**Inferring the distributions of fitness effects and proportions of strongly deleterious mutations**

**Supplementary Information**

**S1. Test of under neutral conditions….....………………………………….2**

**S2. Calibration of for scaling to ……………………………………………...…2**

**S3. Test of the accuracy of inference under cases with ………...…...……..……..4**

**S4. Command lines and init\_files for polyDFE runs……………………………………….5**

**S5. Probability distributions of weakly deleterious DFEs for multi-species tests…...…...5**

**S6. Replication of main results under larger population and sample size………...……...6**

**References…………………………………………………………………………………….8**

**S1. Test of under neutral conditions**

To check that our model behaves as we would expect under a Wright-Fisher population, we measure if the steady-state variance (after the burn-in period) at the neutral linked locus is equal to once all mutations are neutral (Lynch & Hill 1986; Keightley & Otto 2006). Fig. S1 shows an example from a simulation replicate, highlighting how

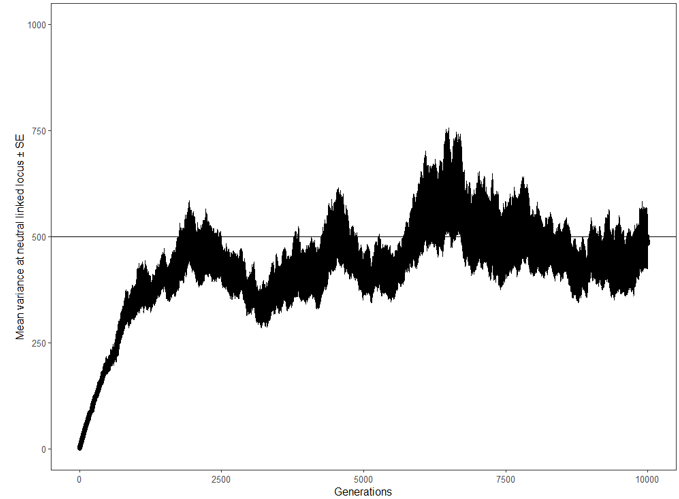
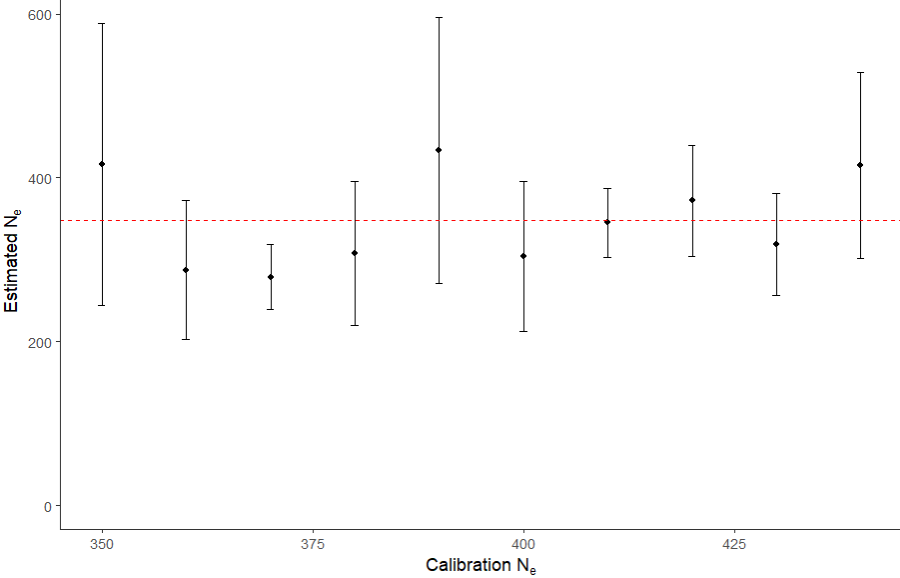


Figure S1: Mean of 50 replicates over time with only neutral mutations SE. The mean after a burn-in period of generations) was . The straight black line denotes .

