to leave the hull; the only step densities outside that class, those that vanish on an interior interval, are provably unrealizable and reach the hull only through the closure, consistent with the theorem. In the other direction, no excluded point below the discretization bound was found to lie inside the class.

C5. Mimicry at weak selection and the first onset bracket

On the weak-selection side the selected spectra are mimicked: at each sampled $S \in \{-10, -1, -0.1, +\frac{1}{2}, +1\}$ (with $h = \frac{1}{2}$) the selected spectrum admits an explicit exact convex representation over sampled neutral curve points, so demography can reproduce weak selection, in the sampled-hull sense of Section 2.5, within the panmictic Kingman class. The sampled mimicked and excluded strengths bracket the onset of distinguishability $S^* := \inf\{S > 0 : \text{excluded}\}$:

$$S^* \in (1,\ 1.375] \qquad \text{on the sampled grid,} \tag{C4}$$

that is, the largest sampled mimicked strength is $S = +1$, the smallest sampled excluded strength is $S = +1.375$, and the intermediate sampled point $S = 1.25$ is unresolved; nothing is inferred between grid points, by interpolation or by a monotonicity assumption. The point $S = 1.25$ is unresolved for a geometric reason, which we state in terms of three quantities used again in Appendices C6 and C7. When no convex representation of a selected spectrum exists over the sampled curve points, the exact computation returns instead a candidate affine functional that is nonnegative at every sampled curve point and negative at the selected spectrum. We scale each candidate so that its largest coefficient, the constant included, has unit magnitude; the *separation shortfall* is then the magnitude of its negative value at the selected spectrum, quoted as a positive number. Between sample points a candidate can fall below zero on the neutral curve: its *dips* are the rational points t of the curve at which we prove it negative, and its *curve-dip depth* is the most negative of these proven values. The *outward-slack shift* raises the constant w_0 by a dyadic amount exceeding the dip depth and proves the shifted functional nonnegative on the whole curve by methods B and R; the constant of Table 3 already includes such a shift. The shift yields a separating functional only if the shortfall exceeds the dip depth, since otherwise it carries the selected spectrum to the nonnegative side as well. At $S = 1.25$ the best shortfall found, 8.235×10^{-14}, is smaller than the proven dip depth, 1.835×10^{-12} (twelve dips; the quoted dip at $t = 123/128$), so none of the candidate functionals found can be shifted into a separating one. Refining the bracket below 1.375 requires additional grid points or a max-margin exact separation step. The additional grid points analyzed in Appendix C6 tighten the sampled bracket to $S^* \in (1.125,\ 1.375]$.

