File S1

Supporting Information

1 Assessment of the methods

1.1 Comparison of the median of the posterior distribution of the uniparental proportion

We address the relative accuracy of estimates of the uniparental proportion s^* produced by our Bayesian method relative to those produced by `RMES` and the F_{IS} method (27) upon application to simulated data (compare Assessment of Accuracy and Coverage using Simulated Data section). While we summarize our posterior distributions of s^* by the *mode* in the main text, we here use the *median*.

We first address application of the methods to simulated data under the large-sample regime ($n = 70$ individuals, $L = 32$ loci). Except in cases in which the true s^* is very close to 0, the error for `RMES` exceeds the error for our method (Figure S1), a trend that is apparent under under both the large- and small-sample regimes. The error for the F_{IS}-based estimate also exceeds the error for our method. It is largest near $s^* = 0$ and vanishes as s^* approaches 1, a pattern distinct from `RMES`.

Both `RMES` and our method show positive bias upon application to data sets for which the true uniparental proportion s^* is close to zero and negative bias for s^* close to unity. This trend reflects that both methods yield estimates of s^* constrained to lie between 0 and 1. In contrast, the F_{IS}-based estimate (27) underestimates s^* throughout the range, even near $s^* = 0$ ($\widehat{F_{IS}}$ is not constrained to be positive). Our method has a bias near 0 that is substantially larger than the bias of `RMES`, and an error that is slightly larger. A major contributor to this trend is that our Bayesian estimate is represented by only the median of the posterior distribution of the uniparental proportion s^*.

Figure S2 indicates that for data sets generated under a true value of s^* of 0 (full random outcrossing), the posterior distribution for s^* has greater mass near 0. We suggest that the bias shown near $s^* = 0$ merely represents uncertainty in the posterior distribution for s^* and not any preference for incorrect values. We note that our method assumes that the data are derived from a population reproducing through a mixture of self-fertilization and random outcrossing. Assessment of a model of complete random mating ($s^* = 0$) against the present model ($s^* > 0$) might be conducted through the Bayes factor.

Figure S3 indicates that all methods show increased error upon application of smaller samples ($n = 10$ individuals, $L = 6$ loci), as expected. Comparison of our method and `RMES` show trends qualitatively similar to the large-sample case: positive bias upon application to data sets for which the true uniparental proportion s^* is close to zero and negative bias for s^* close to unity, with less error exhibited by our method throughout the range of the uniparental proportion (s^*).

1.2 Comparison of the median, mode, and mean

In addition to the mode (Figure 2) and median (Figure S1), error might also be assessed by consideration of the mean of the posterior distribution of s^*. Figure S4 suggests that the bias and root-mean-squared (rms) error of the mode, median, and mean exhibit different properties. For example, the posterior mode shows smaller bias throughout the parameter range, but the median and mean show smaller rms error for s^* near the boundaries (near 0 or 1). That the posterior mode does not display large bias near $s^* = 0$ is consistent with our suggestion that the larger error of the mean in that region reflects higher uncertainty.

1.3 Frequentist coverage

As for the 95% BCIs (Figure 4), Figure S5 indicates that BCIs of different nominal values (0.5, 0.75, 0.9, 0.95, and 0.99) display the same pattern, with coverage exceeding the desired value for intermediate true s^* values and dipping below the desired value for very high values

of s^*. Coverage is closer to the nominal value for the 0.99 and 0.95 levels than for the 0.5 level.

2 Data analysis

2.1 Self-fertilizing vertebrate

Figure S6 shows the posterior distributions of number of generations since the most recent outcross event (13) for each sampled individual in the highly inbred BP population of *K. marmoratus*. Figure S7 shows the posterior distributions for the more outbred TC population.

Figures S8 and S9 present posterior distributions of locus-specific mutation rates for the BP and TC populations, respectively. For each locus, Fig. S10 compares the rank order of its median mutation rate estimated from the BP data set versus that from the TC data set. If a relationship exists between the mutation rates estimated from the datasets, it appears to be diffuse.

2.2 Gynodioecious plant

Figure S11 presents posterior distributions for the uniparental proportion (s_G), the proportion of females among reproductives (p_f), the proportion of seeds set by hermaphrodites by self-pollen ($\tilde{s}$), and the viability of uniparental offspring relative to biparental offspring (τ).

Figure S12 presents the inferred number of generations since the most recent outcross event T_k (13) for each individual k.

Figure S13 presents posterior distributions for locus-specific mutation rates inferred from the *S. salicaria* data set. The loci appear to have similar posterior medians.

