1 Data likelihood

The full data $\mathcal{D}$ consists of the SFS for the n sampled sequences and, possibly, divergence counts, for m and m^d fragments, respectively, of given lengths,

$$\mathcal{D} = \{p_z^j(i), l_z^j : 1 \leq i < n, 1 \leq j \leq m, z \in \{\text{neut}, \text{sel}\}\} \cup \{d_z^j, l_z^{d,j} : 1 \leq j \leq m^d, z \in \{\text{neut}, \text{sel}\}\}$$

Due to the assumed independence of sites, the log-likelihood of the data is a sum of log-likelihoods of counts (from either SFS or divergence) over the fragments and type of counts,

$$\ell(\mathcal{D}) = \sum_{z \in \{\text{neut,sel}\}} \sum_{j=1}^{m} \sum_{i=1}^{n-1} \ell\left(p_z^j(i), l_z^j\right) + \sum_{z \in \{\text{neut,sel}\}} \sum_{j=1}^{m^d} \ell\left(d_z^j, l_z^{d,j}\right),$$

where

$$\begin{aligned} \ell(p_z^j(i), l_z^j) &= \log NB\left(p_z^j(i); l_z^j \mathbb{E}\left[\, P_z(i) \mid \bar{\theta}, r_i, \epsilon, \phi \,\right], a\right), \\ \ell(d_z^j, l_z^{d,j}) &= \log NB\left(d_z^j; l_z^{d,j} \mathbb{E}\left[\, D_z \mid \lambda, \bar{\theta}, r_n, \phi \,\right], a\right). \end{aligned}$$

$\mathbb{E}\left[\, P_z(i) \mid \bar{\theta}, r_i, \epsilon, \phi \,\right]$ and $\mathbb{E}\left[\, D_z \mid \lambda, \bar{\theta}, r_n, \phi \,\right]$ are obtained using equations (1) – (6) from the main text. $NB\,(x; m, a)$ is the density of the negative binomial distribution with mean m and shape a, given by

$$NB\,(x; m, a) = \frac{1}{x\text{Beta}(x, a)} \left(\frac{a}{a+m}\right)^a \left(\frac{m}{a+m}\right)^x,$$

where Beta is the beta function (Abramowitz and Stegun, 1972). The negative binomial distribution converges to the Poisson distribution as $a \to \infty$. If variability in mutation rates between fragments is not modeled, then the likelihood is calculated using the Poisson distribution instead.

2 Implementation

The model was implemented in `C`, with the use of the numerical library `GSL` (Galassi *et al.*, 2009) for numerical integration and optimization. The integrals from $\mathbb{E}\left[\, P_{\text{sel}}(i) \mid \theta, \phi \,\right]$ and $\mathbb{E}\left[\, D_{\text{sel}} \mid \lambda, \theta, r_n, \phi \,\right]$ (equations (1), (2), (5) and (6)) are either evaluated numerically, or as a hypergometric function, as given below.

Let ${}_1F_1(i, n; z)$ be the confluent hypergeometric function (Abramowitz and Stegun, 1972) given by, for $0 < i < n$,

$${}_1F_1(i, n; z) = \frac{(n-1)!}{(i-1)!(n-i-1)!} \int_0^1 u^{i-1} (1-u)^{n-i-1} e^{zu}\, du$$

Given a fixed selection coefficient S, and the fragment specific mutation rate θ, the expectation of $P_{\text{sel}}(i)$, $1 \leq i < n$, is given by

$$\mathbb{E}\left[\, P_{\text{sel}}(i) \mid \theta, S \,\right] = \theta \int_0^1 \binom{n}{i} x^i (1-x)^{n-i} \frac{1 - e^{-S(1-x)}}{x(1-x)\left(1 - e^{-S}\right)}\, dx.$$

The integral can be rewritten as

$$\begin{aligned} \int_0^1 \binom{n}{i} x^i (1-x)^{n-i} \frac{1 - e^{-S(1-x)}}{x(1-x)\left(1 - e^{-S}\right)}\, \mathrm{d}x &= \frac{n!}{i!(n-i)!} \frac{1}{1 - e^{-S}} \int_0^1 x^{i-1} (1-x)^{n-i-1} \left(1 - e^{-S}\, e^{Sx}\right)\, \mathrm{d}x \\ &= \frac{n!}{i!(n-i)!} \frac{1}{1 - e^{-S}} \left[\int_0^1 x^{i-1} (1-x)^{n-i-1}\, \mathrm{d}x \right. \\ &\qquad \left. - e^{-S} \int_0^1 x^{i-1} (1-x)^{n-i-1} e^{Sx}\, \mathrm{d}x \right] \end{aligned}$$

$$= \frac{n!}{i!(n-i)!}\,\frac{1}{1-e^{-S}}\left[\frac{(i-1)!(n-i-1)!}{(n-1)!}\right.$$
$$\left. - e^{-S}\,{}_1F_1(i,n;S)\,\frac{(i-1)!(n-i-1)!}{(n-1)!}\right]$$
$$= \frac{n}{i(n-i)}\,\frac{1}{1-e^{-S}}\,\left[1-e^{-S}\,{}_1F_1(i,n;S)\right],$$

from which we obtain the expectation of $P_{\text{sel}}(i)$ in terms of the confluent hypergeometric function

$$\mathbb{E}\left[\,P_{\text{sel}}(i)\mid\theta,S\,\right] = \theta\,\frac{n}{i(n-i)}\,\frac{1-e^{-S}\,{}_1F_1(i,n;S)}{1-e^{-S}}.$$

The parameters ($\bar{\theta}, a, r_i$ for $1 < i \leq n, \epsilon, \lambda$ and the parameters of the DFE ϕ) are estimated by numerical maximization of the likelihood, using the BFGS method (Fletcher, 2013), with first and second derivatives calculated numerically.

3 Simulation setups

We simulated SFS and divergence counts data using SFS_CODE (Hernandez, 2008). We simulated two populations that split 5 time units ago and subsequently evolved with constant population sizes. We sampled one sequence from one of the populations (the outgroup) and 20 sequences (10 diploid individuals) from the other population (the ingroup). For a full data set, we simulated 10.8MB sites, consisted of of 1000 independent equally long fragments. Each fragment had a constant mutation rate, but the mutation rates between fragments varied according to a gamma distribution with mean $\bar{\theta} = 0.001$ and shape $a = 2$. The fragments where further divided into smaller independent (freely recombining) segments, while the scaled recombination rate $\rho = 4N_e r$ for each segment was set to $\rho = 2e-5$. The segments contained 216 sites for most simulations, but we also simulated segments of 54, 108, 432 and 1080 sites, respectively, in order to vary the amount of linkage present in the data. The low, medium and high linkage scenarios from the main text correspond to using 54, 216 and 1080 sites.

Each fragment was considered to be coding and consisted of 3600 codons. Synonymous mutations evolved neutrally, while non-synonymous mutations where assigned a selection coefficient drawn from a DFE (Table 1 from main text). We used a mixture DFE composed of a reflected gamma distribution for the deleterious mutations, with mean S_d and shape b, and an exponential distribution for the beneficial mutations, with mean S_b. A non-synonymous mutation had $S > 0$ with probability p_b.

Each fragment was obtained from a separate SFS_CODE run. The command line for one such fragment with MAMSD DFE (Table 1 from main text) and segments of 216 sites is

```
./sfs_code 2 5 --sampSize 10 -N 500 --theta 0.001 -a C \
   --selDistType 2 0.02 1 0.5 0.4 0.002 \
   --rho 2e-5 --length 10 216 --linkage p -1 -TS 0.0 0 1 -TE 5
```

From the SFS_CODE output, we calculated the SFS and divergence counts for both synonymous and non-synonymous mutations. The divergence counts included the misattributed polymorphism described in the main text. We additionally calculated the realized α for each simulation by dividing the number of fixed non-synonymous mutations by the total number of fixed mutations (excluding the misattributed polymorphism).

The simulated DFEs (Table 1 from main text) where constructed such that they covered three different values of α:

- low (LA), approx. 20%;
- medium (MA), approx. 50%;

- high (HA), approx. 80%.

For each such DFE, we also simulated data with the same S_d and b, but with no beneficial mutations. For example, LAHSD and DelHSD (Table 1 from main text) share the DFE for deleterious mutations. We additionally simulated one DFE where S_b was set to 800 (HSB), such that only approximately 10% of the beneficial mutations have a fitness of less than $S = 100$.

From the simulated SFS $p(i)$, we added misidentification of the ancestral state to obtain $p_{\text{an}}(i)$, by first simulating the number of sites with misidentified ancestral states and then updating the SFS,

$$p_1 \sim \text{Bin}(p(i), \epsilon), \qquad p_2 \sim \text{Bin}(p(n-i), \epsilon),$$
$$p_{\text{an}}(i) = p(i) - p_1 + p_2, \qquad p_{\text{an}}(n-i) = p(n-i) + p_1 - p_2.$$

When adding misidentification of the ancestral state, we either used one ϵ for all sites, or, alternatively, we used two different ϵ_{neut} and ϵ_{sel} for the sites that have been evolving neutrally or under selection, respectively.

Using the MAMSD and DelMSD DFEs (Table 1 from main text), we simulated four different demography scenarios:

- CONST: constant populations size;
- GROW: the populations doubles in size;
- SHRINK: the populations halves in size;
- EXP: the population size starts growing exponentially to a present size that is double of the ancestral.

All population size changes happen after 2.5 time units since the divergence with the outgroup. The corresponding SFS_CODE demography commands, for the non-constant population sizes, are

```
GROW:   -TN 2.5 P 0 1000
SHRINK: -TN 2.5 P 0  250
EXP:    -Tg 2.5 P 0 0.2773
```

For all simulation set-ups considered, we simulated a total 100 data sets.

4 Inference on simulated data

To ensure that the global optimum was found when estimating the parameters, we performed 10 runs of the BFGS algorithm, with randomly chosen starting values. We also provided the simulated values of the parameters as an additional starting point, to ensure that a failure in finding the simulated values as the true optimum was not caused by a failure of finding the global optimum.

Unless otherwise specified, the DFE inference reported was performed on both polymorphism and divergence data and assuming a full DFE. If only a deleterious DFE is inferred, then divergence data is not used. The distortion parameters r were estimated, while the ancestral misidentification error ϵ was fixed to 0. The mutation rates distribution parameterized by the mean $\bar{\theta}$ and shape a was always estimated. The divergence parameter λ was only estimated when the divergence data was used.

We note here that the shape a was always estimated accurately (data not shown). The estimated $\bar{\theta}$ is not a direct estimate of the mean mutation rate, but a mutation rate scaled by an unknown factor that corrects for any distortions, such as demography, acting on singletons, as r_1 is set to 1 and it is not estimated. The divergence parameter λ can also be written as $T\theta$ (Figure 1 in the main text). The true simulated λ is unknown, as we only have access to the speciation time T_s (simulated to be $5 \cdot 2N_a$, where N_a is the size of the ancestral population), and $T_s < T \leq 2T_s$. The true distortions parameters r_i are not known either, and we therefore focus our inference on the DFE and α. We also consider the estimated ϵ, when this is relevant.

5 Running dfe-alpha

We estimated the deleterious DFE and α_{div} using dfe-alpha (Keightley and Eyre-Walker, 2007; Eyre-Walker and Keightley, 2009; Schneider *et al.*, 2011; Keightley and Eyre-Walker, 2012) as follows

- We first ran est_dfe on the neutral folded SFS to estimate the demography parameters (we used the two epochs model).
- We then ran est_dfe on the selected folded SFS to estimate the deleterious DFE: s_d (not scaled by N_e) and b. The optimization was initialized from the simulated s_d and b. The unscaled simulated s_d was set to $s_d = S_d/(2N_e)$, where S_d is the scaled simulated S_d and N_e is estimated from the demography model from the previous step. In our model, $S_d = 4N_e s$, but to account for the differences in the selection models, we used $2N_e$ instead of $4N_e$. Here, we assume that the heterozygote has fitness of $1 + s$, while dfe-alpha assumes a fitness of $1 + s/2$.
- We finally ran est_alpha_omega, using the divergence counts and the previously estimated deleterious DFE, to estimate α_{div}.