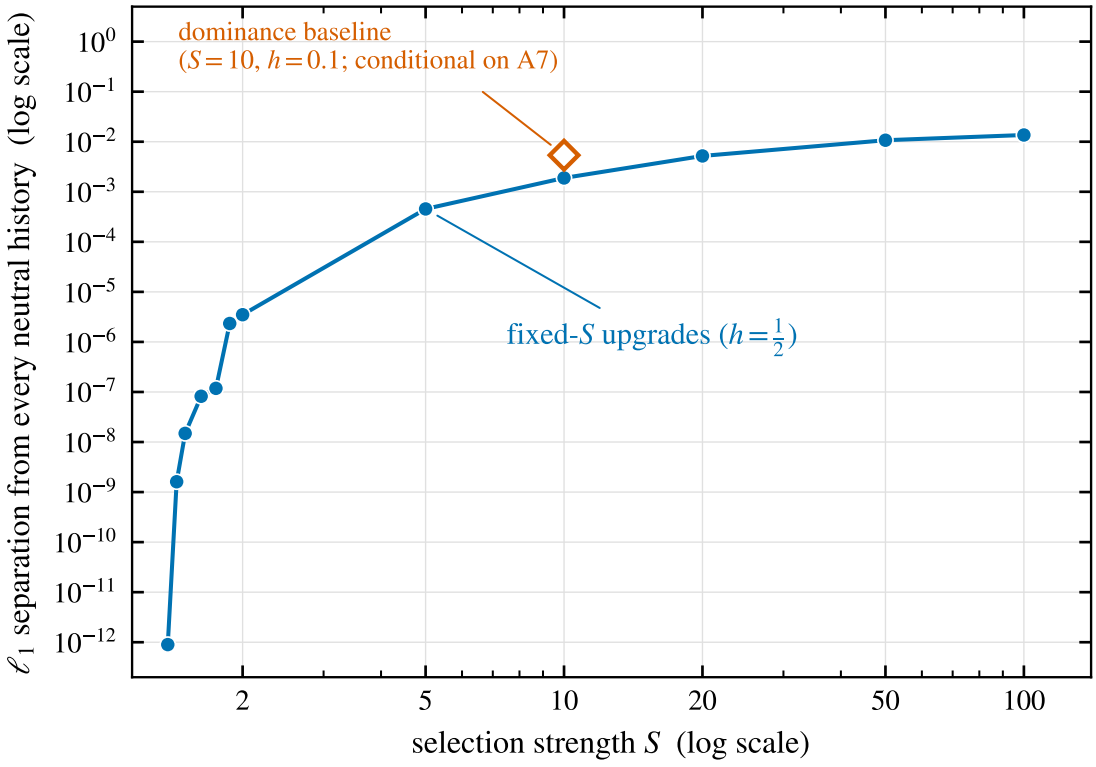

Figure 6 Proven ℓ_1 standoff against the full neutral class for the thirteen exclusions at which it has been computed (both axes log scale). The twelve fixed-S additive strengths ($h = \frac{1}{2}$, filled circles) rise from 8.9×10^{-13} at $S = 1.375$ to 1.3×10^{-2} at $S = 100$; the dominance point at $(S, h) = (10, 0.1)$ from the independent re-analysis (open diamond, standoff 5.3771×10^{-3}) is additionally subject to assumption A7 of Appendix C10, which the twelve fixed-S points are not. All values are proven lower bounds on the ℓ_1 distance over the panmictic Kingman class of Appendix C10, printed rounded down. Every neutral history in the class keeps at least this distance from the selected spectrum.

C6. Independent re-analysis and refinement of the onset bracket

We repeated the mimicry analysis independently on sixteen targets whose identities were masked during the analysis; three of the sixteen were controls with known answers, and all three were recovered correctly. Table 6 gives the result: five targets mimicked constructively (an explicit eighteen-epoch neutral step history whose exact rational ℓ_1 gap to the selected spectrum lies at least five orders of magnitude below the $2^{-20} \approx 9.537 \times 10^{-7}$ acceptance threshold), four exclusions, and four unresolved points with exact bracket gaps. The re-analysis reproduces the structure found in Appendices C3 and C5: mimicry below the onset, exclusion above it, and at strong purifying strengths the same situation as in Table 8, a separation shortfall smaller than the proven dip depth.

We then tightened the onset bracket by direct computation at four dyadic grid points (Table 7), of which only $S = 1.375$ was on the original grid: $S = 1.125$ is mimicked constructively, $S = 1.4375$ is excluded, $S = 1.375$ is excluded with an exact rational margin, and $S = 1.3125$ is unresolved between them. This gives the bracket of Eq. (9), $S^* \in (1.125, 1.375]$ on the grid points analyzed; as before, nothing is inferred between grid points. The two intermediate sampled points, $S = 1.25$ and $S = 1.3125$, both remain unresolved, inside the bracket. Both computations reached the refinement limit with dips remaining, but their final candidate functionals differ: at $S = 1.25$ the shortfall lies below the dip depth, the geometric obstruction of Appendix C5, whereas at $S = 1.3125$ the shortfall, 6.071×10^{-13}, exceeds the proven dip depth, 4.065×10^{-18} (one dip, at $t = 5951/8192$). A slack-shifted separating functional is therefore not precluded at $S = 1.3125$, and a deeper computation may resolve that point. The $(S, h) = (-100, 0.3)$ point of Table 6 was recomputed with a larger computational limit and remained unresolved for the geometric reason described in Appendix C7, with separation shortfall 1.488×10^{-19} against a proven dip of -1.381×10^{-16} at eight dips.

C7. Unresolved points and the boundary of the class

At strong purifying *additive* selection the question of mimicry remains unresolved; the gaps that leave it open can, however, be stated exactly. In the five rows of Table 8 neither a separating functional nor a convex representation was found at the resolution stated: in each row the best proven separation shortfall is smaller than the proven depth of dips of the curve functional, so no outward-slack separating functional exists among the functionals found. These spectra lie exponentially close to the inner sampled hull. Whether they are inside the achievable set (mimicry) or outside by an exponentially small margin is open; deciding it would likely require a margin-maximizing exact separation program or algebraic tangency arguments.

One mixture target also remains unresolved at $n = 20$: the expected spectrum when the strength of each site is drawn from a distribution of fitness effects, here a reflected Gamma of delete-

point ($h = \frac{1}{2}$ unless noted)	result	proven number	note
$S = -10$	mimicked	ℓ_1 gap 9.530×10^{-12}	
$S = -1$	mimicked	ℓ_1 gap 1.789×10^{-15}	
$S = -0.1$	mimicked	ℓ_1 gap 5.998×10^{-16}	
$S = +0.5$	mimicked	ℓ_1 gap 2.297×10^{-14}	
$S = +1$	mimicked	ℓ_1 gap 2.759×10^{-12}	
$S = +10$, $h = 0.1$	excluded	margin -1.123×10^{-4}	A7
$S = +10$, $h = 0.3$	excluded	margin -9.519×10^{-5}	A7
$S = +1.5$	excluded	margin -4.445×10^{-10}	
$S = +1.75$	excluded	margin -3.989×10^{-8}	
$S = +1.25$	unresolved	8.235×10^{-14} vs dip -1.835×10^{-12}	12 dips
$S = -100$	unresolved	2.588×10^{-28} vs dip -2.671×10^{-18}	11 dips
$S = -1000$	unresolved	8.084×10^{-31} vs dip -5.924×10^{-25}	9 dips
$S = -100$, $h = 0.3$	unresolved	reached the computational limit; rerun below	A7