**S2. Calibration of for scaling to**

As shown in S1, was estimated by measuring variance at a neutral linked locus (Hill & Lynch 1986; Keightley & Otto 2006). Since is used to scale mutational effects sampled as population-scaled mutational effect () to selection coefficients (, it is necessary to know approximately which value of will emerge in a simulation. Because the emerging value of is itself affected by the simulation , we estimated an approximate emerging byplotting a standard curve of estimated steady-state against an assumed (the value used to re-scale 2 to ) for a given gamma distribution of selective effects. This allowed us to know the approximate steady-state which would emerge in our Wright-Fisher simulation before we used these values to simulate data for SFS construction.

For each standard curve, 10 x 10 replicate simulations with differing assumed (calibrating ) were simulated and the emerging mean was calculated (using variance at the neutral linked locus, see S1) and used for re-scaling mutational effects in simulations from which SFSs where calculated (for an example of a standard curve, see Fig. S2). The used to scale the mutational effects for the actual simulations was selected as the mean of the emerging of the calibration simulations. The emerging in simulations from which SFSs where calculated was similar to the approximate estimated by this calibration procedure.

  
Figure S2: Example of standard curve used to estimate which would emerge in a simulation given some used for scaling mutational effects and . Estimated values as a function of assumed values (calibration ) used to re-scaled mutational effects from to . Each point is the mean of 10 simulation replicates with standard error. For these simulations, For all simulation sets, the mean emerging was calculated (dashed red line) and this was used to scale mutational effects in the simulations from which SFSs were calculated.

**S3. Test of the accuracy of inference under test cases with .**

Compared to SFS\_CODE (Hernandez 2008), our model implements the idea of and recombination slightly differently (see main text for details). To ensure that polyDFE can still accurately infer DFEs from our output, we replicated Fig. 4A of Tataru et al. 2017 to test the accuracy of inference under standard cases with (Fig. S3).

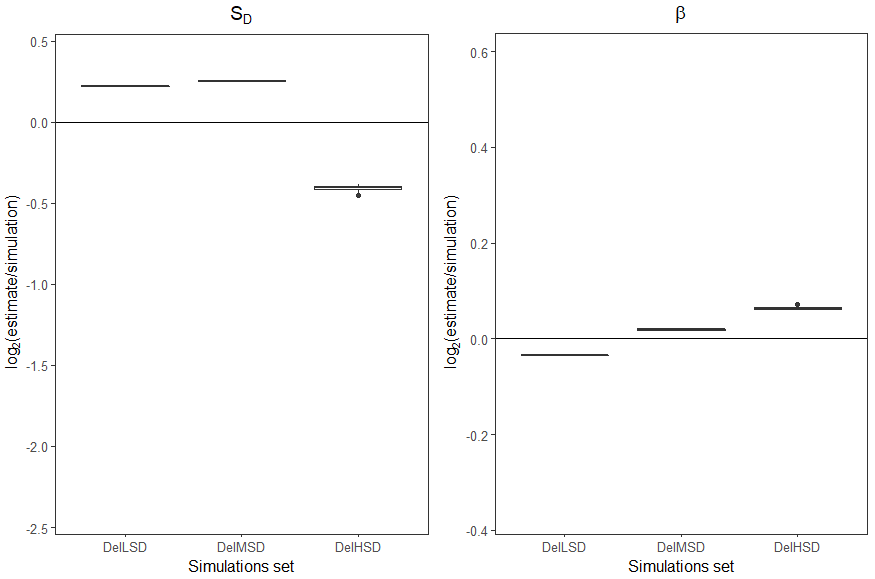


Figure S3: Boxplots of 10 polyDFE runs showing the accuracy of inference of DFE for the gamma distributions DelLSD, DelMSD and DelHSD of Tataru et al. 2017. The left panel shows the accuracy of inference of , while the right panel shows the same for . As expected, the results show slightly lower but comparable accuracy of inference to Tataru et al. 2017 despite a slightly different model.

