The maintenance of self-incompatibility: overdominance of inbreeding depression and S-linked genetic load

Daniel J. Schoen and Sarah J. Baldwin

Department of Biology, McGill University, Montreal, QC H3A 1B1 Canada

Author for correspondence:
Daniel J. Schoen
E-mail: dan.schoen@mcgill.ca

Word Count:
Summary: 159
Introduction: 567
Methods: 1760
Results: 562
Discussion: 1461
Total: 4509
Tables: 2
Figures: 4
Supplementary Information: Table S1, Table S2, Notes (2)

Summary

Inbreeding depression plays a fundamental role in evolution. To help detect and characterize viability loci that underlie inbreeding depression, plants from self-incompatible populations of *Leavenworthia alabamica* were force self-pollinated to produce families of progeny that were genotyped at hundreds of single nucleotide polymorphism (SNP) loci.

We used Bayesian analysis of segregation data for each SNP to explore support for different dominance and selection coefficients at linked viability loci.

There was strong support for overdominance (or pseudo-overdominance) at many viability loci. There was some support for recessivity and underdominance. One recessive viability locus mapped to the genomic region of the novel self-incompatibility locus in *Leavenworthia alabamica.*

The results are consistent with earlier findings showing that inbreeding depression is recalcitrant to purging in *Leavenworthia alabamica*. The results also help account for the maintenance of self-incompatibility in this species and are consistent with expectations from evolutionary genetic theory that recessive, deleterious alleles linked to loci under balancing selection are sheltered from selection.

Keywords: Viability loci, dominance coefficient, selection coefficient, segregation vartiom Bayesian analysis, sheltered genetic load, mating system evolution

Introduction:

Inbreeding depression is the fitness reduction in progeny produced by consanguineous mating versus mating between unrelated parents (Wright, 1977). Inbreeding depression plays a fundamental role in crop production, conservation of natural populations, and mating system evolution (Charlesworth & Willis, 2009; Frankham, 2010; Lloyd, 1979; Lande & Schemske, 1985). While recessive deleterious mutations are thought to play a major role in inbreeding depression, overdominance has also been implicated (Wright, 1977; Charlesworth & Willis, 2009).

Lande & Schemske (1985) considered the evolutionary consequences of inbreeding depression and the mating system. Their theoretical work predicted when there is selective purging of recessive lethal and near-lethal alleles, populations of self-fertilizing organisms should exhibit reduced equilibrium levels of inbreeding. Charlesworth & Charlesworth (1987) extended the study of purging of inbreeding depression to many deleterious alleles of milder effect with a broader range of dominance and found that fairly high levels of inbreeding depression can be maintained in populations of self-fertilizing organisms. For species with self-incompatibility (SI), theory predicts that an additional source of inbreeding depression is due to the linkage of deleterious recessive alleles to self-recognition loci (Uyenoyama, 1997). In gametic self-incompatibility systems, the self-recognition locus (S-locus) region is maintained as heterozygous, thereby hindering the selective purging of recessive deleterious mutations, and leading to "sheltered genetic load" in this region. Likewise, in sporophytic self-incompatibility systems, depending on the dominance levels of the S-allele, the S-locus region is expected to be maintained as heterozygous and harbour sheltered S-linked load (Schierup *et al*., 1997; Llaurens *et al*., 2009). Theory has shown that sheltered load may affect the trajectory of mating system evolution (Porcher & Lande, 2005; Gervais *et al*., 2014).

Few studies have provided information on the number and genomic locations of the viability loci responsible for inbreeding depression, and there is little information on the selection and dominance coefficients of the viability effects, factors important to predicting the evolutionary trajectory of inbreeding depression. For SI species, there is even less information of this kind (though see Karkkainnen *et al*., 1999).

The genus *Leavenworthia* (Brassicaceae) is a group of self-incompatible species that has received considerable attention by evolutionary biologists interested in plant mating system

evolution due to both inter- and intraspecific variation in the mating system (Lloyd, 1965; Busch, 2005; Busch *et al*., 2011; Herman *et al*., 2012; Chantha *et al*., 2013). Self-incompatible species in this genus exhibit significant levels of inbreeding depression (Charlesworth *et al.*, 1994; Busch, 2005). Recently, Baldwin & Schoen (2018) conducted experiments in which plants from self-incompatible populations of *L. alabamica* were subjected to several generations of forced selfing. They found that inbreeding depression expressed in these experiments is resistant to purging.