Table 6 Independent re-analysis at $n = 20$. For "mimicked" rows an explicit eighteen-epoch neutral history exists with the stated exact rational ℓ_1 gap, each gap at least five orders of magnitude below the 2^{-20} acceptance threshold; "excluded" rows state the proven violation margin of the separating functional; "unresolved" rows state the exact separation shortfall against the proven curve-dip depth. Results are relative to the panmictic Kingman class only (Appendix C10); "A7" marks rows subject to assumption A7, as are all $h \neq \frac{1}{2}$ rows; the three control targets are omitted. The re-analysis finds mimicry at weak selection, exclusion above it, and unresolved points at strong purifying selection.

rious strengths (shape 0.35, mean $S = -800$) making up 98% of the mass with the remaining 2% beneficial at $S = +10$, the spectrum handled throughout as a rigorous enclosure. This target is unresolved for the same geometric reason as the points above.

A separate open question concerns the boundary of the class itself: linked selection could still produce mimicry through genealogies that lie outside the panmictic Kingman class, and three such cases arise. First, the exact Hudson–Kaplan structured coalescent for background selection, in which lineages are tracked through selected-background classes, is not literally a time-changed Kingman coalescent; its standard reduction to drift rescaling is the first-order approximation, and the transfer theorem of Appendix C3 applies to that reduction but not to the exact structured coalescent itself. Second, structured-coalescent hitchhiking: a sweep during the sample's genealogy makes the conditional genealogy structured even while all mergers remain binary, so it falls in neither the rescaling class nor the multiple-merger case. Third, multiple mergers themselves: strong draft produces Λ-class marginal genealogies, outside the class by construction. The transfer theorem has two further limitations. The separating functionals are coordinate-bound, so exclusion transfers only to unfolded full-spectrum coordinates at $n = 20$, and folded, truncated, or ascertained observables would require functionals constructed on the image side, which we have not done. The statements are also expectation-level at $\theta \to 0$, whereas at finite θ the normalized estimator carries a covariance-driven ratio bias, which linkage inflates and which only the $\theta \to 0$ limit removes.

We examined the first case, exact background selection, directly by simulation, and that simulation produced the only deviation from the exclusions that we observed. The deviation occurs in the zero-recombination background-selection regime of Section 3.3, matched to the most severe such regime we found in the literature ($r = 0$) and delivered, because of the simulator's fitness encoding described there, at half its nominal strength. First-order theory gives $B = e^{-10}$ at the nominal strength and a more extreme value still at the delivered strength, whereas the measured $\hat{B}$ is 0.078, so interference invalidates the first-order description in either case. In this regime the full structured-coalescent simulation places the pooled $n = 20$ neutral-window spectrum on the excluded side of the hyperplane defined by the shared separating functional. The value of that functional at the pooled spectrum is 3.62 standard errors below zero over 128 replicates (the bootstrap interval excludes zero), with magnitude 2.96×10^{-3}, comparable to the margins of the excluded selected targets. This does not contradict the transfer theorem: the exact structured coalescent lies outside the rescaling class, in the case the theorem leaves open, and the finite-θ ratio bias applies ($\theta_{\text{site}} = 10^{-3}$ in these simulations, whereas the transfer theorem is a $\theta \to 0$ statement). We interpret the deviation as an indication of how strongly full-model background selection may distort finite-θ $n = 20$ spectrum estimates in a data-poor regime (a mean of 22.4 segregating sites per window). The sign quoted above is evaluated at a point estimate of the pooled spectrum, and showing that an entire error box around that estimate lies on the excluded side would require A7-style enclosures, which we leave for future work.