For the comparison with dfe-alpha, our method was also initialized using the simulated parameters, without the additional runs described before.

We would like to note here that dfe-alpha allows for a more complex demographic model, where three epochs of different population sizes are used. As most of our simulated data was performed for a population of constant size and, as such, demography has not constituted a major focus in our analysis, and for computational reasons, we did not not use the three epochs model of dfe-alpha.

dfe-alpha can also infer a full DFE, but as discussed in the text, we have chosen to investigate the inference for a deleterious DFE only.

6 Evaluating estimation accuracy

To enable the comparison of estimation accuracy of the parameters across the different simulated scenarios, we report

$$\log_2\left(\frac{\text{estim}}{\text{sim}}\right),$$

where estim is the estimated value, while sim is the simulated value.

If sim = 0 or estim = 0, the log ratio is not defined. In this case we used instead

$$\log_2\left(\frac{\text{estim}+1}{\text{sim}+1}\right).$$

Almost always sim and estim have the same sign. However, α_{div} can be estimated to negative values. In this case, the log ratio is not defined, and if sim and estim have different signs, we use instead

$$-\log_2\left(\left|\frac{\text{estim}}{\text{sim}}\right|\right),$$

As the range of the log ratio varies over several orders of magnitudes, to allow for an easier visual comparison for both log ratios that are far away from 0, but also close to 0, we used a log-modulus scale in the figures. The log-modulus transformation (John and Draper, 1980) is given by

$$L(x) = \text{sign}(x)\log_{10}(|x| + 1).$$

This transformation preserves 0 and is symmetric around 0, $L(x) = -L(-x)$. For the figures showing p-values obtained from a LRT, we use a log scale instead.

References

Abramowitz, M. and I. Stegun, 1972 *Handbook of mathematical functions with formulas, graphs and mathematical tables (10th Ed.)*. National Bureau of Standard.

Eyre-Walker, A. and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Molecular biology and evolution **26**: 2097–2108.

Fletcher, R., 2013 *Practical methods of optimization*. John Wiley & Sons.

Galassi, M. *et al.*, 2009 *GNU scientific library reference manual (3rd Ed.)*. Network Theory Ltd.

Hernandez, R. D., 2008 A flexible forward simulator for populations subject to selection and demography. Bioinformatics **24**: 2786–2787.

John, J. and N. Draper, 1980 An alternative family of transformations. Applied Statistics pp. 190–197.

Keightley, P. D. and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics **177**: 2251–2261.

Keightley, P. D. and A. Eyre-Walker, 2012 Estimating the rate of adaptive molecular evolution when the evolutionary divergence between species is small. Journal of molecular evolution **74**: 61–68.

Schneider, A., B. Charlesworth, A. Eyre-Walker, and P. D. Keightley, 2011 A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics **189**: 1427–1437.

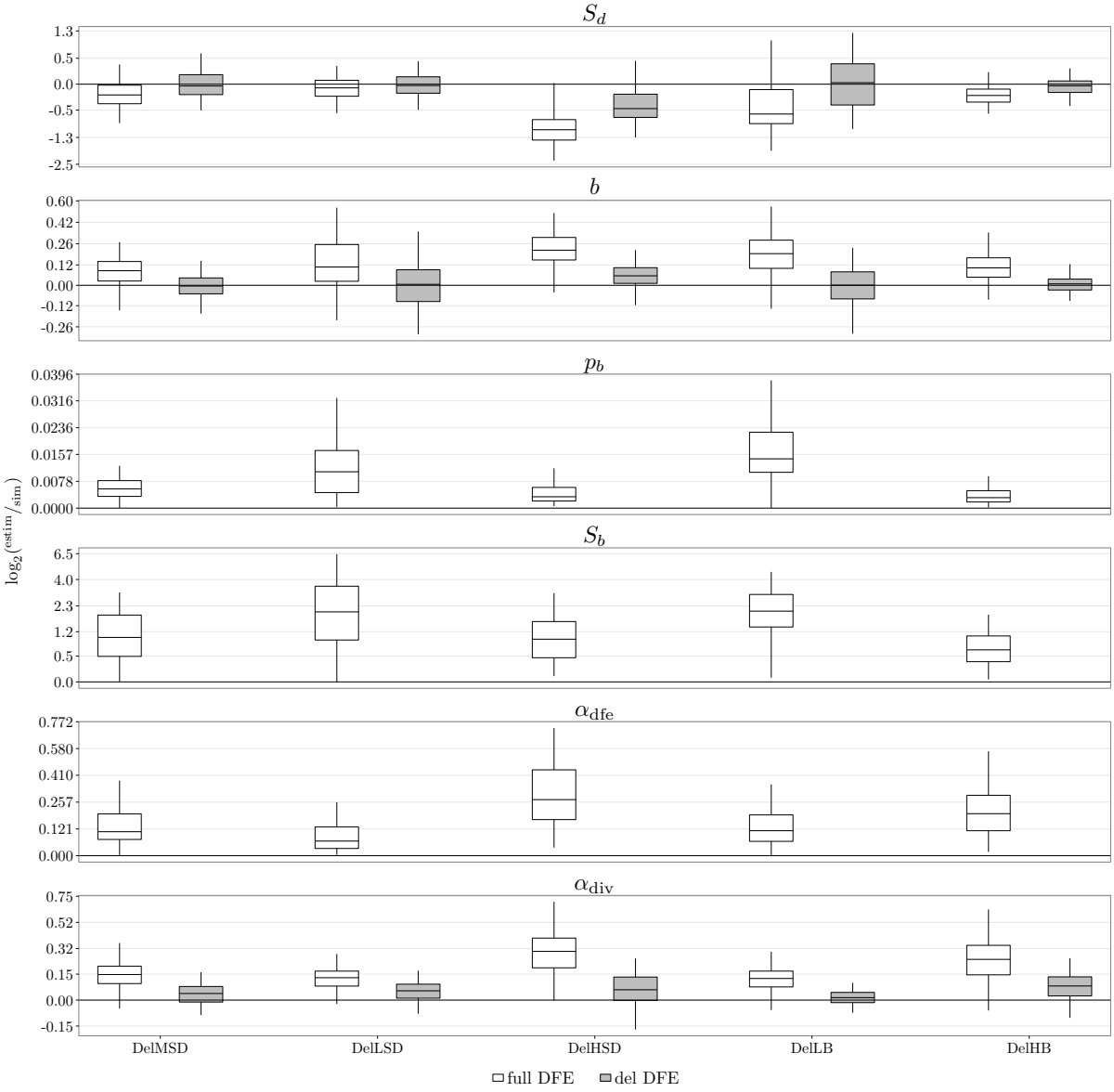

Figure S1: Box plot of inference quality on data simulated under different deleterious DFEs, given on the x-axis and detailed in Table 1 from main text. The DFE parameters are inferred using only polymorphism data.

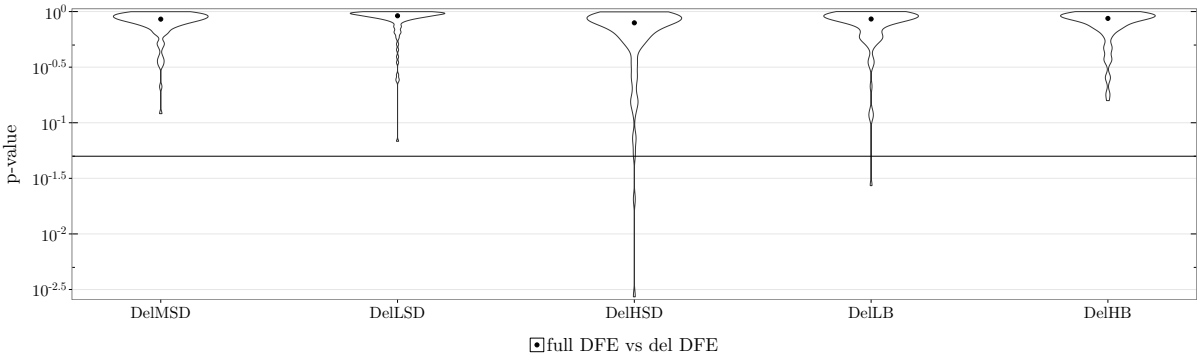

Figure S2: Violin plot of p-values for LRTs for evidence of beneficial mutations in the polymorphism data on data simulated under different deleterious DFEs, given on the x-axis and detailed in Table 1 from main text. The LRT is performed using the maximum likelihoods found when inferring a full or deleterious DFE. For low p-values, the model where a full DFE is inferred is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

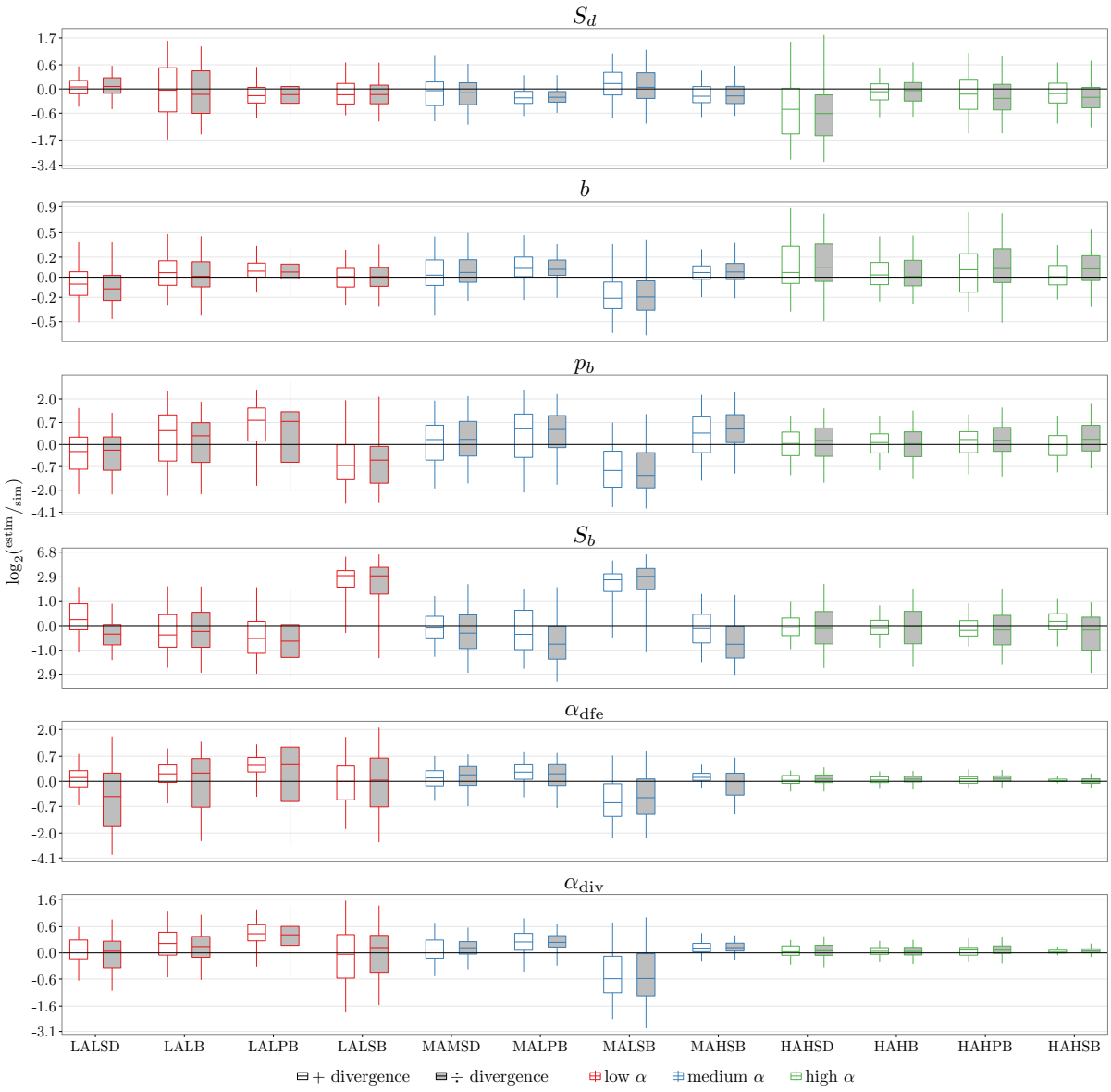

Figure S3: Box plot of inference quality on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text. The divergence data is either used (+ divergence), or not (÷ divergence).