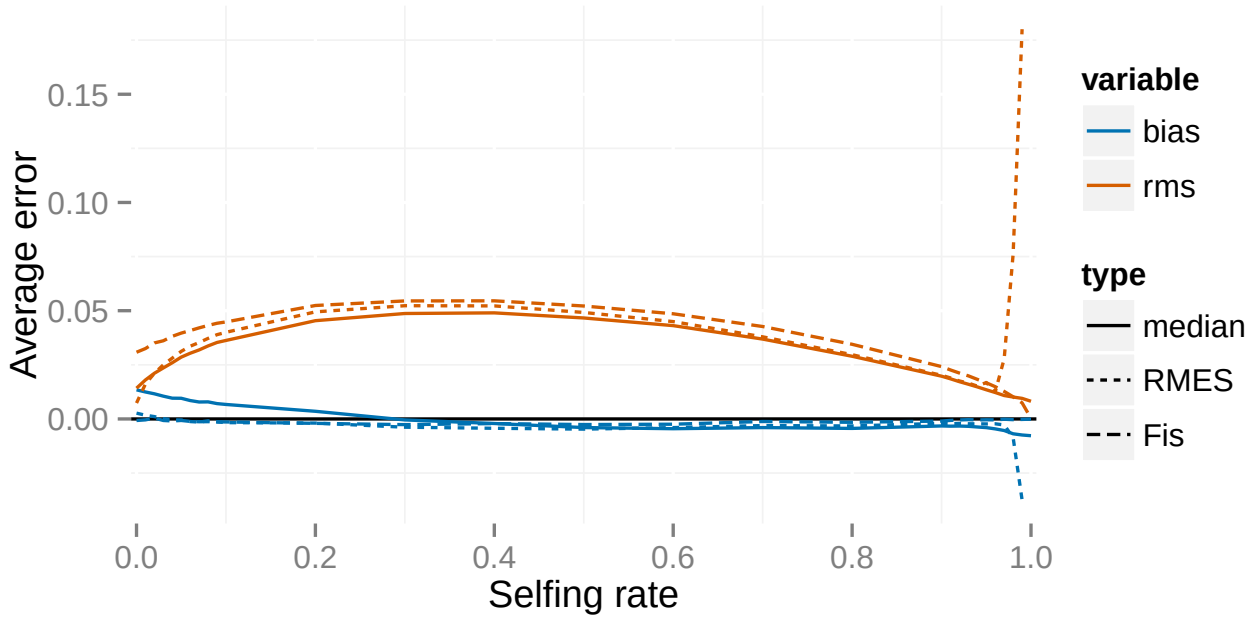


Figure S1 Errors for the full likelihood (posterior median), `RMES`, and F_{IS}-based (27) methods for a large simulated sample ($n = 70$ individuals, $L = 32$ loci). In the legend, rms indicates the root-mean-squared error and bias the average deviation. Averages are taken across simulated data sets at each true value of s^*.

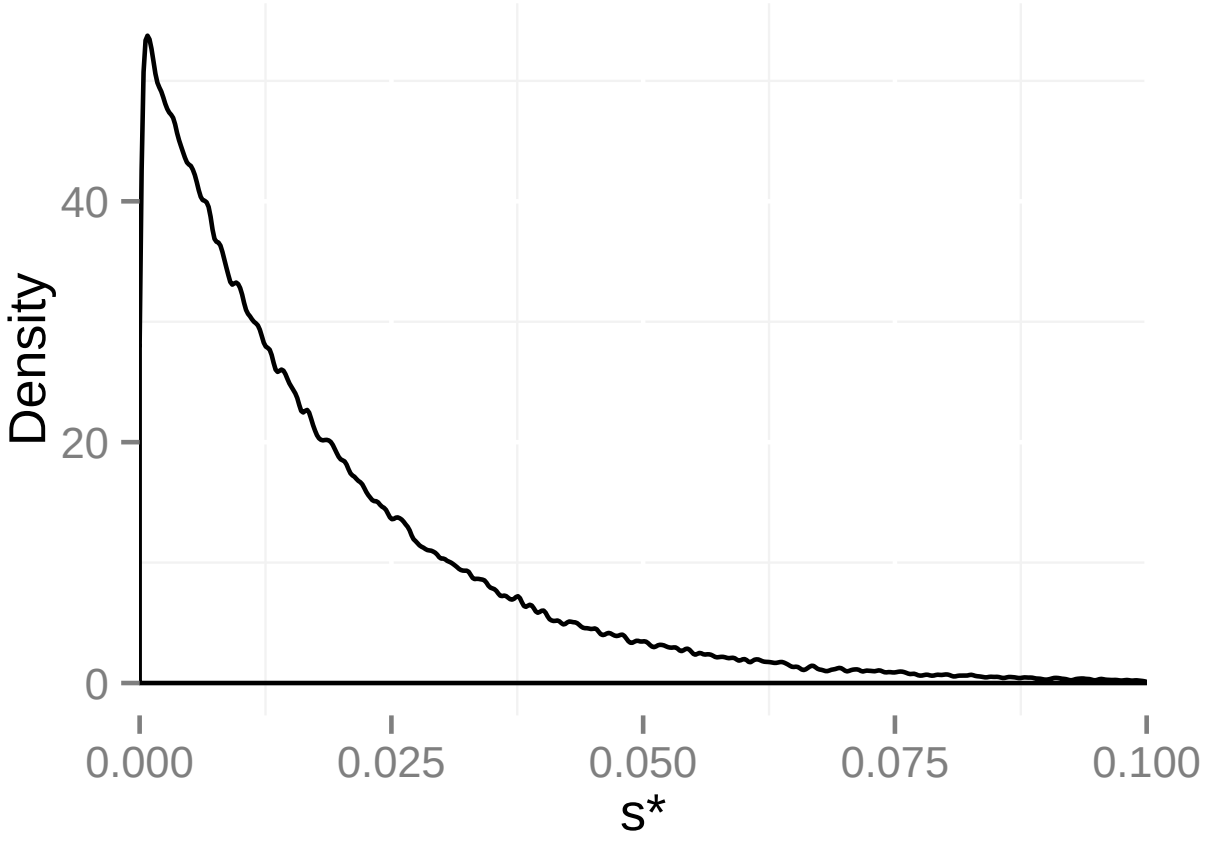


Figure S2 Average posterior density of the uniparental proportion (s^*) inferred from simulated data generated under the large sample regime ($n = 70$, $L = 32$) with a true value of $s^* = 0$. The average was taken across posterior densities for 100 data sets.