In the multiple-merger case the boundary can be located exactly, and the class restriction of Appendix C10 is necessary for the exclusions to hold. On the grids we swept, the half-space of the shared separating functional is crossed by draft-class genealogies at constant population size: every Beta$(2 - \alpha, \alpha)$ coalescent with $\alpha \leq 19/10$ and every Dirac(ψ) coalescent with $\psi \geq 1/100$ has a proven negative margin, by

S (dyadic)	result	proven number	note
$9/8 = 1.125$	mimicked	ℓ_1 gap 1.397×10^{-11}	
$21/16 = 1.3125$	unresolved	6.071×10^{-13} vs dip -4.065×10^{-18}	1 dip
$11/8 = 1.375$	excluded	margin $\approx -4.70 \times 10^{-350}$	exact rational
$23/16 = 1.4375$	excluded	margin -6.060×10^{-10}	

Table 7 Refinement of the onset bracket to $S^* \in (1.125, 1.375]$. Each endpoint is a grid point at which the analysis was run, and nothing is interpolated between grid points; the interior point $S = 1.3125$ is unresolved with the exact bracket gap shown (separation shortfall against proven curve-dip depth, Appendix C5). Here the shortfall exceeds the dip depth, so the point is left unresolved by the refinement limit rather than by the geometric obstruction of Appendix C5. The $S = 1.375$ margin is an exact rational whose decimal summary is shown; it was computed independently of the Table 4 row at the same strength. Row labels as in Table 6.

exact rational signs at the grid points. Exact bisection brackets the crossing, with exactly one sign change on the grid: $\alpha^* \in (79903/40960, 99881/51200) = (1.950757, 1.950801)$, and $\psi^* \in (23/3200, 37/5120)$, so that the functional is already negative at the expected spectrum of a Dirac coalescent in which about 0.72% of the population takes part in each merger event. The expected spectrum under even mild draft therefore lies outside the half-space, and the Kingman-class exclusion cannot be widened to Λ with the existing family of separating functionals. Crossing the half-space is necessary but not sufficient to mimic the excluded targets. A numerical search for mimicking Λ genealogies, which provides no bound, suggests that draft moves the spectrum in the direction opposite to advantageous selection: multiple mergers inflate singletons (x_1 of 0.31–0.44 against Kingman's 0.28) while the twelve excluded additive targets ($h = \frac{1}{2}$, Appendix C4) deplete them ($0.25 \to 0.15$). For each of the twelve targets, the smallest ℓ_1 distance to the target attained over the Beta family and over two-atom Dirac–Kingman mixtures equals the distance from the Kingman spectrum itself to that target, so that no member of either family is closer to any target than the Kingman spectrum. All proven Λ statements here are at constant N, $n = 20$, $\theta \to 0$, in unfolded coordinates. For time-changed Λ-coalescents the question cannot be posed with the present method: the spectrum is then not a moment functional over a single rational curve, so the curve-wide separation argument has no analog, and a Λ-analog of the representation lemma would be needed. One boundary-grazing point, $\alpha = 199/100$, remains unresolved against hull membership, and no later result depends on it.

In a full forward simulation, too, draft moves a neutral window's spectrum away from the excluded targets. The draft regime of Section 3.3, the Neher–Hallatschek Bolthausen–Sznitman regime, produces a pooled $n = 20$ spectrum with singleton share 0.686, which is 2.43× Kingman's 0.282, together with that regime's characteristic non-monotonic high-frequency uptick, and the selection strength a standard single-S fit assigns to this neutral window is negative ($\hat{S} = -6.09$, the purifying direction). The displacement has the same direction as in the numerical search above, and like that search the simulation is suggestive and provides no bound. The rescaling description of background selection, then, is excluded by the transfer theorem of Appendix C3, while exact background selection and hitchhiking during the sample's genealogy remain open. Mimicry by multiple-merger genealogies is not excluded, but such genealogies appear to move the spectrum away from the targets in both exact search and forward simulation. The remaining possibility for mimicking the excluded targets is therefore a time-changed multiple-merger genealogy that reverses the singleton surplus; we found no constructive candidate of this kind, and the present method cannot test one.