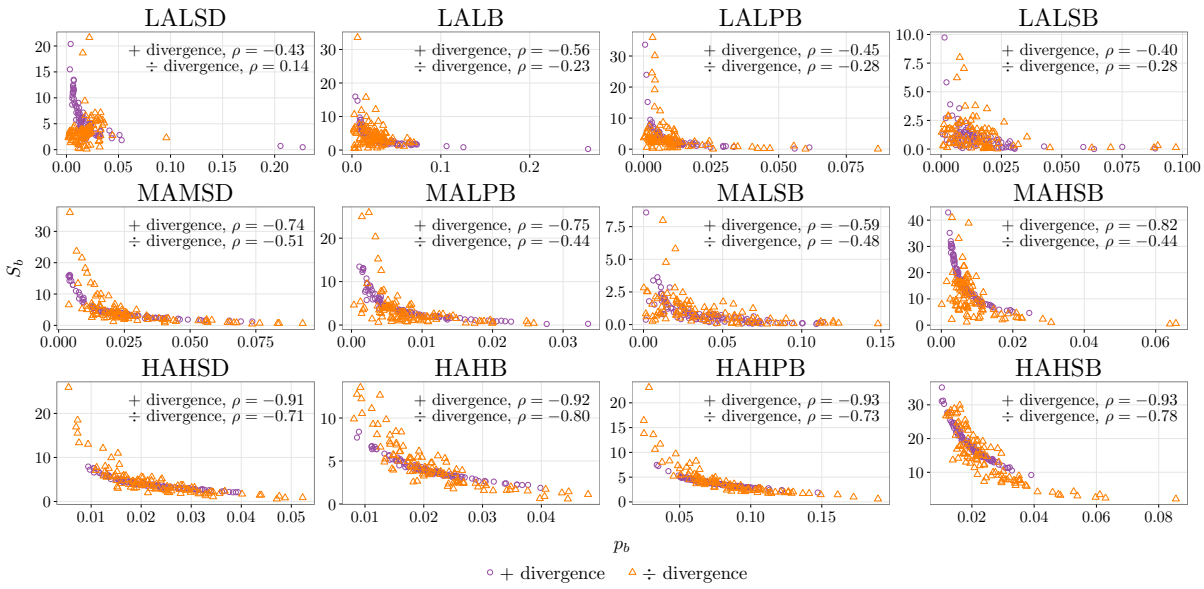

Figure S4: Negative correlation of estimated p_b and S_b from data simulated under different full DFEs, detailed in Table 1 from main text. Here, ρ is Pearson's correlation coefficient. The divergence data is either used (+ divergence), or not (÷ divergence).

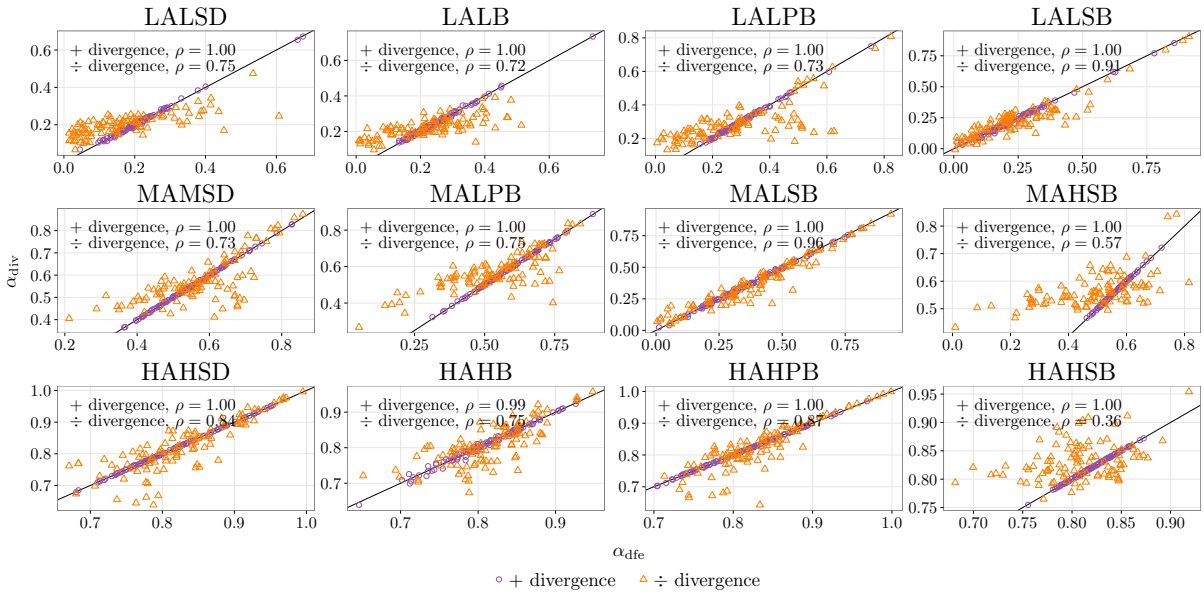

Figure S5: Correlation of estimated α_{dfe} and α_{div} from data simulated under different full DFEs, detailed in Table 1 from main text. Here, ρ is Pearson's correlation coefficient. The identity line is given in black. The divergence data is either used (+ divergence), or not (÷ divergence).

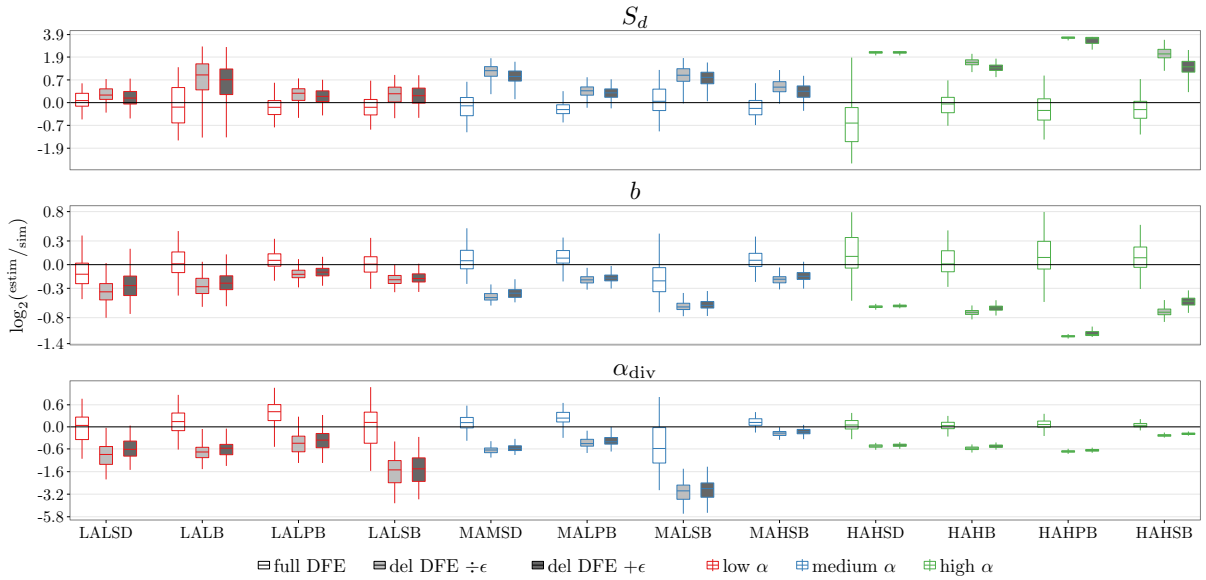

Figure S6: Box plot of inference quality on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text. The DFE parameters are inferred using only polymorphism data. When a deleterious DFE was inferred, ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$).

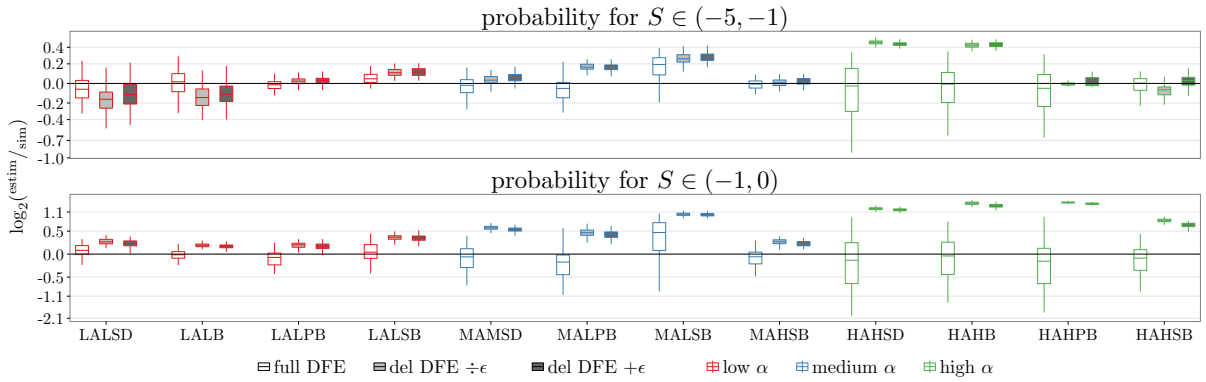

Figure S7: Box plot of inference quality on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text. The DFE parameters are inferred using only polymorphism data. When a deleterious DFE was inferred, ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$).

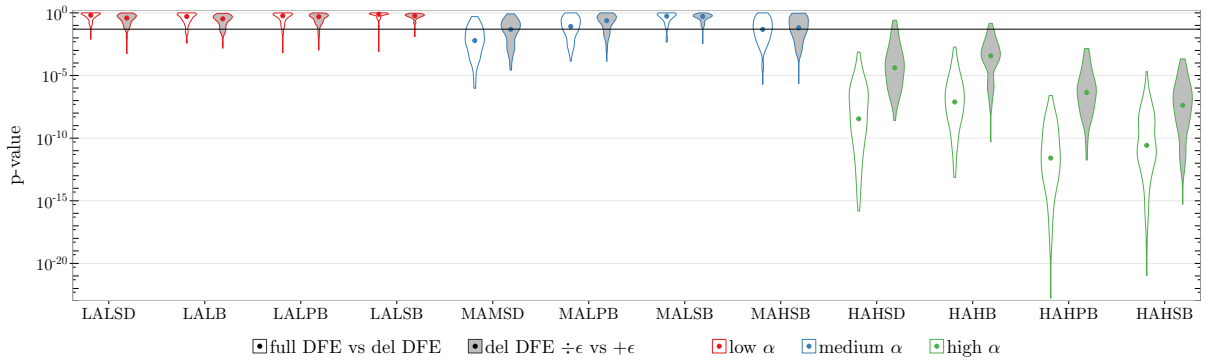

Figure S8: Violin plot of p-values for LRTs for evidence of beneficial mutations and ϵ in the polymorphism data on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text. The LRT is performed using the maximum likelihoods found when inferring a: full or deleterious DFE (white); or a deleterious DFE where ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$) (gray). For low p-values, the model where: a full DFE is inferred (white); or $\epsilon \neq 0$ (gray), is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

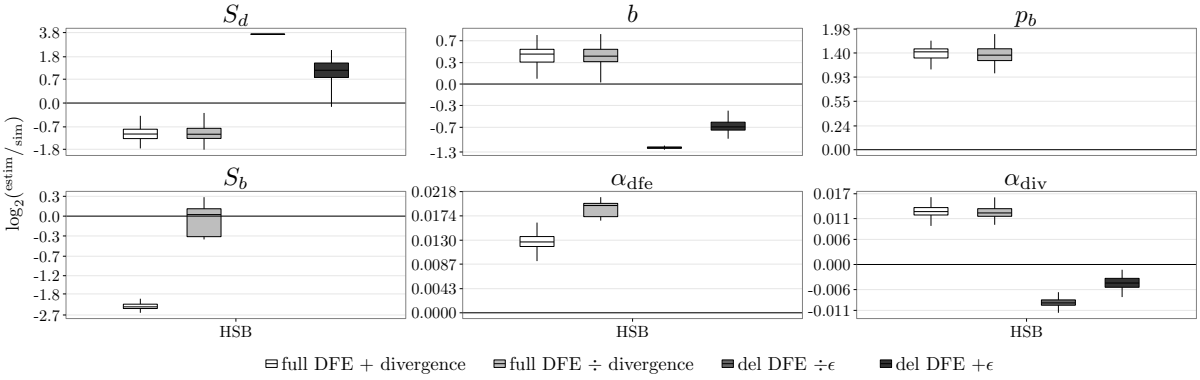

Figure S9: Box plot of inference quality on data simulated under HSB DFE, detailed in Table 1 from main text. When inferring a full DFE, the divergence data is either used (+ divergence), or not (÷ divergence), while when a deleterious DFE was inferred, ϵ was either estimated (+ϵ) or fixed to 0 (÷ϵ).