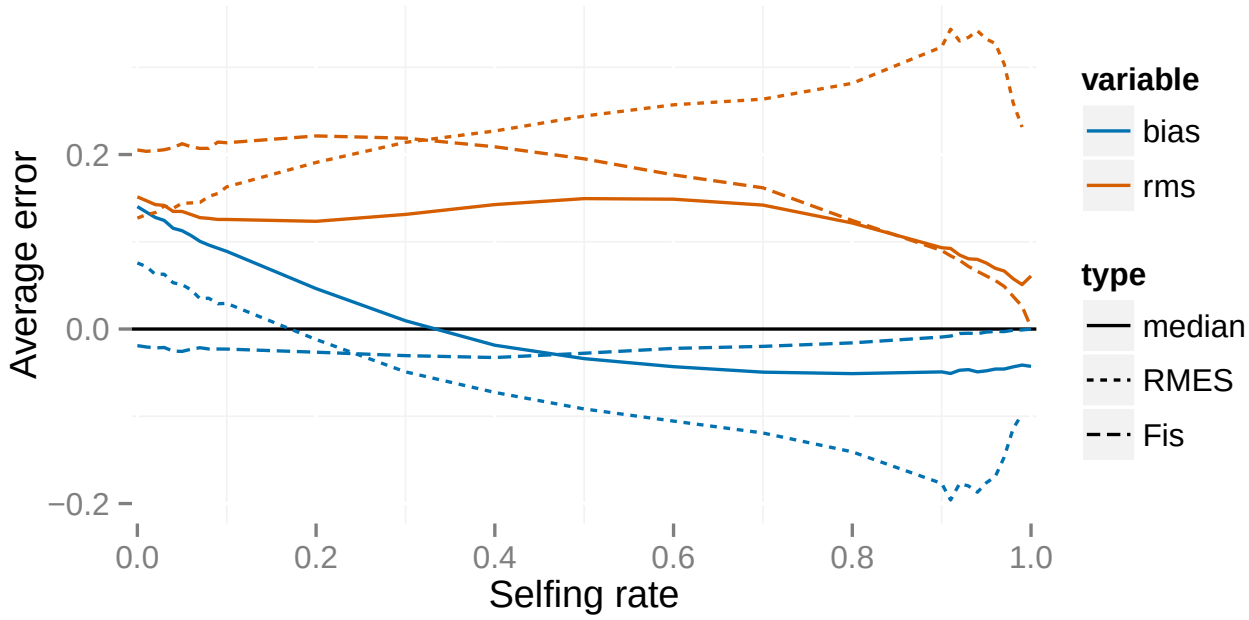
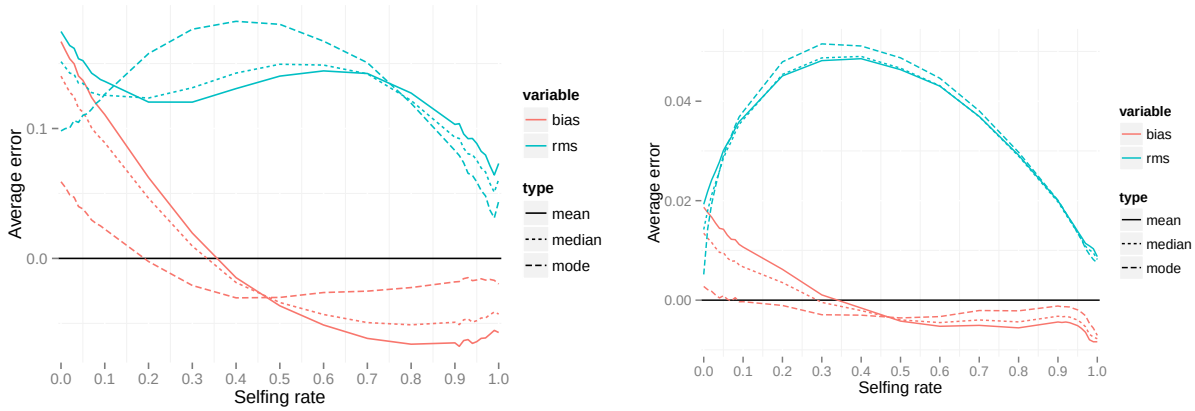


Figure S3 Errors for the full likelihood (posterior median), `RMES`, and F_{IS} methods for a small sample ($n = 10$ individuals, $L = 6$ loci). In the legend, rms indicates the root-mean-squared error and bias the average deviation. Averages are taken across simulated data sets at each true value of s^*.



(a) $n = 10$, $L = 6$ **(b)** $n = 70$, $L = 32$

Figure S4 Errors for the posterior mean, posterior median, and posterior mode. Blue curves (rms) indicate the root-mean-squared error, and red curves (bias) the average deviation. Averages are taken across simulated data sets at each true value of the selfing rate s^*.