C8. Local identifiability of the demography–selection–dominance map

Dominance and selection are commonly described as confounded in frequency data: an implicit equilibrium hs degeneracy is invoked without being characterized (Balick et al. 2015), and a recent biobank-scale analysis reports dominance as only weakly identifiable in practice inside one parametric fit (Sendrowski et al. 2026). The object we test is the composite demography–selection–dominance map of Section 4, in which demography acts on the selected spectrum through per-cell multipliers, the r_i nuisance structure of POLYDFE and FASTDFE. Its parameters are $(\eta, b, \bar{S}, h)$: a relative size history that is piecewise constant in the depth coordinate $t = e^{-g}$ of Eq. (8), with free epoch sizes and breakpoints η and the most recent size fixed at one (the normalized spectrum is invariant under rescaling the whole history); the shape b and mean $\bar{S} < 0$ of a Gamma distribution of scaled selection coefficients, discretized onto the twelve nodes $S_j = -2^k$, $k = -5, \ldots, 6$, with weights $f_j(b, \bar{S})$ proportional to the Gamma density of $|S|$ at $|S_j|$ times $|S_j|$; and the dominance coefficient h. Its output is the normalized expected spectrum $\Phi = v / \sum_i v_i$ over the $n - 1 = 19$ polymorphic cells at $n = 20$, where $v_i = r_i(\eta)\, m_i(b, \bar{S}, h)$, $m_i = \sum_j f_j\, \xi_i(S_j, h)$ is the equilibrium spectrum of Eq. (5) averaged over the discretized distribution, and $r_i(\eta)$ is the ratio of the normalized neutral spectrum under the history to that under constant size. The Jacobian is the $19 \times p$ matrix of derivatives of Φ with respect to the p free parameters, assembled from the exact rational derivatives of r_i, the analytic derivatives of the weights in $(b, \bar{S})$, and $\partial \xi_i / \partial h$ at the nodes; an exact confounding ridge through a point would appear as rank below p. The twelve points are the combinations of three histories, constant size ($p = 3$), tenfold growth ($N = \frac{1}{10}$ for $t < \frac{1}{2}$ and 1 thereafter; $p = 5$) and a tenfold bottleneck ($N = \frac{1}{10}$ for $\frac{1}{2} \le t < \frac{3}{4}$ and 1 otherwise; $p = 7$), with two effect distributions, Gamma with shape 0.2 and mean -100 and Gamma with shape 0.5 and mean -10, each at $h = \frac{1}{2}$ and at $h = 0.1$. At none of these twelve points does an *exact* ridge exist: when the consistency-based error boxes are replaced with two proven-ball routes (exact arithmetic in $\mathbb{Q}(e^{-S_j})$ at $h = \frac{1}{2}$; rigorous quadrature of closed-form kernels

point	resolution	shortfall	dip depth (at t)	dips
$S=-100,\ h=\frac{1}{2}$	first pass	9.79×10^{-19}	-1.30×10^{-15} (499/512)	8
$S=-1000,\ h=\frac{1}{2}$	first pass	2.99×10^{-23}	-1.19×10^{-21} (8177/8192)	9
$S=-100,\ h=\frac{1}{2}$	deep retry	2.588×10^{-28}	-2.671×10^{-18} (245/256)	11
$S=-1000,\ h=\frac{1}{2}$	deep retry	8.084×10^{-31}	-5.924×10^{-25} (1021/1024)	9
$S=-100,\ h=0.3$	first pass	2.98×10^{-15}	-3.22×10^{-14} (253/256)	8

Table 8 Unresolved points at strong purifying selection. In each row the best proven separation shortfall (column 3) is smaller than the proven curve-dip depth (column 4), which arithmetically precludes an outward-slack separating functional among the functionals found. Values are decimal summaries of exact rationals; "first pass" and "deep retry" are successive hull resolutions (for $S=-100, h=0.3$ the row gives the first pass; the deeper computation described in Appendix C6 also remained unresolved). None of these rows is assigned to either side. Results are relative to the panmictic Kingman class only (Appendix C10).

at $h=0.1$), the joint Jacobian of this map is proven full-rank at all twelve points, at each point by a minor whose rigorous interval enclosure excludes zero. Eight apparent rank deficiencies suggested by a preliminary floating-point computation are refuted by these proofs. Local identifiability holds within this class of maps; the confounding seen in practice is real but is a matter of *conditioning*, with rigorous smallest-singular-value enclosures reaching as low as $[1.13\times10^{-20},\ 4.047\times10^{-9}]$ at the worst of the twelve points (the bottleneck with the shape-0.2 distribution at $h=0.1$) and up to $[1.60\times10^{-5},\ 1.81\times10^{-4}]$ at the best (constant size, shape 0.5, $h=0.1$), and not an exact structural ridge. The tail-insensitivity direction is likewise proven nonzero, though small, at each point. Three restrictions apply to this claim: the proven object is the composite map, with demography acting multiplicatively on the selected spectrum, and not the exact time-inhomogeneous joint theory, whose treatment remains open; the family is the discretized twelve-node Gamma family and not the continuous family; and the statement is local, at twelve points at $n=20$, with no global and no other-n statement. Assumption A7 does not apply here, because these proofs replace the consistency-based boxes outright with proven balls. (As a control, the same procedure correctly rejects an exactly singular constructed Jacobian.)