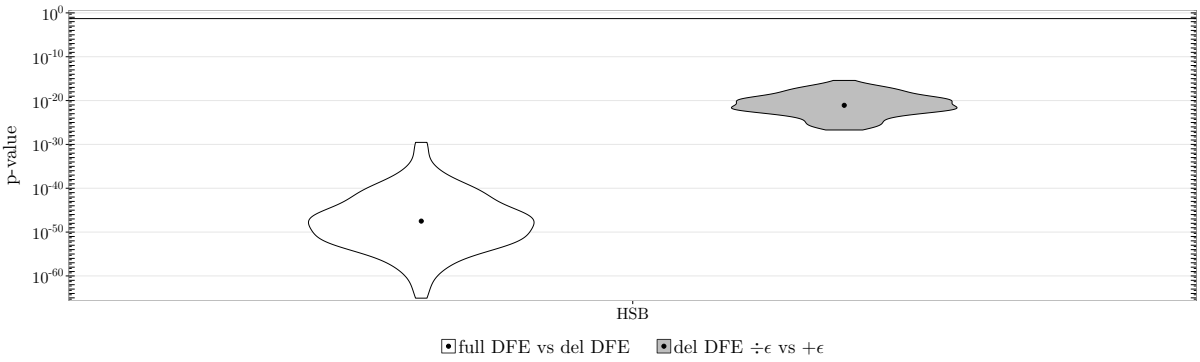

Figure S10: Violin plot of p-values for LRTs for evidence of beneficial mutations and ϵ in the polymorphism data on data simulated under HSB DFEs, detailed in Table 1 from main text. The LRT is performed using the maximum likelihoods found when inferring a: full or deleterious DFE (white); or a deleterious DFE where ϵ was either estimated (+ϵ) or fixed to 0 (÷ϵ) (gray). For low p-values, the model where: a full DFE is inferred (white); or $\epsilon \neq 0$ (gray), is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

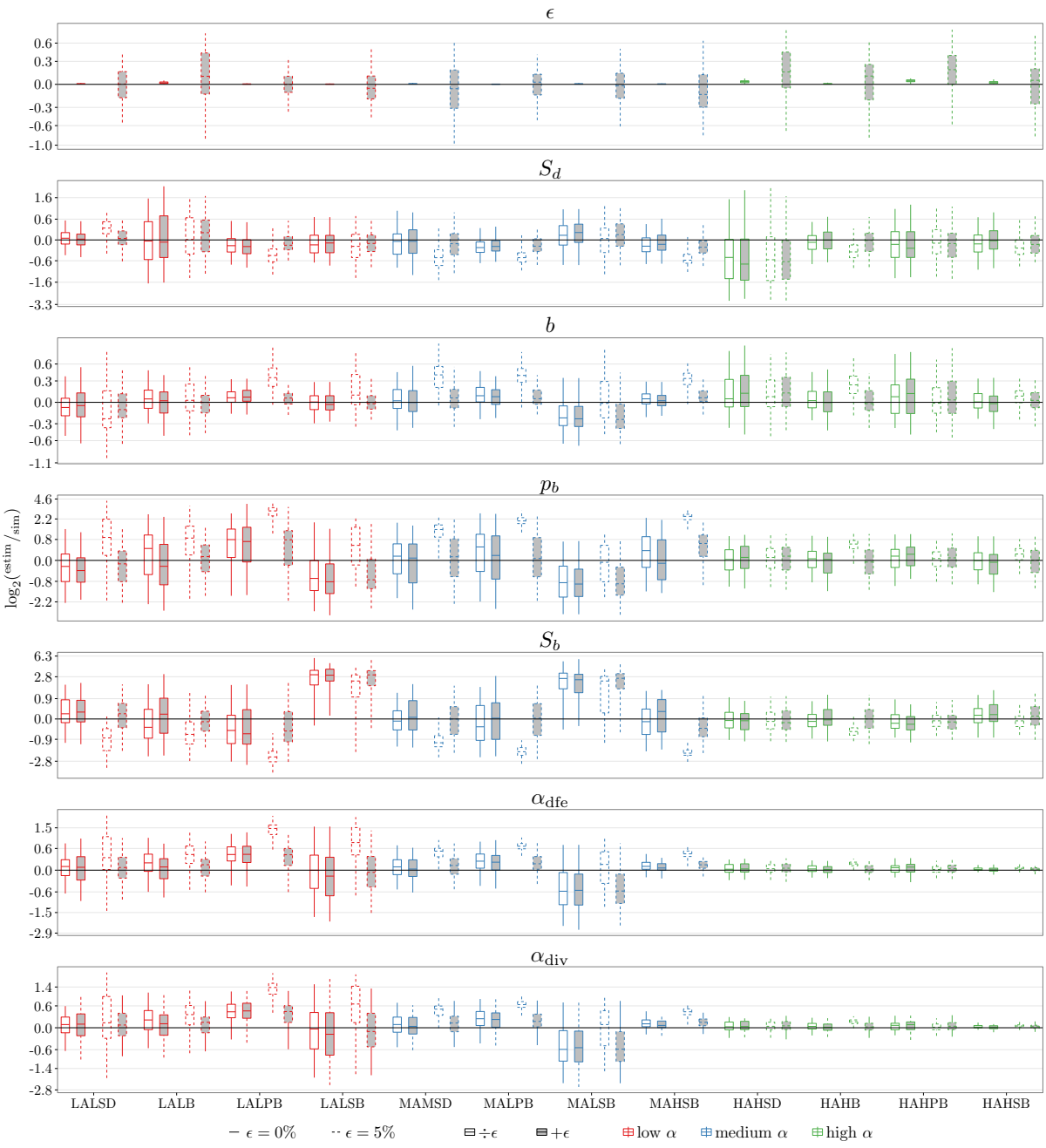

Figure S11: Box plot of inference quality on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text, where divergence data was used. Misidentification of ancestral state was added with $\epsilon = 0.05$. The ancestral error parameter ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$).

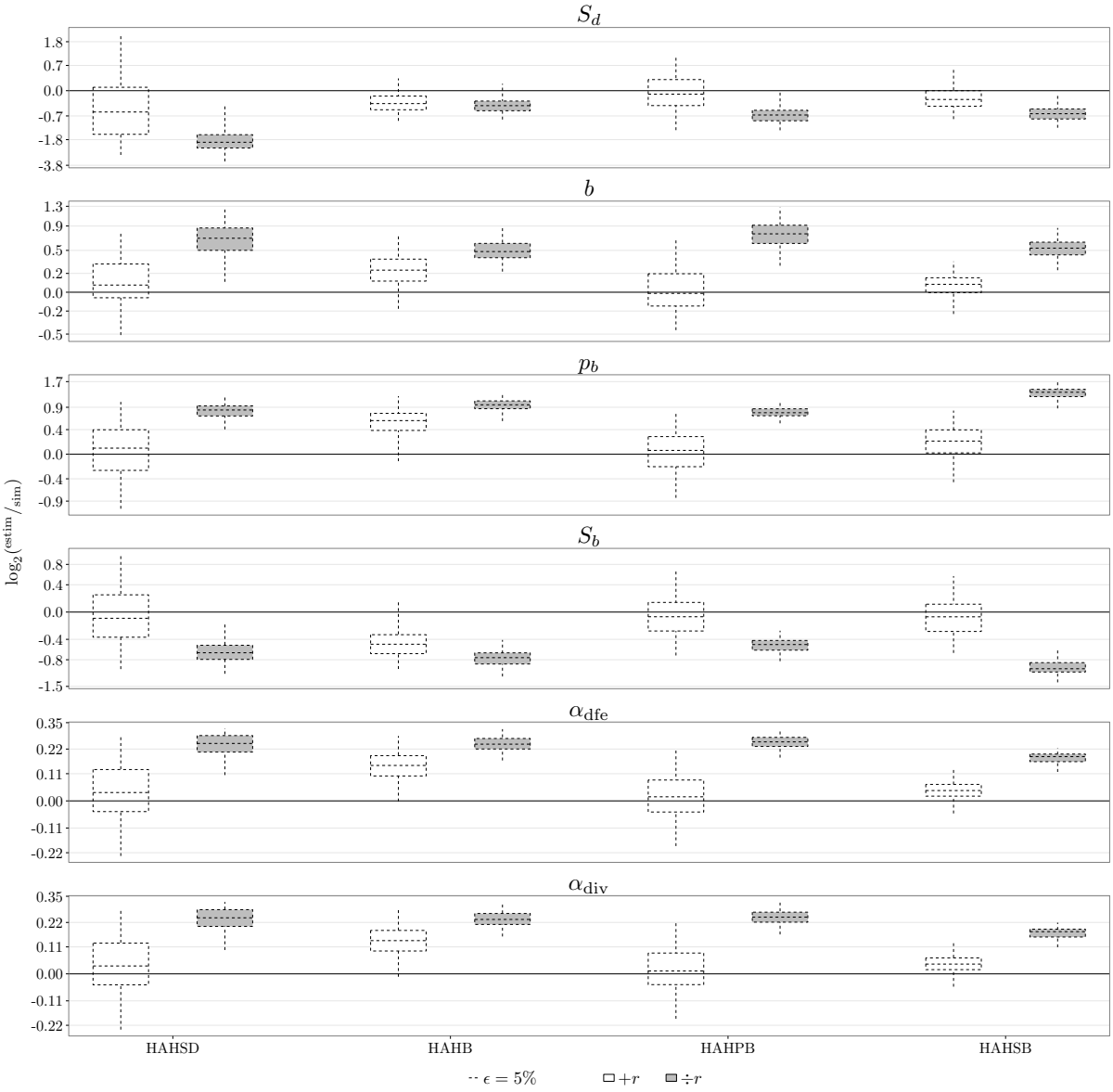

Figure S12: Box plot of inference quality on data simulated under different full DFEs with high α, given on the x-axis and detailed in Table 1 from main text, where divergence data was used. The nuisance parameters r were either estimated ($+r$, white) or fixed to 1 ($\div r$, gray). Misidentification of ancestral state was added with $\epsilon = 0.05$, but in the inference, ϵ was set to 0 and was not estimated.