Marker-based approaches to studying the genetic architecture of inbreeding depression can provide information on the dominance and selective coefficients at individual viability loci (Hedrick & Muona, 1990; Fu & Ritland, 1994a,b), which in turn can help shed light on how inbreeding depression evolves. When available, such information can be coupled with recombination and physical maps to reveal the approximate genomic location of viability loci (Mitchell-Olds, 1995; Remington & O'Malley, 2000; Karkkainnen *et al*., 1999). Here we extract information on the genetic architecture of inbreeding depression in *Leavenworthia alabamica* using a genetic marker-based approach. We discuss the relevance of our results to mating system evolution and to previous findings that have used the marker-based approaches to examine the genetic architecture of inbreeding depression in plants.

Methods

Plant growth and production of self-fertilized progeny

Two self-incompatible *Leavenworthia alabamica* plants from separate "a1 race" populations in Alabama (Lloyd, 1965) were used in this work. These plants were grown in separate 9" pots in the McGill University Phytotron and forced to self-fertilize by bud or salt pollinations as described in Chantha *et al*. (2013). Up to 30 pollinations were carried out per plant two times per week. All seeds were collected at maturity, stratified in a dry environment for at least two months, and germinated. The progeny from these two separate seed families (Families 1 and 2, hereafter) were grown to the flowering stage.

Genotyping-by-sequencing (GBS)

We collected young leaf tissue samples from the progeny of flowering stage plants and stored them at -80°C. Qiagen plant mini-DNA kits were used to extract DNA from the leaf

111 tissue. Extractions were repeated if concentrations were lower than 20 ng/ul and were diluted to
112 30-50 ng/ul if over 50 ng/ul.
113 We used the genotyping-by-sequencing (GBS) (Elshire *et al*., 2011) method to produce
114 fragments for genotyping using the protocol of Wallace & Mitchell (2017). The standard protocol
115 was used to generate libraries with 1.44 ng PstI per 100 ng DNA and 15-18 cycles of PCR.
116 Fragments were then sequenced on one lane of an Illumina High-Seq run (Family 1, N = 91 and
117 Family 2, N = 90). FASTQC was used to evaluate the basic quality of the sequencing runs.
118 Reads were demultiplexed using Process Radtags in STACKS v. 2.59. The *Leavenworthia*
119 *alabamica* reference assembly version 2.5 (Haudry *et al.*, 2013, CoGe platform: Genome ID
120 20247) was indexed and sorted and aligned to the reference genome with the Burrows-Wheeler
121 Aligner. The average read depth across the loci was 276x for Family 1, 277x for Family 2. SNP
122 locus segregation ratios derived from the two families by producing a catalog of segregating loci
123 with the GSTACKS and POPULATIONS routines in STACKS v. 2.59 (Catchen *et al*., 2013).
124 In-house scripts for the processing steps and associated information are available in the
125 Supplementary Materials section (Table S1).
126
127 Treatment of potential genotyping errors
128 Genomic analysis of *Leavenworthia alabamica* indicates that this species underwent
129 whole genome triplication during its evolutionary history, and chromosome painting with
130 *Arabidopsis thaliana* BAC probes suggests that that some triplicated genome segments were
131 retained as duplicates or triplicates (Haudry *et al*., 2013). The reference genome assembly in this
132 species is only partially complete (N50 = 309 kb); while the haploid chromosome number of *L.*
133 *alabamica* is 11, the genome assembly consists of 3,321 scaffolds. It comprises *ca.* 174 million
134 bp in total of the expected 315 million bp genome size, the latter estimated from flow cytometry
135 (Brian Husband, pers. comm.). Over half of the scaffolds are short, with lengths 5,000 bp or
136 smaller, though larger ones are > 1M bp. Due to sequence divergence of the duplicated (or
137 triplicated) regions, longer scaffolds in the reference genome likely represent separate genomic
138 regions (Haudry *et al*., 2013). But many shorter duplicated (or triplicated) regions of the genome
139 with extended identity have likely been collapsed into single (artifactual) contigs. GBS products
140 from separate genomic regions that map to such short, artifactual contigs will yield only (or
141 mostly) heterozygote genotype calls. To minimize these artifacts we: (1) restricted our analysis