C9. Invisibility of a neutral point mass and the open converse

How much strictly neutral sequence there is cannot be determined from the site-frequency spectrum, in the following exact sense: fix any $n=6$ through 10 (the underlying span condition is also proven at $n=3,4,5$) and any atom model placing mass $\pi_0\in\{\frac{1}{10},\frac{3}{10}\}$ on exactly neutral sites with the rest on a Gamma distribution of deleterious effects. Then there exist at most $n-1$ Gamma distributions and positive weights whose mixture, a smooth density with no atom anywhere, has *exactly* the same normalized expected spectrum. The identity is exact: the atom spectrum lies in the interior of the achievable cone, the Gamma-mixture cone is dense in it, and a relative-interior argument (Rockafellar 1970) converts closure membership into membership. It follows that, with the mutation-rate scale free, the two models induce likelihood families that coincide up to the reparametrization absorbing their relative scale: every maximum-likelihood fit, likelihood-ratio test, or scale-invariant procedure is unable to distinguish them on any dataset, however many sites it contains. (A Bayes factor under a fixed proper prior on the scale, or a parameter-count-penalized criterion such as BIC, is not literally invariant; the identifiability content stands as stated.) The nearest prior statements are verbal or empirical: that near neutrality "cannot be precisely defined" was posed conceptually (Akashi et al. 2012); sensitivity of inference to multimodality was measured by simulation (Kousathanas and Keightley 2013); and a point mass at neutrality has been fit and compared by likelihood on three datasets (Kim et al. 2017), with no distinguishability statement, which the theorem above supplies. We constructed explicit matching mixtures at each $n=3$ through 10, and a numerical search for counterexamples found no atom spectrum outside the mixture cone at any $n\le10$; the only targets found outside the cone had single-point support, which places them outside the theorem's hypotheses.

The converse remains open: against the *rigid* single-Gamma family we have no proven positive lower bound on the separation at any tested (n,π_0). Every completed branch-and-bound stage, across $n=6$ through 10 at $\pi_0\in\{\frac{1}{10},\frac{3}{10},0\}$ plus two smaller-π_0 cases at $n=6$, returned a proven global lower bound of exactly zero, and in each case exhausted the computing time allotted to it before reaching the parameter floor. The proven state everywhere is the bracket [0, chord ceiling], with ceilings from $\approx2.999\times10^{-3}$ ($n=6$, $\pi_0=\frac{1}{100}$) up to $\approx6.425\times10^{-2}$ ($n=6$, $\pi_0=\frac{3}{10}$), a ceiling linear in π_0. On the most favorable instance the bound remained exactly zero after a computation more than five times as long. Floating-point evidence suggests a small positive separation, but we make no rigorous separation claim, and further progress probably requires an algorithmic change in the bound-tightening step; more computing time alone appears unlikely to help.

C10. Scope and assumptions

All mimicry, exclusion, and identifiability claims in this paper are subject to the following assumptions.

- **Class.** All results are relative to the panmictic Kingman class $\mathcal{N}_n$ only; island/structured/Λ-coalescent mimicry is not excluded. The boundary this restriction draws is examined in Appendix C7. Per-locus time-changes of Kingman under unconditional pooling are excluded alongside single histories wherever a curve-wide separating functional is proven; quantitative standoffs exist at thirteen points, all at $n=20$, and nothing is excluded at any other n (Appendices C3 and C4). The Λ crossing of the separating halfspace is proven on the swept grids, and the structured conditionals, exact background selection and hitchhiking, remain open, with the simulation result on exact background

selection described in Appendix C7.

- **A1 (sampling).** "Inside" means inside the convex hull of curve *samples* only: it is exact as a statement that no separating functional exists over those samples, up to sample refinement, and it implies mimicry insofar as sampled hull points are achievable.
- **A2 (representation).** Every exclusion rests on the representation lemma (Lemma 1, proved in Appendix C1), which is a theorem about the Kingman coalescent with variable population size and carries no further assumption within that class.
- **A4 (transcendence).** Sign tests on quantities rational in e^{-S} terminate by Lindemann's theorem.
- **A5 (exact arithmetic).** Floating-point computation is used only to propose candidates; every result stated is established in exact rational or rigorous interval arithmetic.
- **A7 (dominance error radii).** For all dominance ($h \neq \frac{1}{2}$) results: the error radii are consistency estimates from the dominance reference evaluator of Section 2.4 rather than proven balls; each result covers the entire box defined by those radii and is conditional on their being valid upper bounds on the error.

D. Software comparison: details

This appendix collects the per-configuration measurements behind Section 5, for the software versions named there. For FIT*dadi* we include both the 2017 rule and the current rule for DFE mass beyond the cached grid.