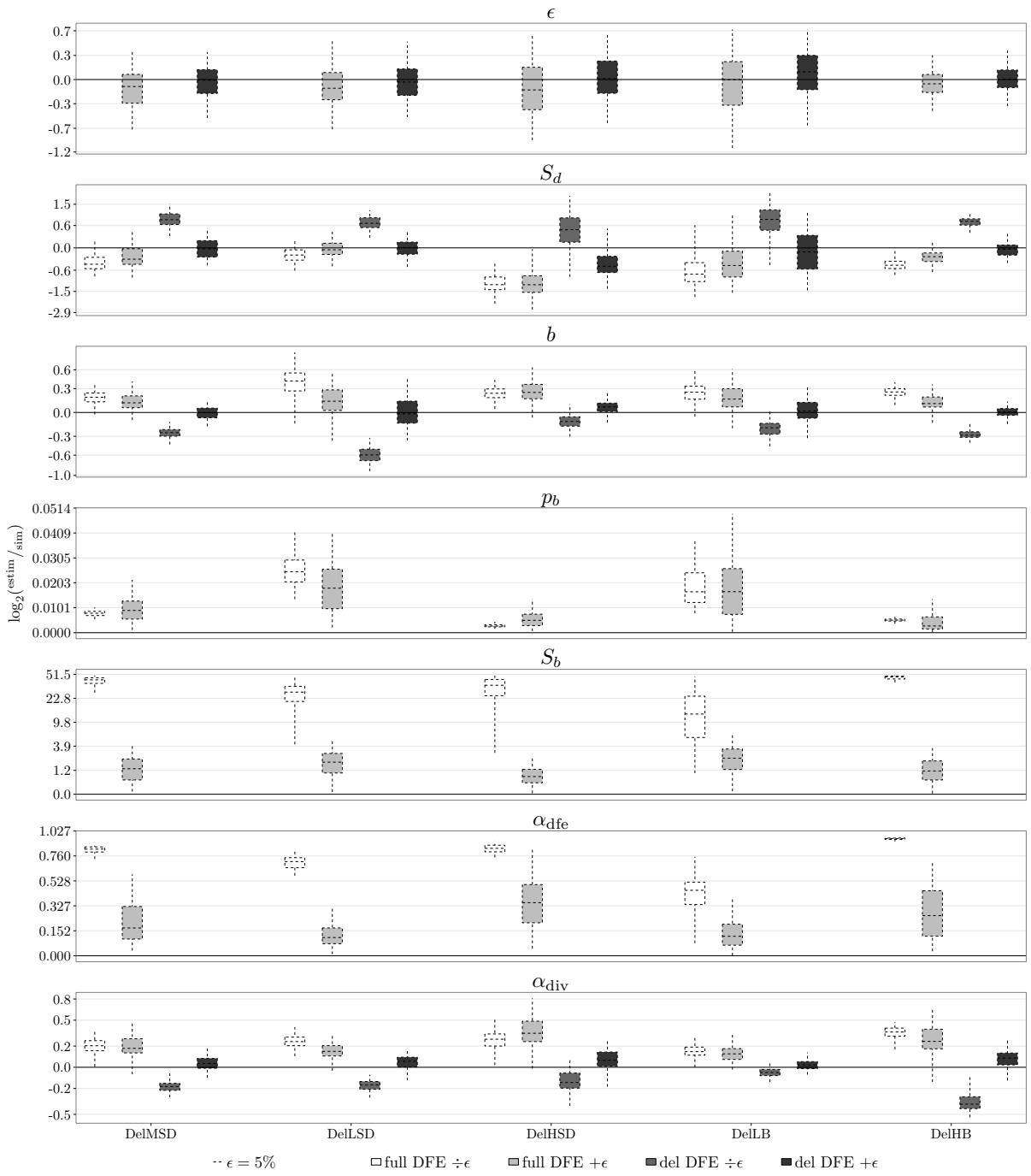

Figure S13: Box plot of inference quality on data simulated under different deleterious DFEs, given on the x-axis and detailed in Table 1 from main text. Misidentification of ancestral state was added with $\epsilon = 0.05$. The DFE parameters are inferred using only polymorphism data. The ancestral error parameter ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$).

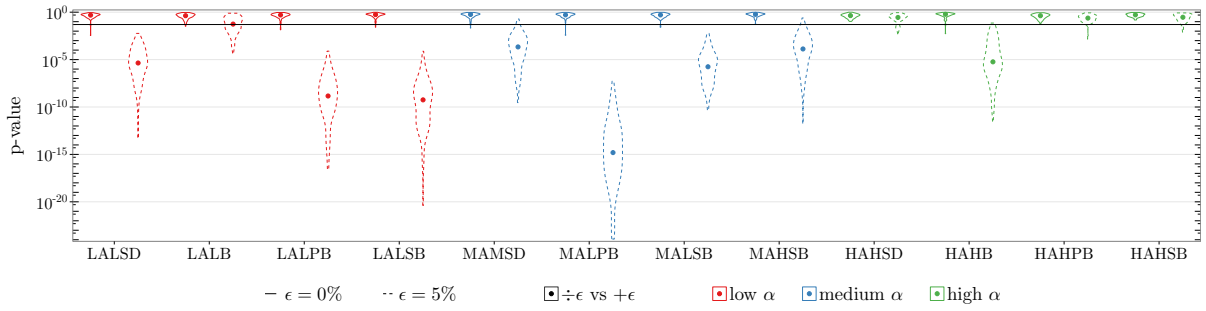

Figure S14: Violin plot of p-values for LRTs for evidence of ϵ on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text, where divergence data was used. Misidentification of ancestral state was added with $\epsilon = 5\%$. The LRT is performed using the maximum likelihoods found when inferring a full DFE where ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$). For low p-values, the model where $\epsilon \neq 0$ is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

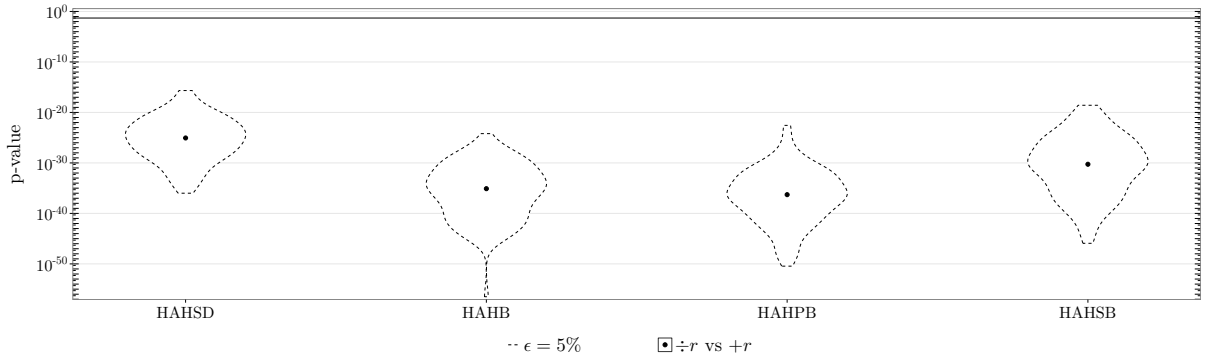

Figure S15:

Violin plot of p-values for LRTs for evidence of r parameters on data simulated under different full DFEs with high α, given on the x-axis and detailed in Table 1 from main text, where divergence data was used. Misidentification of ancestral state was added with $\epsilon = 0.05$. The LRT is performed using the maximum likelihoods found when inferring a full DFE and ϵ was fixed to 0, where the r_i parameters were either estimated ($+r$) or fixed to 1 ($\div r$). For low p-values, the model where $r_i \neq 1$ is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

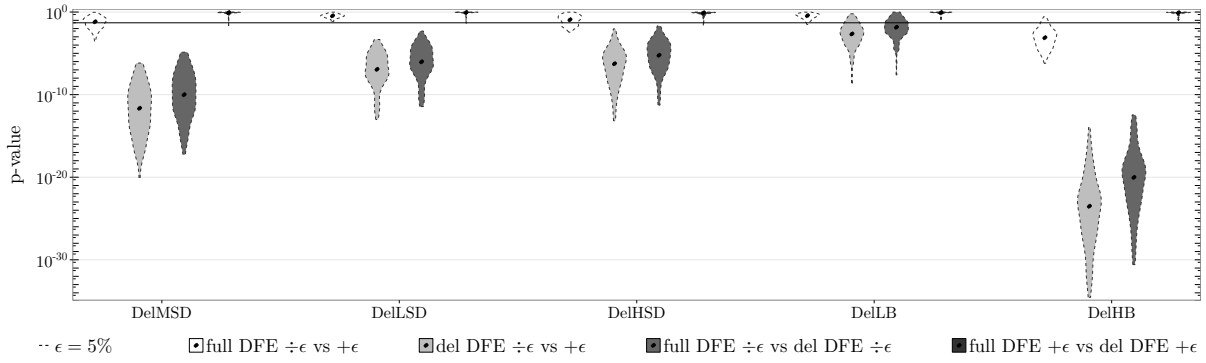

Figure S16: Violin plot of p-values for LRTs for evidence of ϵ and beneficial mutations in the polymorphism data on data simulated under different deleterious DFEs, given on the x-axis and detailed in Table 1 from main text. Misidentification of ancestral state was added with $\epsilon = 0.05$. The LRT is performed using the maximum likelihoods found when inferring a: full DFE where ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$) (white); deleterious DFE where ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$) (light gray); full or deleterious DFE where ϵ was fixed to 0 ($\div\epsilon$) (medium gray); or full or deleterious DFE where ϵ was estimated ($+\epsilon$) (dark gray). For low p-values, the model where: $\epsilon \neq 0$ (white and light gray); or a full DFE is inferred (medium and dark gray), is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

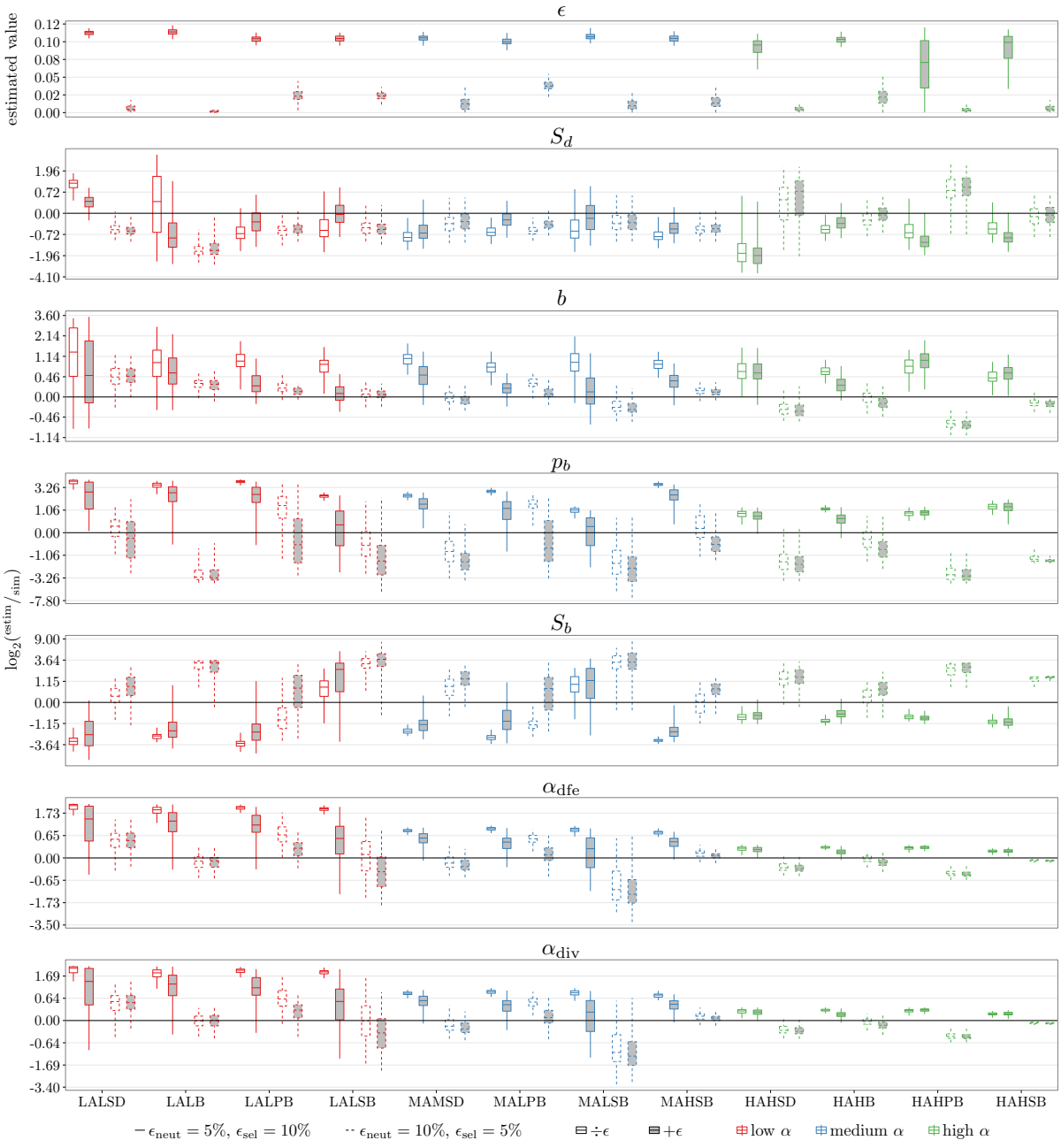

Figure S17: Box plot of inference quality on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text, where divergence data was used. Misidentification of ancestral state was added with two different error rates, ϵ_{neut} and ϵ_{sel}, on the sites assumed to be evolving neutrally and under selection, respectively. The ancestral error parameter ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$).