For , the inferred means for the three groups DelLSD, DelMSD and DelHSD were -11.62, -476.77 and -7531.32 and the true values were -10, -400 and -10000. In all cases, the true value of β was 0.40. The remaining discrepancy between true and inferred values under our model likely stems from the fact that SFS\_CODE (which is what Tataru et al used) treats as a parameter (chosen by the user) and scales mutation and recombination rates accordingly, whereas in our model just emerges given the chosen (see also S2).

**S4. Command lines and init\_files for polyDFE runs**

Below are examples showing a polyDFE command line and a polyDFE init file used for the analysis resulting in Fig. 1 in the main text. For details on polyDFE command line and init files, please see the polyDFE documentation; Tataru et al. 2017.

polyDFE was run using the following command line:

./polyDFE -d 0\_SFS.txt -i init\_model\_BandH.txt 1 -v 100 -w > output

For the init files, which contains the following parameters

#ID eps an eps cont lambda theta bar a S d b p b S b r (at least #groups-1 of them)

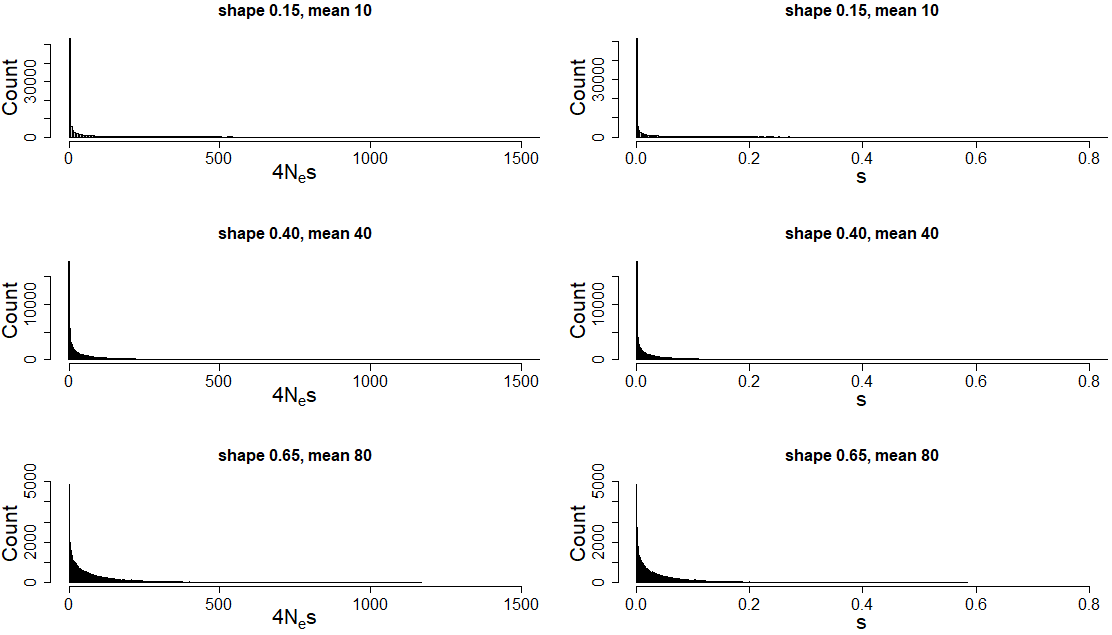
the following line was used

1 1 0.00 1 0.00 0 0.005 0 0.001 1 -1 0 -17694.63 0 0.22 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

where -17694.63 and 0.22 in this case are examples of randomly sampled initial values for and , respectively. For more details on random sampling of initial values, please see the main text.