POLYDFE. The near-neutral window in which the GSL Kummer routine loses all significant figures is $|S| \in [1.0, 5.0] \times 10^{-5}$ in complete runs of the program ($[1.08, 5.0] \times 10^{-5}$ on the fixed-S grid). At even $n = 100$ the distributed binary exits during initialization on the refit data of Section 5.2, with the same signature as GitHub issue #3, and POLYDFE's bundled simulator does not generate an $n = 100$ dataset, so the even-n configuration could not be run in full with the distributed tools. A build modified to pass through the window converges, on the $n = 100$ data with truth $(b, S_d, p_b, S_b) = (0.4, -400, 0.05, 40)$, to $b = 0.99$, $S_d = -46$; the fitted shape b is then 83 times the 95% interval from truth. At odd $n = 999$ the deleterious quadrature evaluates double-precision residues of the wrong sign ($\sim -10^{-288}$) at deep-deleterious nodes, 11,406 evaluations per likelihood call, and the distributed binary again exits. At $n = 101$ with the same truth, the individual values fitted with the distributed numerics are $b = 0.455$ (7.8× the 95% confidence interval at genome scale), $S_d = -251$ (2.7×), $p_b = 0.0685$ (8.0×), and $S_b = 18.5$ (11.8×); α_{dfe}, the fraction of relative fixations attributable to the beneficial component, computed from the fitted DFE through the fixation kernel $S/(1 - e^{-S})$ as POLYDFE defines it, fits at 0.9499 against a true 0.9642 (15.6×). In the S_b sweep at fixed (b, S_d, p_b) (Figure 7) the beneficial mass above the threshold, $p_b\, e^{-38/S_b}$, is 5×10^{-8} at $S_b = 2.84$, $p_b = 0.031$, where all parameter differences are below 10^{-3} of a confidence interval at genome scale; the largest difference is 0.085 of an interval at $S_b = 5$ and 2.3 intervals at $S_b = 10$, and at $S_b = 20$ the effect reaches exome-scale data sizes. The boundary corresponds to $p_b\, e^{-38/S_b}$ crossing $\sim 10^{-3}$ (genome scale) or $\sim 7 \times 10^{-3}$ (exome scale).

FIT*dadi* and *dadi*. On the distributed example's DFE the 2017 rule's fitted tail mass $P(|S| > 4000)$ is 1.6 standard errors from truth at the example's own data size. On a steeper DFE (mean $|S| = 10^4$) the fitted scale under the 2017 rule runs to the optimizer's box bound, and the current rule underestimates the scale by 92%. The −11% profiled scale shift of the current rule on the example is specific to the example's $|\gamma| \leq 500$ cache grid, which leaves 7.5% of the DFE mass beyond its edge; the default $|\gamma| \leq 2000$ grid leaves far less. On the real example data, with the mutation rate fixed and a single realization, the same disagreement appears as an 8.6% scale difference between the two rules, 0.6σ at the example's size; of the gap between this split and the profiled expected-data split, 94% is attributable to the estimand convention, 4.6% to demography and 1.5% to the realization. The diffusion-solver comparisons are made at demographic equilibrium because that is the setting in which the exact answer is independently known; they isolate the solver and extrapolation layer under conditions favorable to it, and runs with demography solve the same discretized operator and inherit at least this error.

FASTDFE. At the default 2×1000-bin kernel a fitted-parameter difference of one standard error would require $L \sim 10^{11}$–10^{12}. At 100 bins per sign the $|S_d|$ inflation is 2.4σ at $L = 10^8$ and 7.7σ at 10^9; the error scales quadratically with the bin width under bin halving, as expected, and the shape–scale ridge (correlation ≈ -0.9) amplifies the kernel error by a factor of about seven into the scale parameter. Arbitrary-precision software run at a fixed precision bounds per-operation roundoff but does not bound the accumulated error of a tabulation (Section 2.2); in the terms of Appendix B, FASTDFE's use of `mpmath` at 15 significant figures falls under library trust.

Dominance evaluation. FASTDFE's dominance evaluation uses 1000×200 trapezoid nodes in double precision. Its largest measured errors, 1.6×10^{-3} at $(S, h) = (-100, 0.2)$ and 2.3×10^{-2} at $(-1000, 0.2)$, are up to a thousand times its semidominant kernel error (Figure 8). The 21-point linear h-interpolation adds a further layer that we have not measured. An exact dominance kernel in the style of Section 2.3 would remove both, and a full dominance refit is the natural next step.