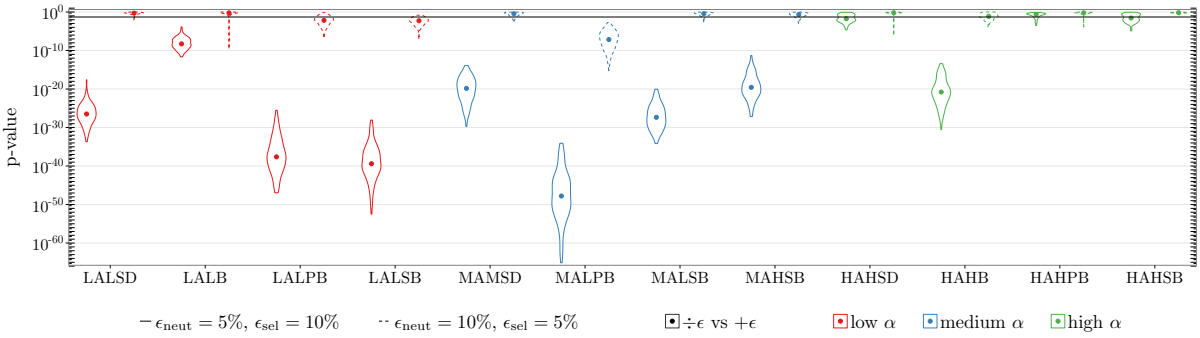

Figure S18: Violin plot of p-values for LRTs for evidence of ϵ on data simulated under different full DFEs, given on the x-axis and detailed in Table 1 from main text, where divergence data was used. Misidentification of ancestral state was added with two different error rates, ϵ_{neut} and ϵ_{sel}, on the sites assumed to be evolving neutrally and under selection, respectively. The LRT is performed using the maximum likelihoods found when inferring a full DFE where ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$). For low p-values, the model where $\epsilon \neq 0$ is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

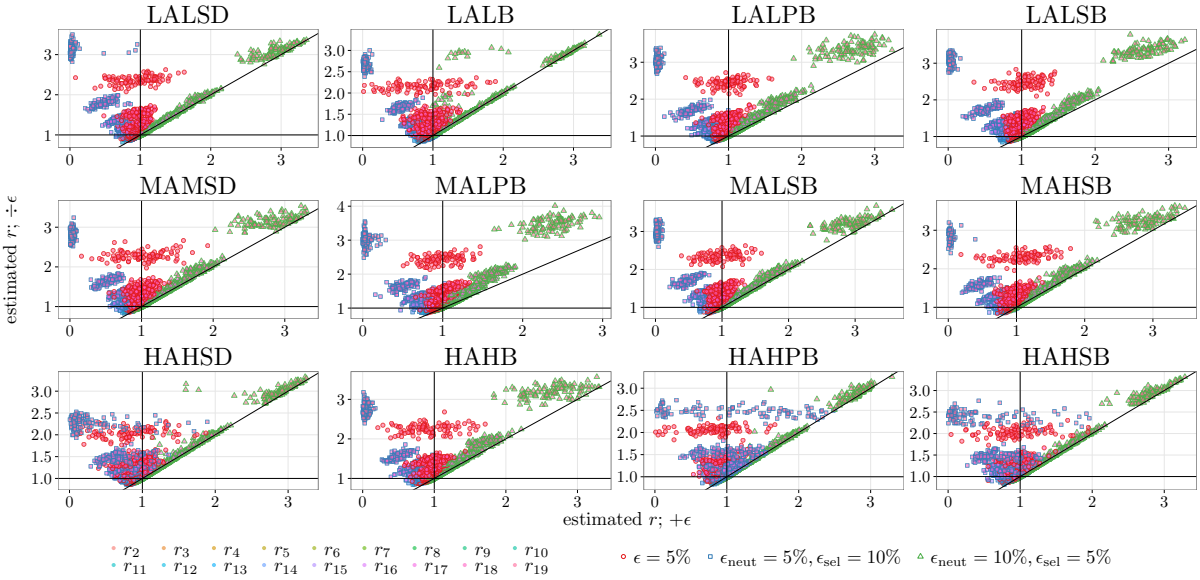

Figure S19: Estimated r parameters on data simulated under different full DFEs, detailed in Table 1 from main text, where divergence data was used. Misidentification of ancestral state was added either with $\epsilon = 0.05$, or with two different error rates, ϵ_{neut} and ϵ_{sel}, on the sites assumed to be evolving neutrally and under selection, respectively. The ancestral error parameter ϵ was either estimated ($+\epsilon$, given on the x-axis) or fixed to 0 ($\div\epsilon$, given on the y-axis). The r parameters correspond to the estimates given in Figures S11 and S17.

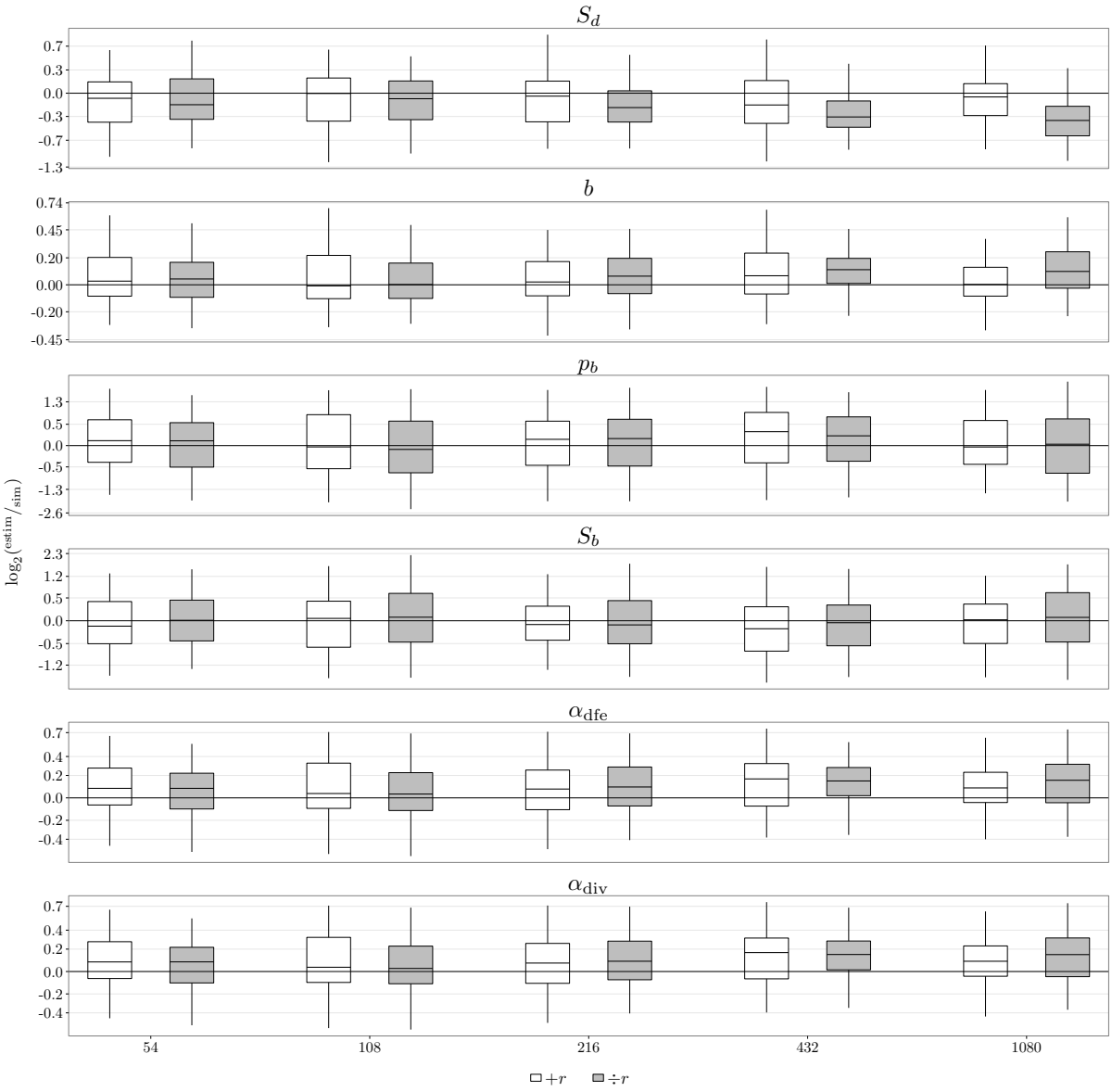

Figure S20: Box plot of inference quality on data simulated under MAMSD DFE, detailed in Table 1 from main text, where divergence data was used, and different levels of linkage. The x-axis indicates the number of linked sites used, as described in the *Simulation setups* section. The nuisance parameters r were either estimated ($+r$, white) or fixed to 1 ($\div r$, gray).

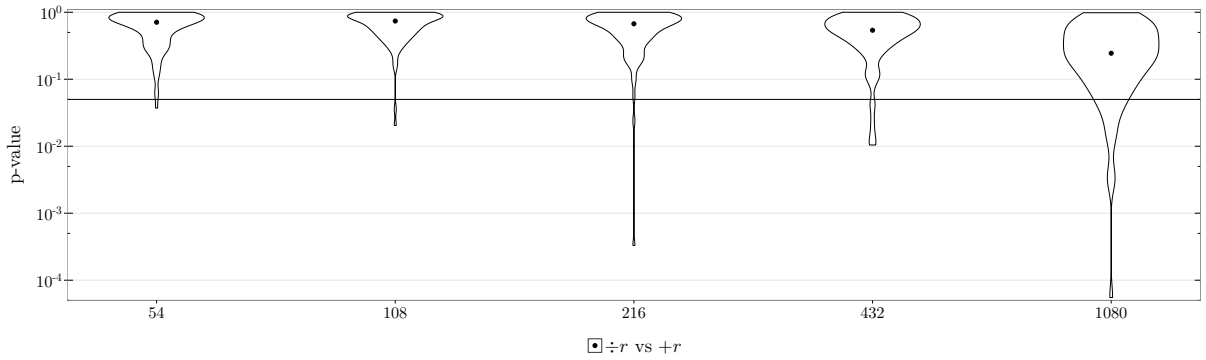

Figure S21: Violin plot of p-values for LRTs for evidence of r parameters on data simulated under MAMSD DFE, detailed in Table 1 from main text, where divergence data was used, and different levels of linkage. The x-axis indicates the number of linked sites used, as described in the *Simulation setups* section. The LRT is performed using the maximum likelihoods found when inferring a full DFE, where the r_i parameters were either estimated ($+r$) or fixed to 1 ($\div r$). For low p-values, the model where $r_i \neq 1$ is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

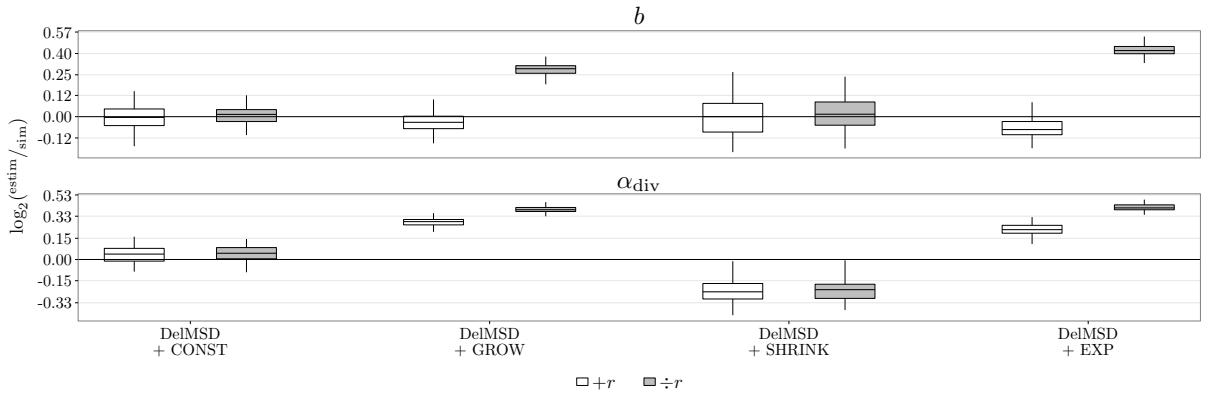

Figure S22: Box plot of inference quality on data simulated under DelMSD DFEs detailed in Table 1 from main text, and different demographic scenarios as described in the *Simulation setups* section. The DFE parameters are inferred using only polymorphism data and assuming a deleterious DFE. The nuisance parameters r were either estimated ($+r$, white) or fixed to 1 ($\div r$, gray).