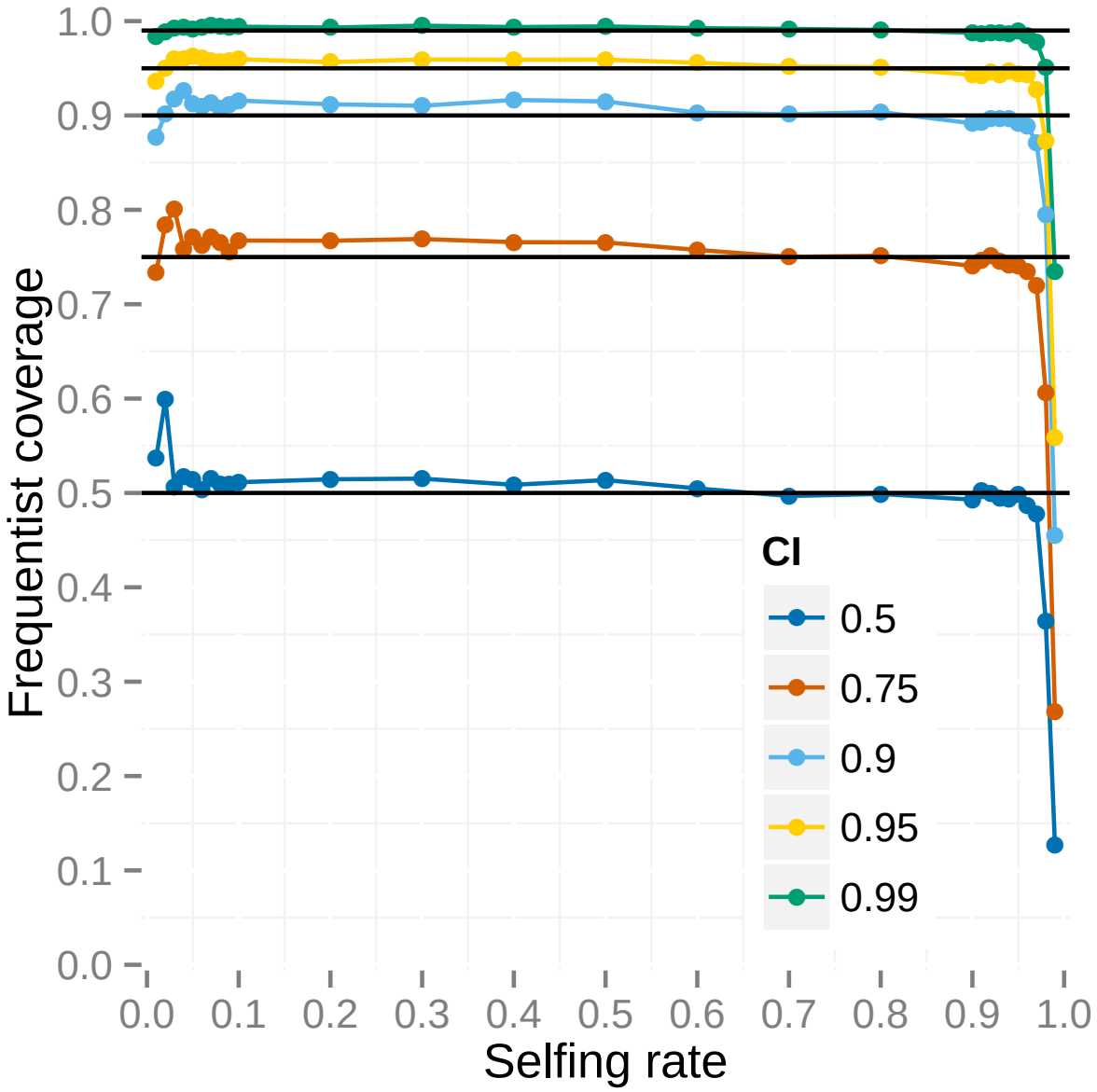


Figure S5 Frequentist coverage for Bayesian credible intervals at different levels of credibility under the large sampling regime ($n = 70, L = 32$).

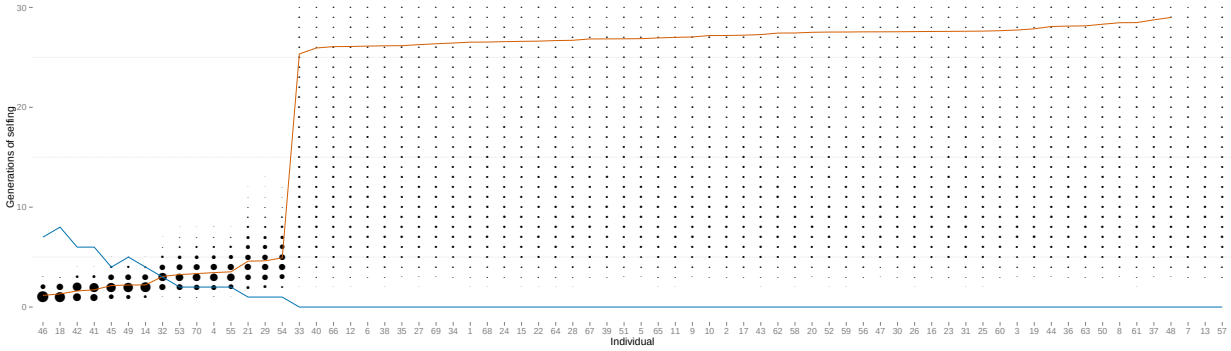


Figure S6 Number of generations since the most recent outcross event in the ancestry of each individual in the sample from the BP population of *K. marmoratus*. The area of each dot indicates the posterior probability that an individual (X-axis) has the indicated number (Y-axis) of consecutive generations of selfing in its immediate ancestry. The red line indicates the posterior mean number of selfing generations and the blue line indicates the number of heterozygous loci across individuals. The Y-axis is truncated to $[0, 30]$.