142 to SNPs that map to scaffolds > 100 kbp in length (these account for 73% of the *L. alabamica*
143 assembly); (2) rejected SNP loci that yielded only heterozygotes if adjacent SNP loci on the
144 same scaffold exhibited segregation ratios that were not significantly different from 1:2:1 by
145 Chi-square tests, on the basis that evidence of strong selection should be shared by adjacent SNP
146 loci; and (3) restricted our analysis to loci where data were available from > 90% of the progeny
147 of a cross since genotyping errors are more likely to have occurred when few progeny per family
148 are genotyped successfully.
149
150 Linkage map construction and tests for departure from normal segregation
151 Linkage groups (each containing a minimum of three SNP loci) were constructed with
152 the software package 'qtl' version 1.50 (Broman *et al*., 2003) running in R (version 4.1.1).
153 Separate linkage maps were created from the Family 1 and 2 data. These maps allowed us to
154 examine segregation ratios of each SNP in relation to that of other SNPs. To arrive at a
155 conservative estimate of the number of viability loci, we limited our analysis to those SNPs that
156 exhibited departures from the expected (null) 1:2:1 segregation ratio at a false discovery rate
157 (Benjamini & Hochberg, 1995) of $q = 0.05$. Moreover, a viability locus located adjacent to a
158 given SNP locus will likely affect not only the segregation ratio at that SNP but also at other
159 nearby SNPs, and so within each 50 cM window if several such SNPs were found, we selected
160 the one exhibiting the most extreme departure from 1:2:1 for analysis of the selection and
161 dominance coefficients at the linked viability locus.
162
163 Selection and dominance coefficients at linked viability loci
164 The analysis of marker locus segregation ratios to estimate selection and dominance at
165 viability loci linked to marker loci assumes that departure from the expected 1:2:1 Mendelian
166 segregation ratio is due to viability selection, rather than to meiotic drive and/or gametic selection.
167 The parent plant that is self-fertilized is assumed to be heterozygous at a selectively neutral marker
168 locus (M) and a linked viability locus (V)—e.g., the parent has the genotype M_1V_1/M_2V_2, and the
169 genotypes V_1/V_1, V_1/V_2, and V_2/V_2 are associated with the fitness values 1-s, 1-hs, and 1,
170 respectively, where s is the selective coefficient or the proportional amount by which selection at
171 this locus reduces viability in the homozygote V_1/V_1 ($0 \leq s \leq 1$), and h is the dominance coefficient.
172 A value of $h < 0$ implies overdominance, $h = 0$ implies fully recessive, $0 \leq h \leq 0.5$ implies partially

recessive, $h = 0.5$ implies additive, $0.5 \leq h \leq 1.0$ implies partial dominance, and $h > 1$ implies underdominance. The marker and viability loci are assumed to be linked with recombination fraction c. A Bayesian statistical genetic approach was developed to explore the posterior support for different values of h and s at viability loci based on the observed numbers of marker locus genotypes.

Fu & Ritland (1994a) showed that following self-fertilization, the post-selection genotype frequencies expected for three marker genotypes (p_{11} for M_1/M_1, p_{12} for M_1/M_2, and p_{22} for M_2/M_2) are:

$$p_{11} = \frac{c^2\, s(2h-1) - 2\, cs(h-1) + (1-s)}{4 - s(2h+1)}$$

$$p_{12} = \frac{2(-c^2\, s(2h-1) + 2cs\,(2h-1) - hs + 1)}{4 - s(2h+1)} \qquad (1)$$

$$p_{22} = \frac{c^2\, s(2h-1) - 2\, csh + 1}{4 - s(2h+1)}$$

In these equations p_{11} denotes the frequency of the least common homozygote genotype. Note that when $s = 0$, the marker genotype frequencies give the null 1:2:1 ratio.