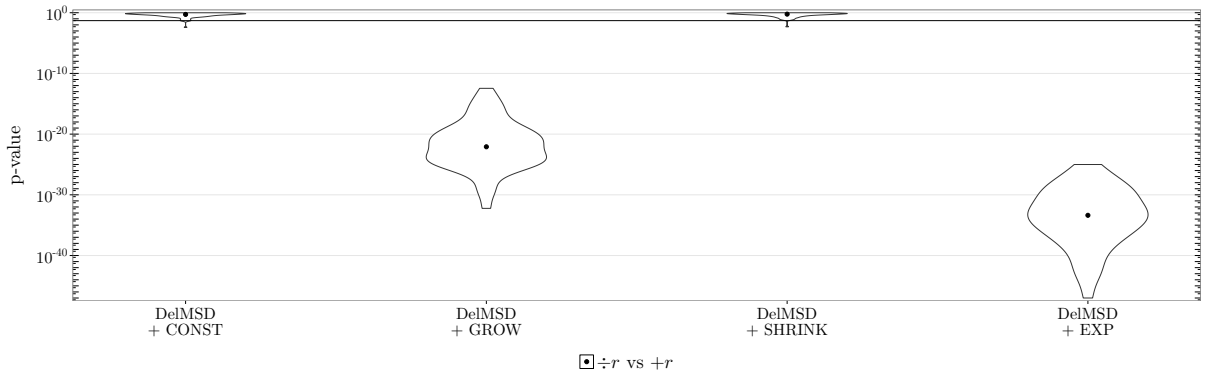

Figure S23: Violin plot of p-values for LRTs for evidence of r parameters on data simulated under DelMSD DFE, detailed in Table 1 from main text, and different demographic scenarios, as described in the *Simulation setups* section. The LRT is performed using the maximum likelihoods found when inferring a deleterious DFE, where the r_i parameters were either estimated ($+r$) or fixed to 1 ($\div r$). For low p-values, the model where a $r_i \neq 1$ is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

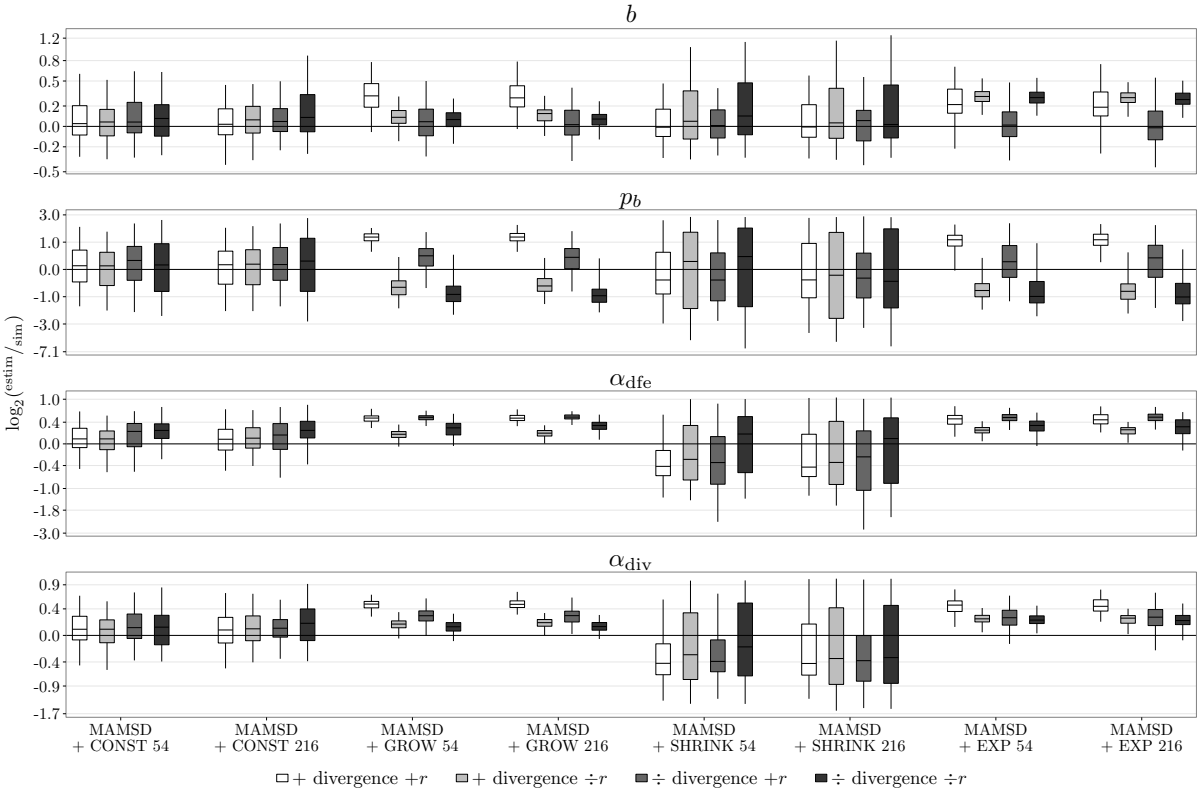

Figure S24: Box plot of inference quality on data simulated under MAMSD DFE, detailed in Table 1 from main text, different demographic scenarios and two different levels of linkage, as described in the *Simulation setups* section. The divergence data is either used (+ divergence), or not ($\div$ divergence, and the nuisance parameters r were either estimated ($+r$) or fixed to 1 ($\div r$).

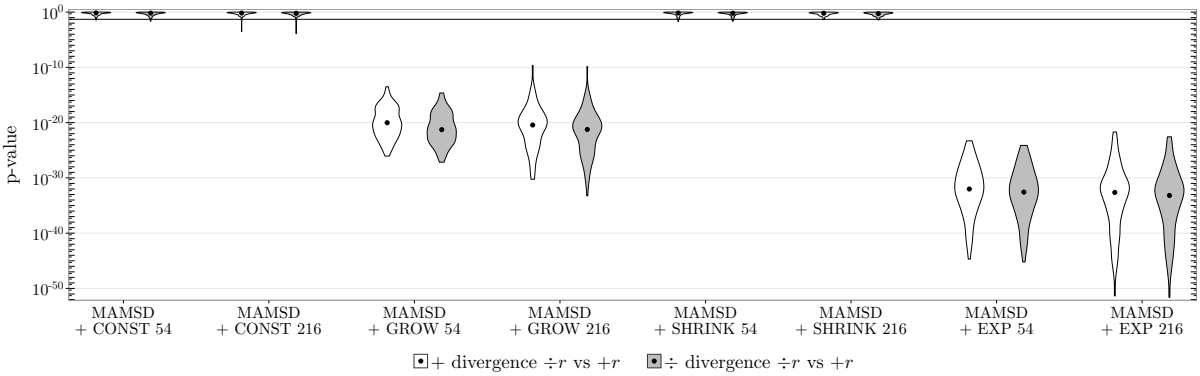

Figure S25: Violin plot of p-values for LRTs for evidence of r parameters on data simulated under MAMSD DFE, detailed in Table 1 from main text, different demographic scenarios and two different levels of linkage, as described in the *Simulation setups* section. The LRT is performed using the maximum likelihoods found when inferring a full DFE, while using (+ divergence) or not ($\div$ divergence) divergence data, where the r_i parameters were either estimated ($+r$) or fixed to 1 ($\div r$). For low p-values, the model where $r_i \neq 1$ is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

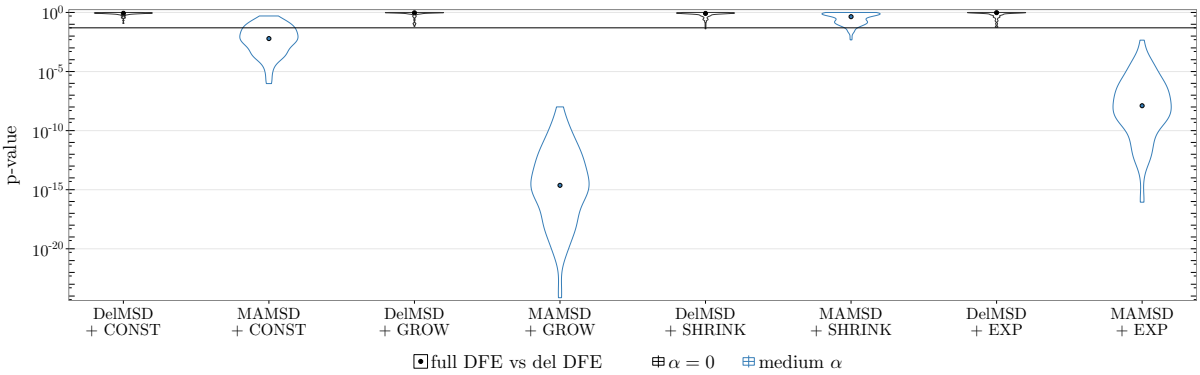

Figure S26: Violin plot of p-values for LRTs for evidence of beneficial mutations in the polymorphism data on data simulated under MAMSD and DelMSD DFE, detailed in Table 1 from main text, and different demographic scenarios, as described in the *Simulation setups* section. The LRT is performed using the maximum likelihoods found when inferring a full or deleterious DFE from polymorphism data alone. For low p-values, the model where a full DFE is inferred is preferred. Dots indicate the median. Horizontal line shows the 5% threshold.