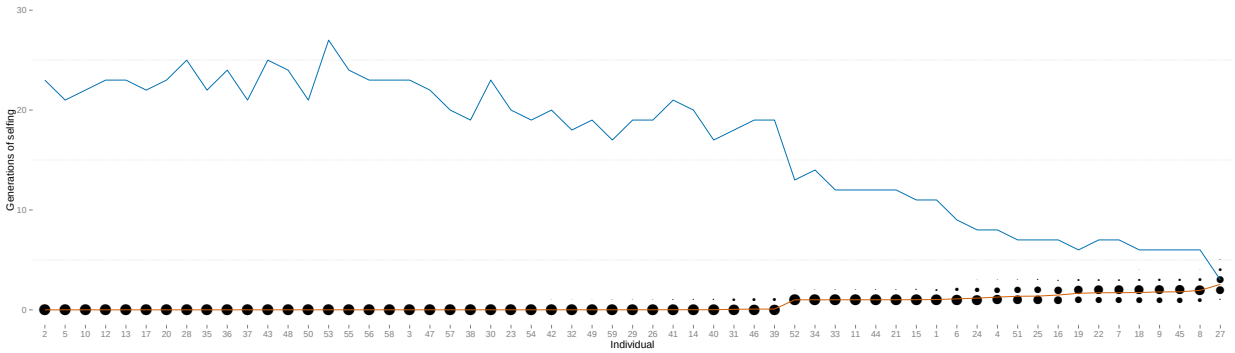


Figure S7 Number of generations since the most recent outcross event in the ancestry of each individual in the sample from the TC population of *K. marmoratus*. Symbols as in Figure S6.

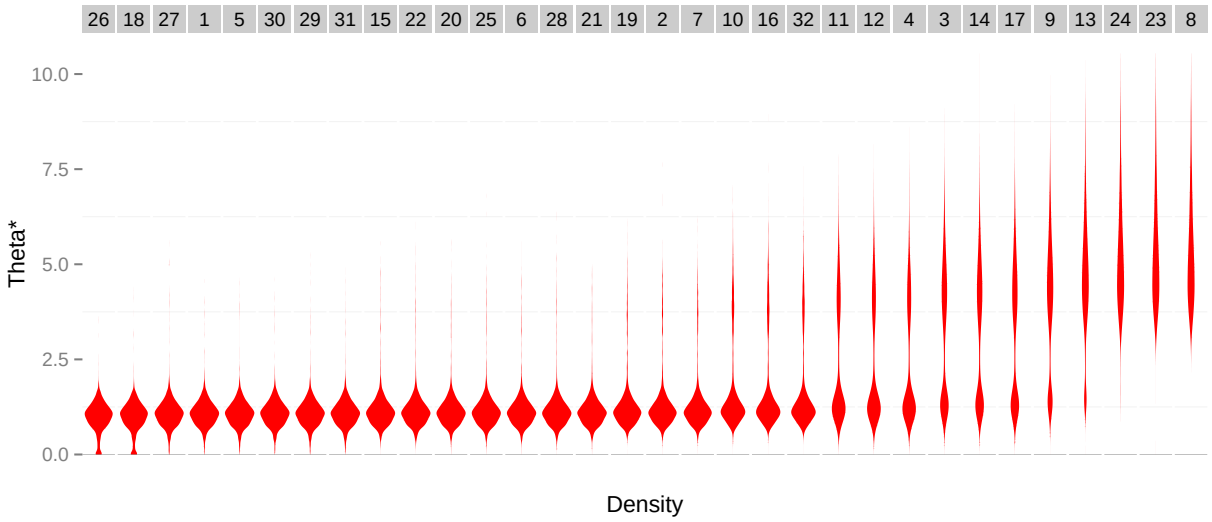


Figure S8 Posterior distributions for mutation rates at each locus in *K. marmoratus* (BP population). For each distribution. the locus name is indicated in the grey shaded box.