The data are the numbers of each of the three marker genotypes observed among the surviving progeny of the self-fertilized parent (N_{11} = number of M_1/M_1, number of observed, N_{12} = number of M_1/M_2 observed, and N_{22} = number of M_2/M_2 observed). The probability (likelihood) given the data can be written:

$$Pr\,(N_{11}, N_{12}, N_{22} \mid s, h, c) = \left(\frac{N!}{N_{11}!\, N_{12}!\, N_{22}!}\right) p_{11}{}^{N_{11}} p_{12}{}^{N_{12}}\, p_{22}{}^{N_{22}} \qquad (2)$$

where N is the sum of N_{11}, N_{12}, and N_{22}. The joint density and posterior probability are, respectively:

$$Pr(s, h, c, N_{11}, N_{12}, N_{22}) = Pr(s)\, Pr(h)\, Pr(c)\;\; Pr(N11, N12, N22 \mid s, h, c) \qquad (3)$$

$$Pr(s,h,c\,|N_{11}N_{12}N_{22}) = \frac{Pr(s)\,Pr(h)\,Pr(c)\,Pr(N11,N12,N22\,|\,s,h,c)}{\int_{s,h,c} Pr(s)\,Pr(h)\,Pr(c)Pr(N11,N12,N22\,|\,s,h,c)\;ds\,dh\,dc} \quad (4)$$

To explore values of the parameters using a Bayesian approach, we chose uniform priors with $0 < s < 1$, $-100 < h < 100$, and $0 < c < 0.2$. The broad prior for h allows for the possibility of extreme overdominance and underdominance. The prior for c (less than the theoretical upper limit of $c = 0.5$) is justified based on the average distance between markers of *ca.* 2 cM in this study. A program written in the R statistical language (version 4.1.1) to conduct Markov Chain Monte Carlo method with the Metropolis-Hastings sampler (Weir, 1996) (Supplementary Materials: Note 1). This was used to "integrate out" the nuisance parameter c, allowing the posterior probabilities to be estimated for different values of s and h. For the required burn-in period, we ran the sampler 5.5 million times before examining the posterior probability distributions. Tracking the parameter values during the runs showed that the sampler had stabilized by the end of this burn-in period.

The S-locus region of *Leavenworthia alabamica* and locations of viability loci

The *Leavenworthia alabamica* S-locus region has previously been cloned and sequenced from the same plant material that was used to produce the sequence information in the genome assembly in the present work (GenBank accession KC981242.1) (Chantha *et al*., 2013). This allowed us to locate the individual *L. alabamica* genome assembly scaffold containing the S-locus. We used the BLAST algorithm (Altschul *et al.,* 1990) in the software package Geneious Prime (version 2.2) to search of *L. alabamica* reference genome version 2.5 using both *L. alabamica* S-locus genes (*Lal2* and *LaSCRL*; Chantha *et al.*, 2013). We estimated selection and dominance coefficients for all potential viability loci linked to SNPs in S-locus region that showed significant departures from 1:2:1.

Statistical power test and different modes of gene action

The analysis of the number and mode of action viability loci hinges on the ability to detect departure from the expected 1:2:1 ratio obtained when self-fertilizing a plant with

heterozygous marker loci. Given that the sample size of progeny per family in our work was limited to a maximum of 91 progeny (the maximum number we could obtain by forced selfing of each of the SI parent plants used), we were interested in the statistical power of our study to detect selection and characterize the mode of gene action. Specifically, we did this by producing expected segregation ratios for a range of different values of the parameters in equation (1) followed by standard methods (Cohen, 1988) for examining statistical power, as implemented in the R program package 'pwr' version 1.3.

Results

SNP segregation ratios, linkage map locations, and significance testing

We found 469 segregating SNP loci in Family 1, and 378 segregating SNP loci in Family 2. In Family 1 we detected 20 linkage groups, ranging in size from 5 to 182 cM, for a total map length of 1,066 cM. In Family 2 we found18 linkage groups, ranging in size from 2 to 109 cM, for a total map length of 817 cM (see Supplementary Table S2 for linkage group, scaffold positions, and segregation patterns of SNPs).

SNP segregation patterns along with and their significance levels based on the Chi-square goodness-of-fit to the hypothetical 1:2:1 segregation ratio are shown in Figures 1 and 2. With a false discovery rate of $q = 0.05$, we detected eight putative viability loci in Family 1, and eight more in Family 2 (Table 1).

Selection and dominance of viability loci linked to SNPs

The 16 putative viability loci detected were further characterized for their dominance and selection coefficients using the Bayesian approach outlined above. For 12 out of 16 of SNPs, the posterior probability mass for inferred linked viability loci indicated overdominant gene action, with Bayesian estimates of h ranging from -17.47 to -1.07, and s ranging from 0.32 to 0.96 (Table 1, Fig. 3). Three SNPs showed evidence of linkage to viability loci exhibiting underdominance, with Bayesian estimates of h ranging 1.30 to 1.48, and s ranging from 0.51 to 0.68 (Table 1, Figs. 3). The remaining locus showed evidence of recessive gene action with Bayesian estimates $h = 0.45$ and of $s = 0.86$ (Table 1, Fig. 3).