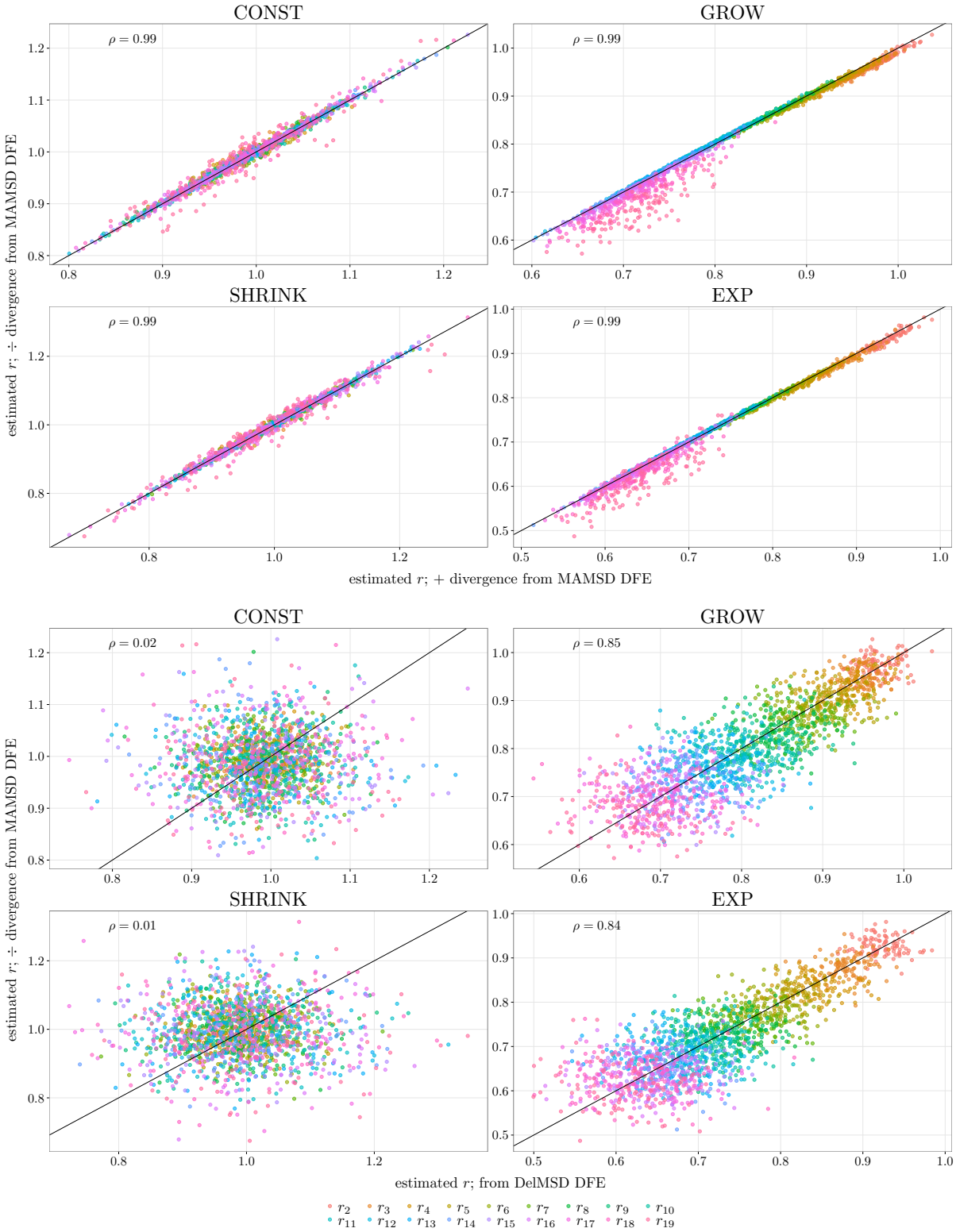

Figure S27: Correlation of estimated r parameters for different demographic scenarios, as described in the *Simulation setups* section. Top: comparison of inference from data simulated under MAMSD DFE, with (on the x-axis) or without (on the y-axis) divergence. Bottom: comparison of inference from data simulated under MAMSD DFE (on the y-aixs) and DelMSD DFE (on the x-axis). The r parameters correspond to the estimates given in Figures S24 and S22. Here, ρ is Pearson's correlation coefficient.

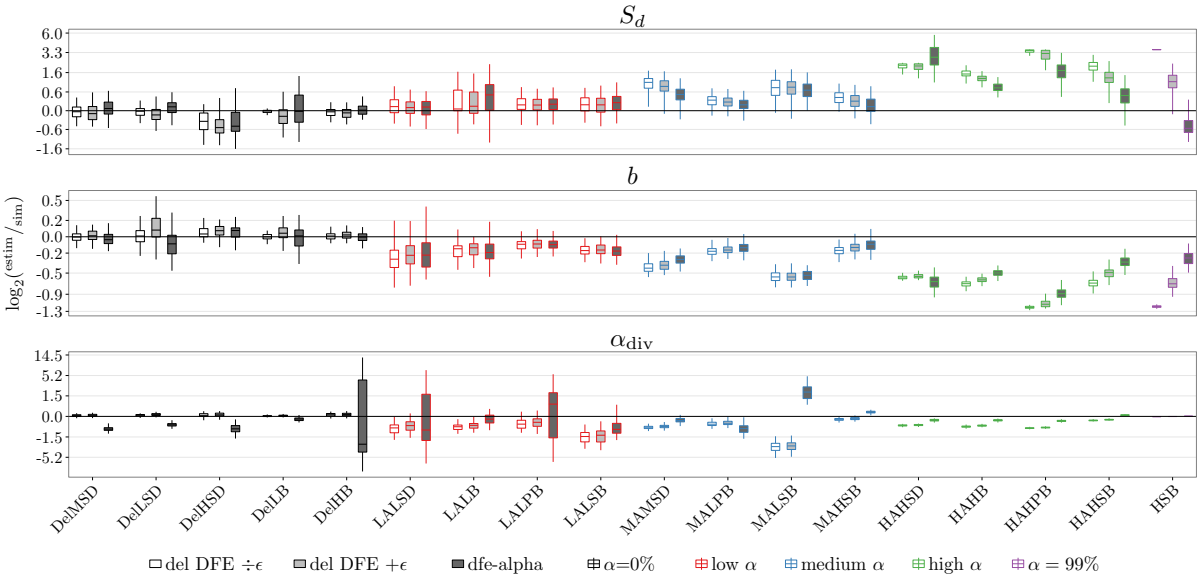

Figure S28: Comparison with dfe-alpha on data simulated under different full and deleterious DFEs, given on the x-axis and detailed in Table 1 from main text. The DFE parameters are inferred using only polymorphism data. The ancestral error parameter ϵ was either estimated ($+\epsilon$) or fixed to 0 ($\div\epsilon$).

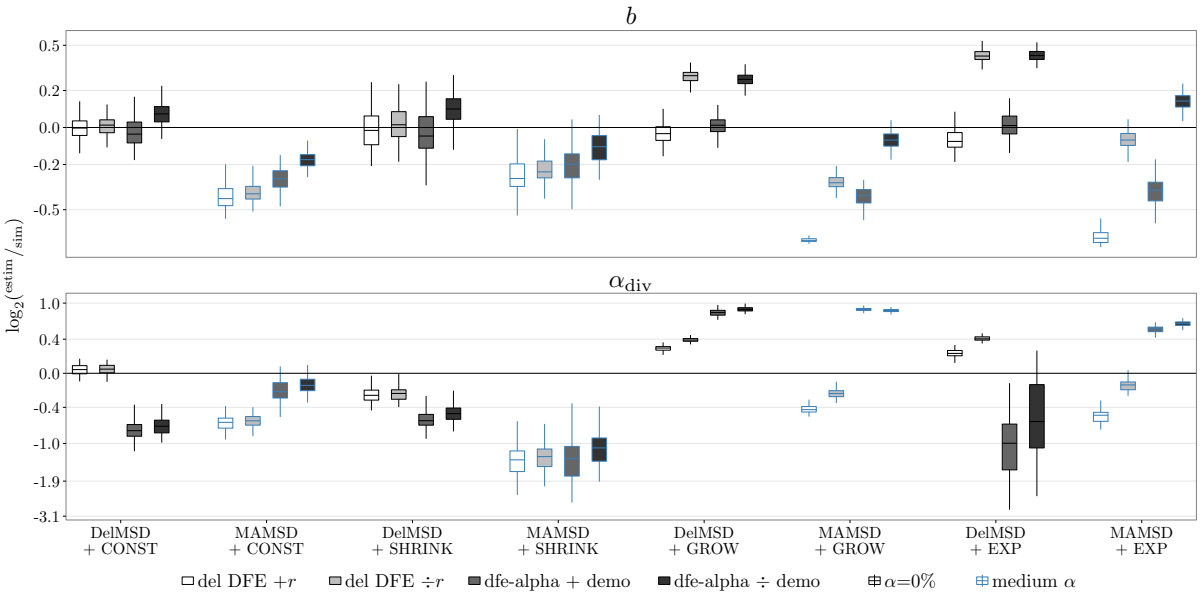

Figure S29: Comparison with dfe-alpha on data simulated under MAMSD and DelMSD DFEs, detailed in Table 1 from main text, and different demographic scenarios, as described in the *Simulation setups* section. The DFE parameters are inferred using only polymorphism data. The nuisance parameters r were either estimated ($+r$) or fixed to 1 ($\div r$), and correspondingly, demography was estimated (+ demo) or not ($\div$ demo) when running dfe-alpha.

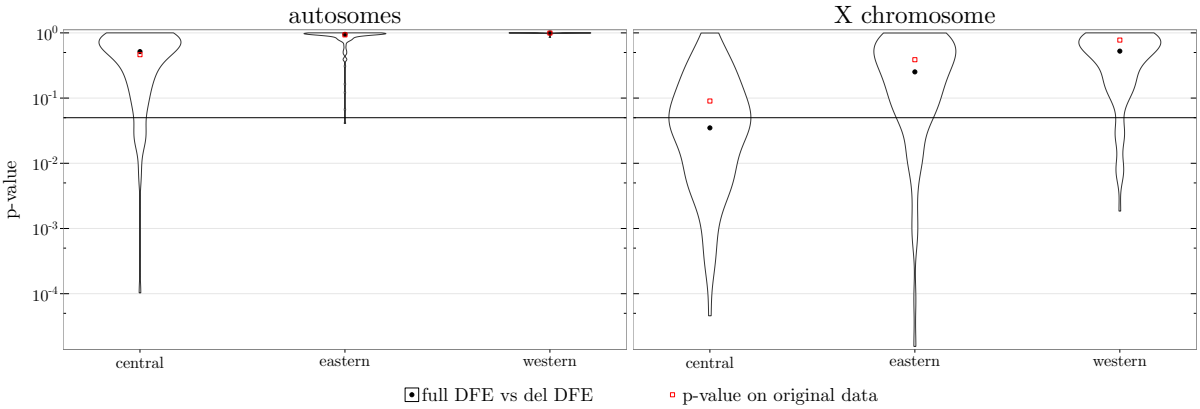

Figure S30: Violin plot of p-values for LRTs for evidence of beneficial mutations in the polymorphism data on the bootstrap chimpanzee exome data, for all three subspecies, given on the x-axis. The LRT is performed using the maximum likelihoods found when inferring a full or deleterious DFE. For low p-values, the model where a full DFE is inferred is preferred. Dots indicate the median. Horizontal line shows the 5% threshold. The red squares mark the p-values obtained when performing the LRT on the original full data sets.

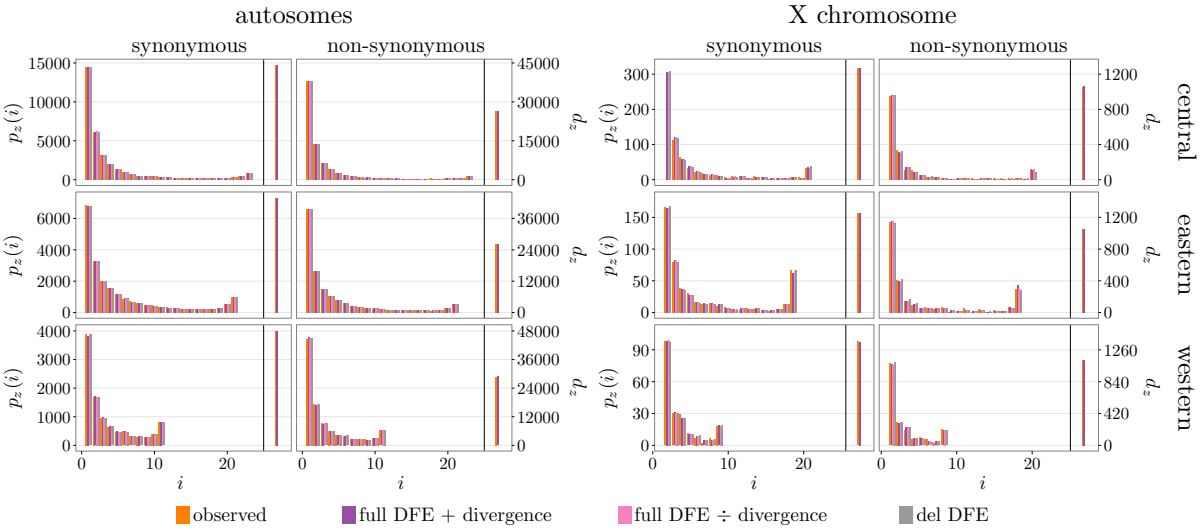

Figure S31: Observed and expected SFS ($p_z(i)$) and divergence (d_z) counts for the three chimpanzee subspecies. A full DFE was inferred from both polymorphism and divergence data, while also both a full DFE and deleterious DFE were inferred from polymorphism data only.