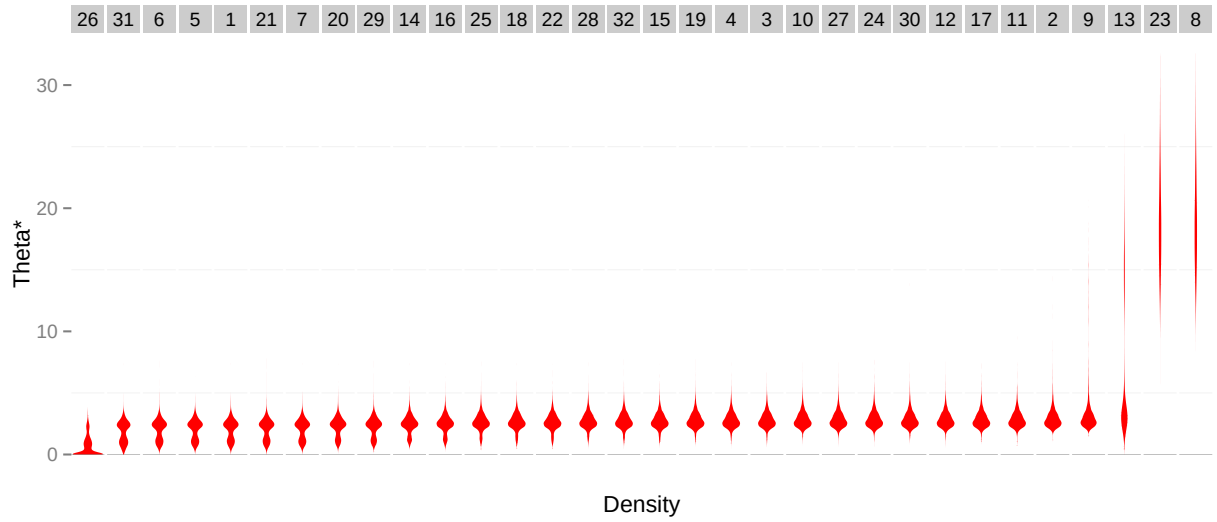


Figure S9 Mutation rates at each locus for *K. marmoratus* (TC population). For each distribution. the locus name is indicated in the grey shaded box.

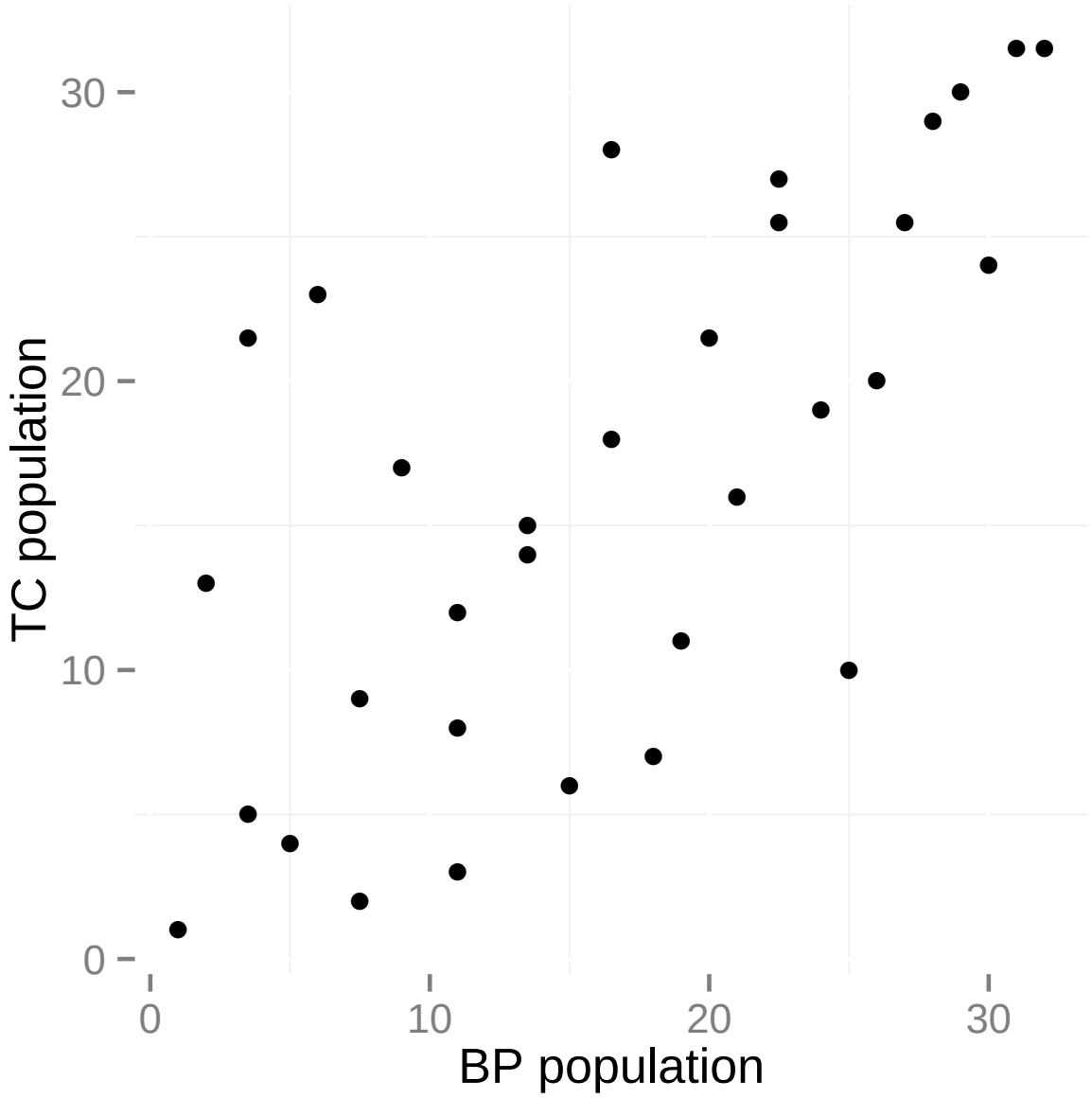


Figure S10 Comparison of rank order of estimated locus-specific mutation rates between the BP and TC populations of *K. marmoratus*. Each dot represents the rank order of the median of the mutation rate of a given locus estimated from the BP data set versus that from the TC data set.