260 The S-locus region of *Leavenworthia alabamica* and locations of viability loci

261 The tightly linked S-locus genes *Lal2* and *SCRL* along with their intervening non-coding
262 region are located on scaffold 46, at base pair positions 2,035,545 to 2,044,046. Several genomic
263 regions on this scaffold in Family 1 showed a strong deviation from the 1:2:1 ratio (Fig. 2 and
264 Table 2). The segregation data here support recessive gene action and strong selection $h = 0.18$ to
265 0.45, and $s = 0.80$ to 0.86 (Table 2). In Family 2, however, this region did not harbour any
266 detectable viability loci.
267
268 Statistical power to detect different modes of gene action

269 Statistical power to detect different modes of gene action in experiments of the type
270 conducted here depends on the value of the recombination fraction c. For close linkage, e.g., c
271 near 0 as is likely for most SNPs in this study (Figs. 1 and 2, Table S2), overdominant and
272 underdominant action are easier to detect than dominant and recessive action (Fig.4). As the
273 recombination fraction between marker and inbreeding loci increases, statistical power to detect
274 underdominant selection increases, whereas statistical power to detect overdominant gene action
275 decreases (Fig. 4). For dominant and recessive gene actions, statistical power of detection is
276 reduced with increasing c, but compared with overdominant and underdominant action the
277 changes are less affected by recombination fraction. Given a sample size of 90 progeny (the
278 modal value in this study), close linkage, and overdominant or underdominant genes actions,
279 selection coefficients of $ca.\ s = 0.3$ at viability loci are detectable with statistical power of 80%.
280 For dominant and recessive actions, viability loci with selection coefficients of $ca.\ s = 0.6$ are
281 detectable with statistical power of 80%.
282
283 **Discussion**
284
285 The genetic architecture of inbreeding depression.

286 We detected significant departures from the 1:2:1 segregation ratio at SNPs in the
287 progeny of parental plants from two SI populations across many regions in the *Leavenworthia*
288 *alabamica* genome. The segregation data were used to explore support for different values for
289 selection and dominance coefficients associated with linked viability loci. The approach we used
290 assumes that meiotic drive, gametic selection, and epistasis do not contribute to departures from

291 normal segregation. We cannot distinguish the effect of these factors from the action of viability
292 selection acting early on embryo and seed development, which is a limitation of this study.
293 Nevertheless, there are several reasons to believe that most departures from the 1:2:1 segregation
294 ratio observed in this study are associated with inbreeding depression rather than with meiotic
295 drive or gametic selection. First, where evidence of meiotic drive has been studied in detail (e.g.,
296 in *Arabidopsis thaliana*), the effects were generally found in specific genomic regions, rather
297 than genome wide as observed in the present study (Salomé *et al.*, 2012; Seymour *et al.*, 2019).
298 Second, gametic selection favouring a specific allele linked to markers would not produce a
299 heterozyote excess at linked marker loci, and so would not be interpreted as overdominance
300 acting at viability loci as we observed in most cases.
301 Several past studies have used marker data to explore the genetic architecture of
302 inbreeding depression in plants. Hedrick & Muona (1990) used the marker-based approach in
303 *Pinus sylvestris* and detected a single viability locus with a strong contribution to the fitness of
304 progeny. When we applied the Bayesian approach outlined above to their data we found
305 overdominant action at this locus. In another study, using progeny from self-pollinated plants
306 collected from a wild *Mimulus guttatus* population, Fu & Ritland (1994b) applied a graphical
307 approach that divides the space of possible segregation ratios into different modes of gene action.
308 They detected 24 viability loci, with the majority showing underdominance or dominance
309 effects. Mitchell-Olds (1995), studying inbreeding depression in the progeny of parent of an
310 *Arabidopsis thaliana* F1 hybrid between different accessions, reported a single viability locus
311 with overdominance that mapped to chromosome 1. Remington and O'Malley (2000) examined
312 progeny segregation ratios at AFLP loci in selfed *Pinus taeda* plants and detected both recessive
313 and overdominant action at 19 embryonic viability loci. In the only other study to apply a
314 Bayesian approach to estimating selection and dominance at loci associated with inbreeding
315 depression, albeit with a different parameterisation that the one employed here, Kärkkäinen *et al*.
316 (1999) in *Arabidopsis lyrata* ssp. *petraea* reported linkages of viability loci to seven separate
317 allozyme loci, mostly with overdominant effects. Finally, Hedrick *et al.* (2016) examined
318 segregation ratios for nearly 10,000 SNPs in 28 progeny descended from a self-polliinated plant
319 of *Eucalytus grandis*, and found widespread heterozygote excess, consistent with
320 overdominance, though their simulation analysis suggested that pseudo-overdominance is the