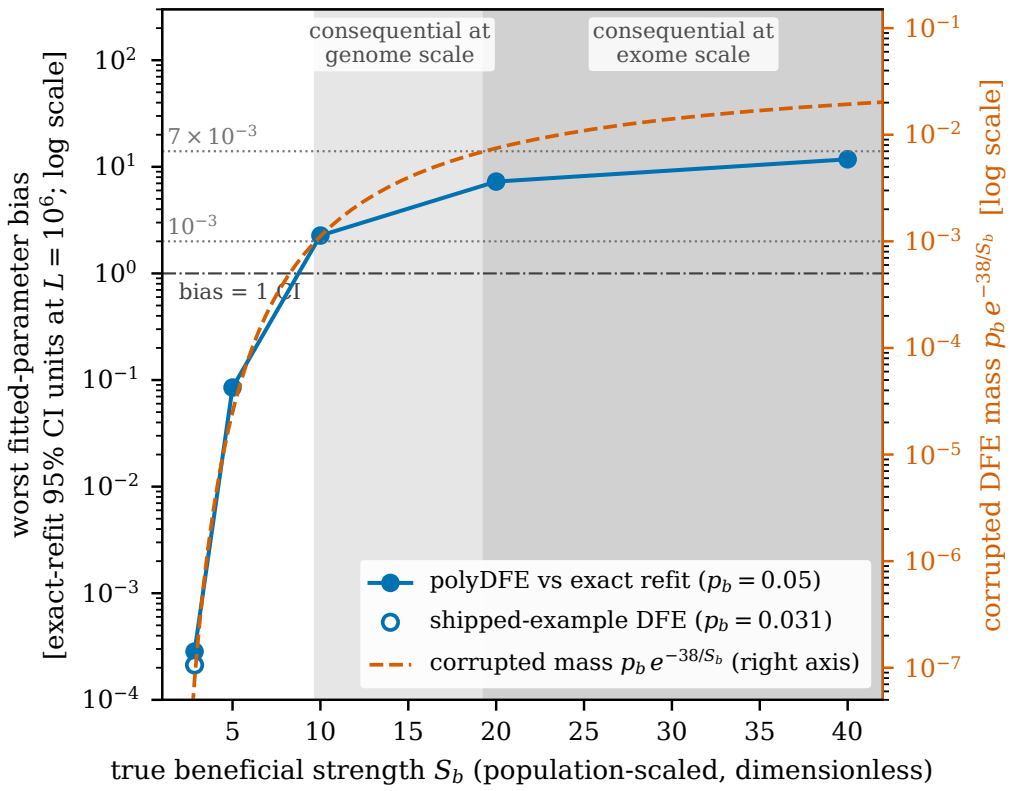

Figure 7 Largest fitted-parameter difference between POLYDFE's numerics and the exact likelihood, in units of the exact-likelihood 95% confidence half-width, versus the true S_b at fixed $b = 0.4$, $S_d = -400$, $p_b = 0.05$, measured at $S_b = 2.84$, 5, 10, 20, and 40 ($n = 101$). The overlaid boundary is the beneficial mass above the threshold, $p_b\, e^{-38/S_b}$, crossing $\sim 10^{-3}$ (genome scale) and $\sim 7 \times 10^{-3}$ (exome scale). Fits with modest beneficial strength ($S_b \lesssim 5$) are unaffected; fits at $S_b \gtrsim 10$ at this n are affected.

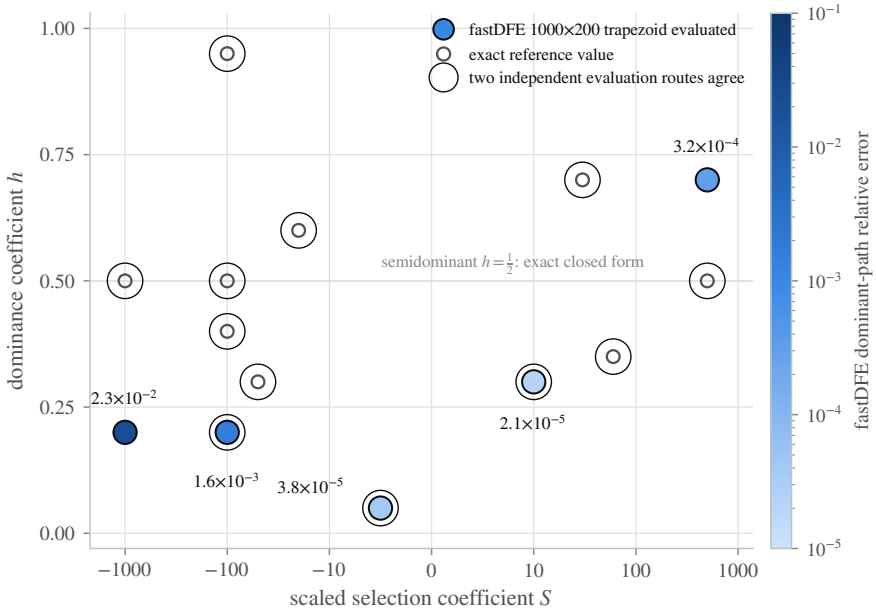

Figure 8 Relative error of FASTDFE's nested-trapezoid dominance evaluation (double precision) against the reference values of Eq. (5) (Section 2.4) at each measured (S, h) point ($n = 20$; marker color). For ringed markers the reference value was confirmed by a second route (the series evaluator, or the exact $h = \frac{1}{2}$ closed form); for unringed markers it rests on the agreement of the reference evaluator's two quadrature degrees (Appendix B). Errors reach 2.3% at strong purifying selection with low dominance, about a thousand times the tool's semidominant kernel error. These are spectrum-entry errors; their effect on fitted parameters has not been measured (Section 5.2).