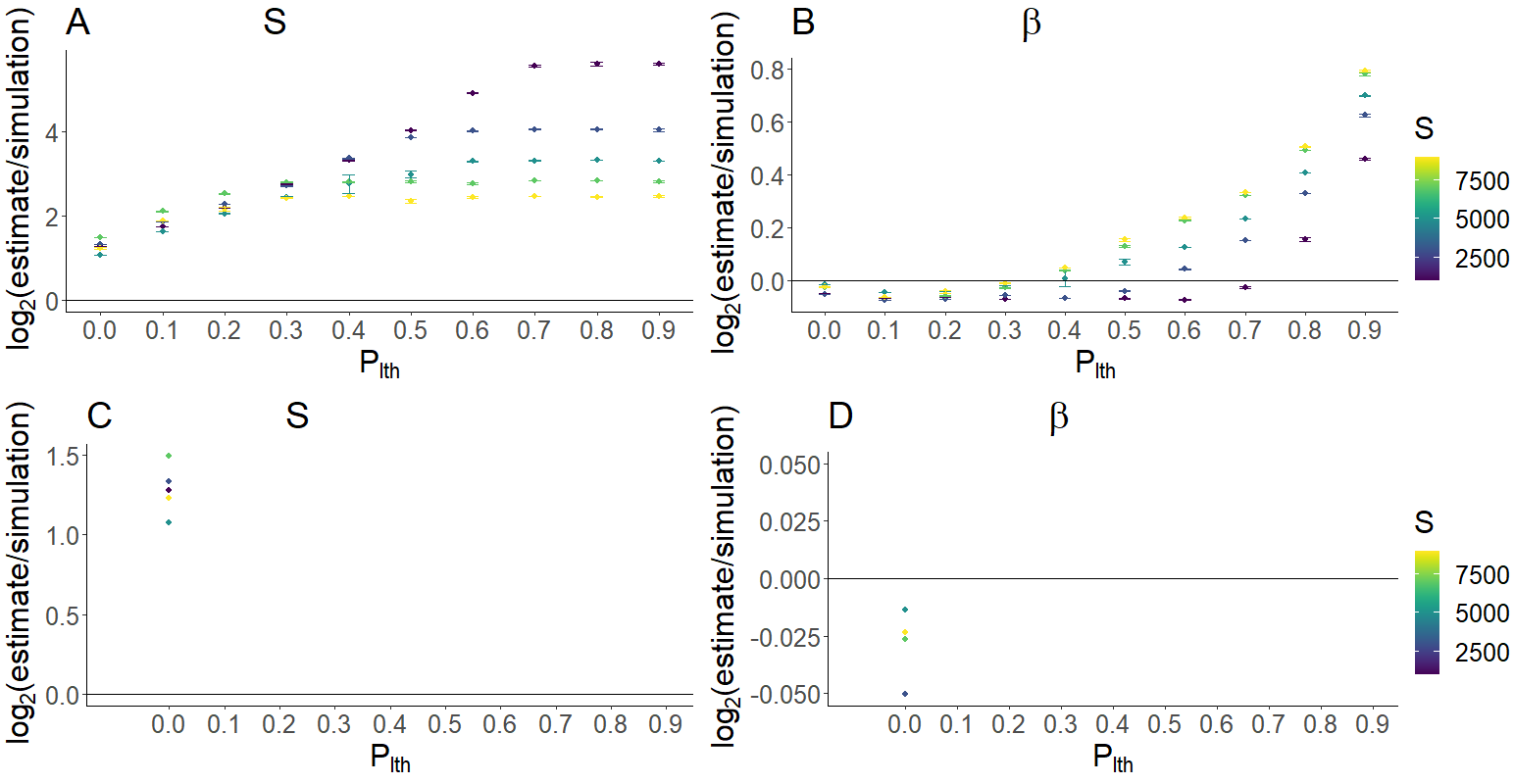
**S5. Probability distributions of weakly deleterious DFEs for multi-species tests**

To test whether high could maximize the likelihood of a DFE in model even in cases where essentially zero strongly deleterious mutations are present, we defined three weakly deleterious gamma-distributed DFEs with different shape parameters and means . We then fitted DFE models to SFSs from all three distributions (a multi-species DFE model) and studied how the likelihood of the model was affected by changing (see main text). The probability distribution of mutational effects from these three DFEs are visualized in the following (Fig. S4).