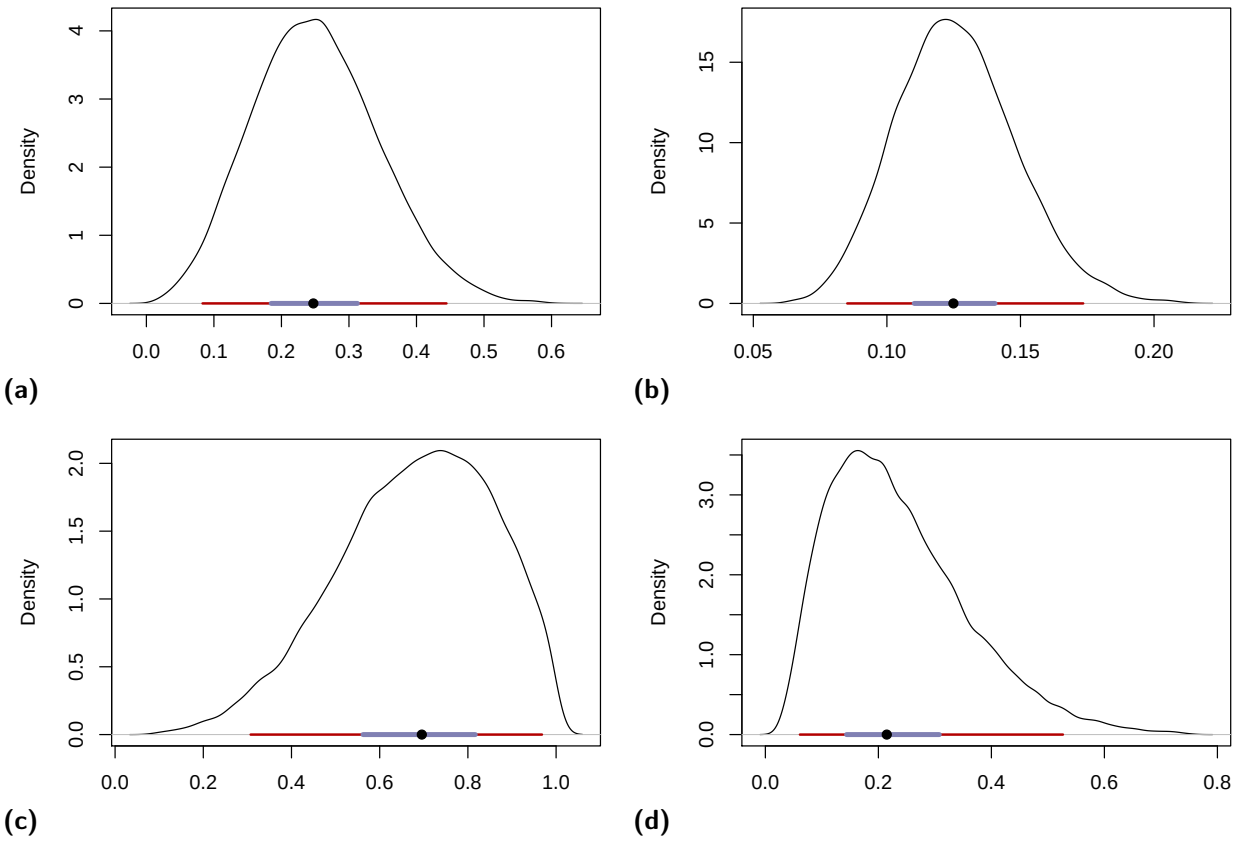


Figure S11 Posterior distributions on (a) s_G, (b) p_f, (c) $\tilde{s}$, and (d) τ for the *Schiedea salicaria* data set. Also shown are 95% BCI (maroon), 50% BCI (slate), and median (black dot).

