Appendix I: Simulation model and likelihood functions

Simulation model: The simulation was initiated by generating eight diploid virgin females with n_{loci} unlinked CSD loci that are all heterozygous. Each virgin female produced a single haploid genome through meiosis to obtain a son for the mother-son mating. Subsequently, diploid offspring were produced by combining the son's genome and one of both maternal genome copies (randomly sampled for each offspring). As in the experiment, all diploid offspring produced by a mother were sired by the same son. Diploid males are produced if both genome copies in a newly produced diploid offspring are identical. Diploid male survival, *s*, was implemented by comparing a random number, drawn from a uniform distribution, against *s*. We continued to generate adult diploid offspring from a single replicate until we matched the number of diploid offspring that was produced for a particular replicate in the actual experiment. Hence, while diploid family size was equal to the observed values, the number of surviving diploid males varied according to n_{loci} and *s*.

For each mother in the mother-son generation, we then generated the same number of brother-sister matings as in the experiment, unless a mother had produced only diploid sons, which is a realistic consequence of the stochasticity resulting from CSD-allele segregation with a limited number of CSD loci. In that case, no brother-sister matings were performed for that particular mother. This happened only rarely in our simulations: in the most likely case of having 100% male broods (n_{loci} =1, *s*=1), this occurred in 386 simulations out of 50,000 (0.72% of all replicates). Brother-sister matings were generated by randomly sampling a daughter from the mother's female offspring, and by generating a haploid son from that same mother. Again, a mated daughter produced the same number of adult diploid offspring as in the actual experiment.

Likelihood functions: We denoted the proportion of diploid males x produced by a particular mother k by x_k. We ran 50,000 replicate simulations of the inbreeding experiment, resulting in 50,000 simulated deviates of each data point, $\hat{x}_k$, for each set of model parameters $\mathbf{v} = \{n_{\text{loci}}, s\}$. From a histogram of these simulated deviates $\hat{x}_k$, we obtained a simulated density function $f_k(x_k|\mathbf{v})$ that informs us of the probability of the actual datapoint x_k given the current parameters. For each data point x_k, the density function f_k was obtained from the frequency histogram of the simulated deviates, which was smoothened using R's `approxfun()` method (R version 2.12.1, R Development Core Team 2011). Figure S1 shows an example of the density function f_k. The function is discrete since a female's fecundity values can only consist of integers, but nonetheless provides us with a likelihood value that reflects the simulated outcome. Hence, the likelihood function for an individual datapoint x_k is $f_k(x_k|\mathbf{v})$, and the total likelihood for the vector $\mathbf{x}$ of all datapoints resulting from the experiment is $\mathrm{L}(\mathbf{x} \mid \mathbf{v}_i) = \prod_k^m f_k(x_k \mid \mathbf{v}_i)$. The overall likelihood (taking logs and summing) is shown in Figure S2, the values of $\mathrm{L}(\mathbf{x} \mid \mathbf{v})$ are shown for varying s and $n_{\text{loci}} = \{1; 2; 3\}$.

Comparisons between different models were carried out with likelihood-ratio tests (LRTs). LRTs are conventionally used to compare nested models (i.e., situations where one of the models is a special version of the other, having additional parameters), with the null hypothesis that the data are drawn from the simpler of the two models. However, LRTs can also be applied to models that are non-nested (i.e., where one model does not have additional parameters compared to the other), as is the case in our study. To do this, we used the following approach (for details see Lewis et al. 2011): First, when comparing two non-nested models (say, model A and model B), one cannot simply assign one of both models as a null model (unless prior information is available). Instead, two reciprocal model comparisons are necessary, so that both models A and

B are considered as a null model. The observed value of the likelihood ratio test statistic $L(\mathbf{x}|\mathbf{v}_1)/L(\mathbf{x}|\mathbf{v}_0)$ (see main text) falls into one of the following categories:

1. An LRT with A as the null model is non-significant, but an LRT with model B is significant. Model A is therefore preferred over model B.
2. An LRT with B as the null model is non-significant, but an LRT with model A is significant. Model B is therefore preferred over model A.
3. Both LRTs (A as a null model, B as a null model) are significant: neither model can be considered appropriate.
4. Neither of the LRTs (A as a null model, B as a null model) are significant: no discrimination between the models is possible.

In case of a comparison between non-nested models, significance of the likelihood ratio test statistic cannot be calculated from the chi-squared distribution. Instead, we generated the appropriate test distribution from the simulations of the experiment, assuming that the null hypothesis is true. To generate the test distribution for a null hypothesis (which assumes the particular parameter values $\mathbf{v}_0$), a set of 5,000 replicates was randomly sampled from the full set of 50,000 replicate simulations for the parameter combination $\mathbf{v}_0$. Every single datapoint, $\tilde{x}_k$, within each of these sampled replicates is now used as a datapoint to calculate a likelihood ratio using the density function mentioned above, above, i.e. $\tilde{L}(\tilde{x}_k \mid \mathbf{v}_1) / \tilde{L}(\tilde{x}_k \mid \mathbf{v}_0) = \sum_k^m (\ln f_k(\tilde{x}_k \mid \mathbf{v}_i) - \ln f_k(\tilde{x}_k \mid \mathbf{v}_0)$. This step was repeated for all 5,000 sampled simulations, resulting in a distribution of 5,000 likelihood ratio test values that were then used for null hypothesis testing, summarized in Table 3. An example of a distribution $q\left(\tilde{L}(\tilde{x}_k \mid \mathbf{v}_1) / \tilde{L}(\tilde{x}_k \mid \mathbf{v}_0)\right)$ of likelihood ratio test values, in comparison to the actual likelihood ratio is given in Figure S3.

Although significance values are not corrected for multiple comparisons, a Bonferroni correction by multiplying significance values by $1/n=1/6$ does not alter our conclusions.