Figure S4: Visualisation of the probability density of 3 weakly deleterious DFEs with the shape parameters and means of , and for simulated Wright-Fisher populations with . Mutational effects are shown as (left column) and -coefficients (right column).

**S6. Replication of the main results under larger population and sample size**

To show examples of the main results under larger population size and larger number of sampled haplotypes, we ran Wright-Fisher simulations with and (see main text for details). We then tested how the accuracy of inference was affect by when a DFE was inferred based on SFSs calculated from a single species (Fig. S5). We also tested whether would still artificially maximize the likelihood given higher and when a single DFE was inferred for SFSs from multiple species (Fig. S6).

Figure S5: Mean accuracy of inference for (A) and (B) under the assumption of different values of for different Wright-Fisher simulation sets (differing in the simulation , 2000, 6000, 10000, 140000, 18000). Each point shows the mean value of 10 polyDFE replicates. Bars denote 95% confidence intervals (in most cases, these are not visible as they lie within the point). For each simulation set (color), bottom panels show the subset that had the highest accuracy of inference of . For these subsets, the accuracy of inference for (C) and (D) are shown. is equivalent to perfect accuracy of inference.

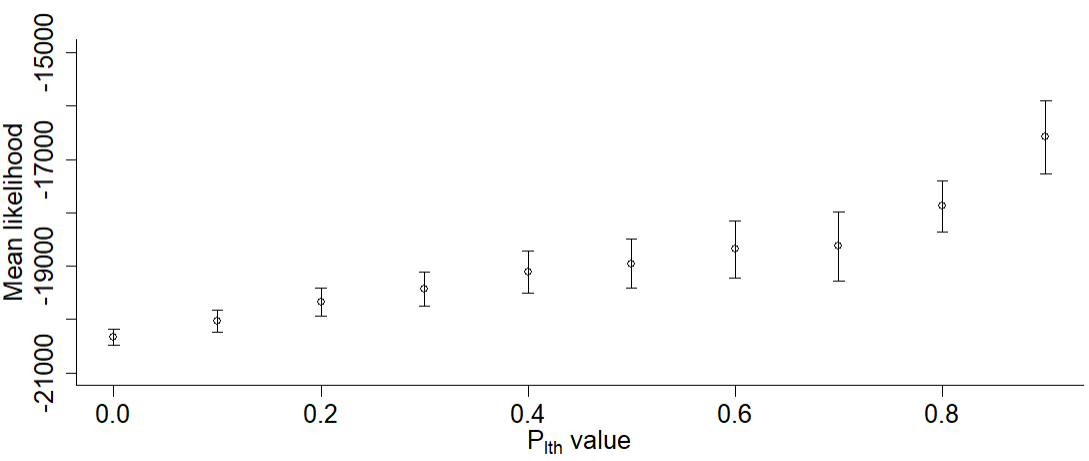


Figure S6: Mean likelihood with 95% confidence intervals returned by polyDFE for a multi-species DFE model against 10 values. was inferred and was assumed fixed at a randomly sampled value. 20 neutral and selected SFSs from 3 different simulated deleterious DFEs with the shape parameter and of and , respectively, were combined into one data input file. Each point represents the mean value of 10 polyDFE runs.

Under higher population size and higher number of sampled haplotypes, the results remain qualitatively the same: may improve the accuracy of inference if is high (Fig. S5A). When a DFE is fitted to SFS data from multiple species, high values of artificially maximize the likelihood of the DFE model (Fig. S6) despite no proportion of mutations being present in the dataset (Appendix A in the main text).

**References**

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

Keightley, P. D., and S. P. Otto. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443:89–92.

Lynch, M. Hill, W. 1986. Phenotypic Evolution by Neutral Mutation. Evolution 40(5):915-935.

Tataru, P., M. Mollion, S. Glémin, and T. Bataillon. 2017. Inference of Distribution of Fitness Effects and. Genetics 207:1103–1119.