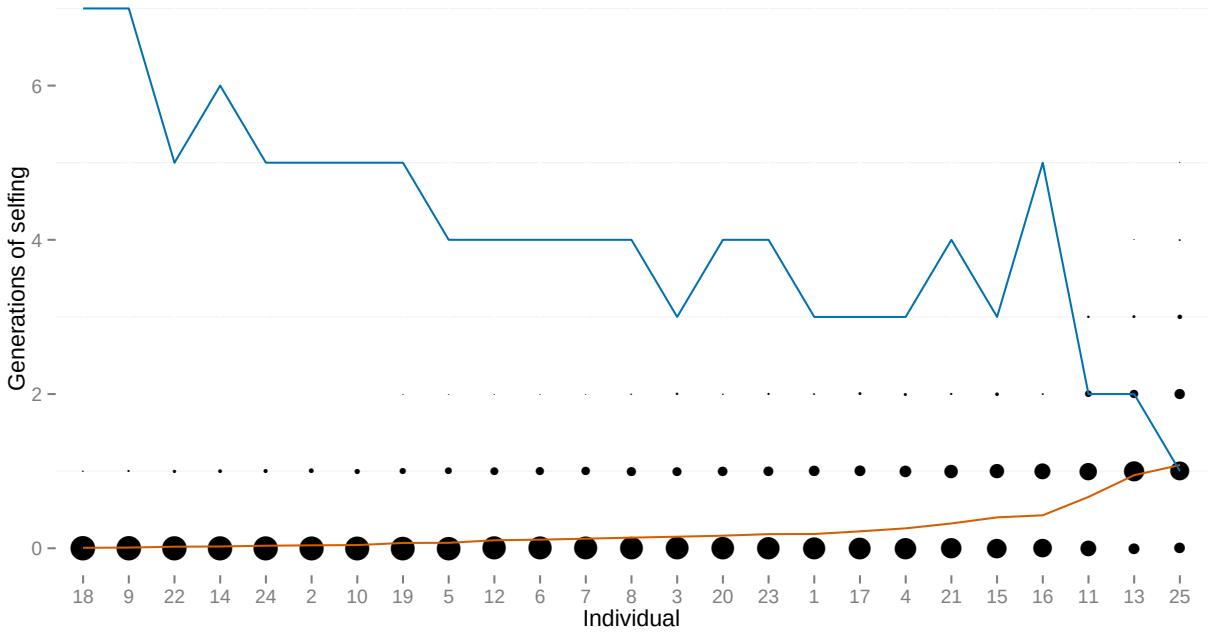


Figure S12 Estimated number of selfing generations for each individual for *S. salicaria*. The area of each dot indicates the posterior probability that a numbered individual (x-axis) has been selfed for a given number of generations (y-axis). For each individual the red line indicates the posterior mean number of selfing generations and the blue line indicates the number of heterozygous loci.

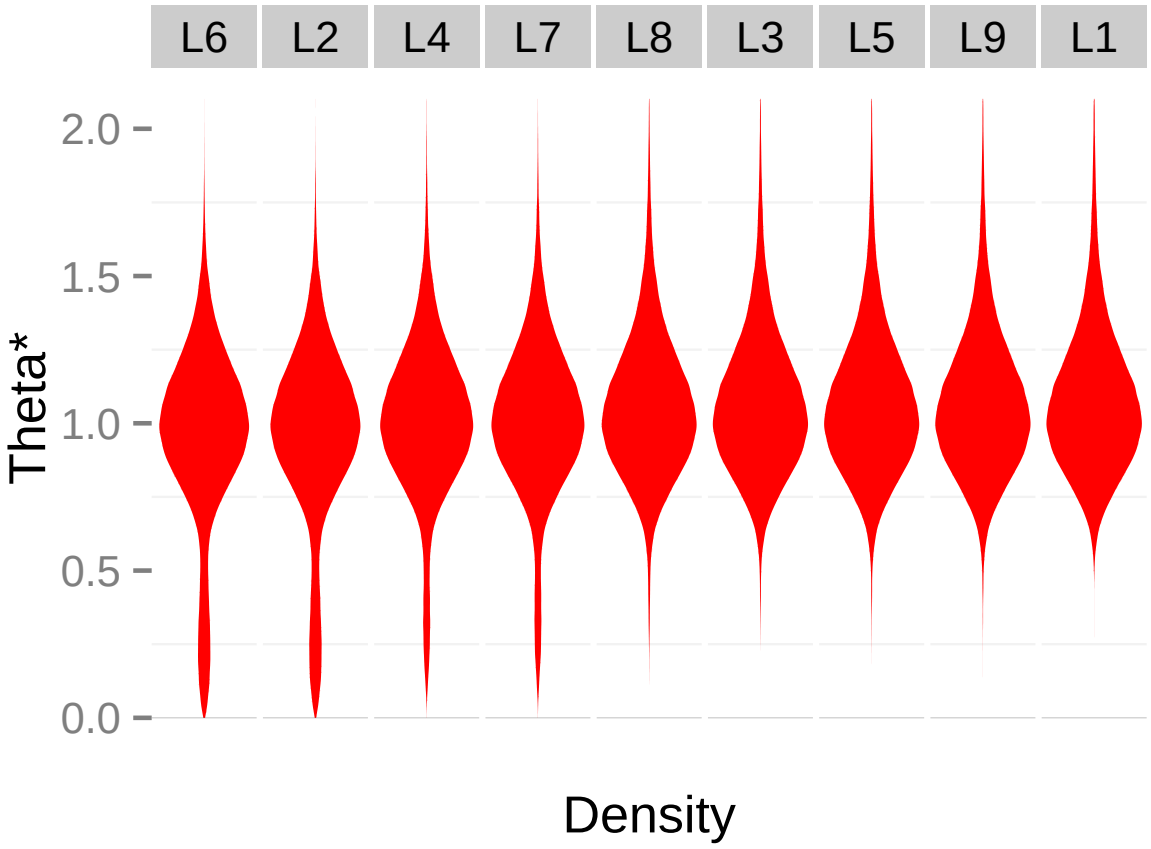


Figure S13 Posterior distributions for mutation rates at locus in *S. salicaria*. For each distribution, the locus name is indicated in the grey shaded box.