more likely explanation. Overall, these empirical studies suggest that modes of gene action for linked viability loci may vary from species to species.

It is notable that a few of the above-cited studies have reported underdominant viability loci linked to markers (e.g., especially Fu & Ritland, 1994b, but also Kärkkäinen *et al*. 1999). Underdominance was also inferred at two marker-linked viability loci in the present investigation. Underdominance at viability loci cannot contribute to inbreeding depression. We note from the results of the power analyses we conducted, underdominance is detectable with smaller sample sizes compared with other modes of action that require larger sample sizes to detect selection of the same magnitude when $c < 0.3$ (Fig. 4). This could help account for the unexpected reports of underdominance in studies that have employed few markers (e.g., Fu & Ritland, 1994b).

Overdominance or pseudo-overdominance of inbreeding depression

For 12 of the 16 SNPs showing departures from normal segregation, our analysis suggests that overdominance at linked viability loci contributes to inbreeding depression. Alternatively, these results could be explained as being due to pseudo-overdominance (Ohta, 1971). Psuedo-overdominance would be detected by the methods used here if deleterious, recessive viability mutations are linked and in repulsion. In general, compared with recessive or partly recessive action at viability loci, there is little support from past studies for true overdominance as an underlying cause of inbreeding depression (Charlesworth and Willis 2009). Distinguishing between true overdominance and pseudo-overdominance is difficult when there is a single family of progeny, but becomes progressively easier with multiple families, as in the present study. True overdominance, if present, should often be detected at the same viability loci in the progeny of multiple selfed parents, since selection there should render it frequent. In contrast, viability loci exhibiting pseudo-overdominance arising from deleterious recessive or partly recessive alleles are less likely to be shared among parents, as the genomic regions are maintained by the interaction of mutation and selection. Our study was limited to two separate families, but we did, in fact, infer overdominant viability loci in the same two genomic regions in both Families 1 and 2 (SNPs 68_160676 and SNPs 68_16082; and SNP 72_1428232 in both families). No other such cases were detected (Table S2). This leaves open the possibility that both true and pseudo-overdominance are present.

352

353 Sheltered S-linked load

354 An inferred viability locus (or possibly, multiple linked loci) with recessive action was
355 detected only in one genomic region, and one family (Family 1). Interestingly, the region is
356 within 80 kb from the site of the S-locus. These results can be interpreted as evidence that there
357 is sheltering from selection of one (or more) recessive, deleterious mutations in the S-locus
358 region (Table 2).

359 In Family 2, no such effects were found in the S-locus region. The finding is not
360 necessarily at odds with the evidence of sheltered load in the Family 1 data. There are likely
361 several levels of dominance for S-locus alleles in *L. alabamica* (Busch *et al.*, 2010), and in
362 sporophytic systems such as this one, the strongest sheltering from selection is expected for those
363 viability loci linked to S-alleles that exhibit the highest degree of dominance (Llaurens *et al.*,
364 2009).

365

366 Relationship of results to earlier findings on inbreeding depression and the evolution of the
367 mating system in *Leavenworthia alabamica*

368 Busch (2005) discovered significant levels of inbreeding depression in self-incompatible
369 populations of *Leavenworthia alabamica* for germination and survival rate, biomass, flower
370 number, and pollen viability. He estimated that inbreeding depression for total lifetime fitness
371 either as a male or female parent exceeds 0.5. Baldwin & Schoen (2019) studied inbreeding
372 depression for a similar set of fitness components in *L. alabamica* and again found inbreeding
373 depression for lifetime fitness exceeds 0.5. Baldwin & Schoen extended their study to include
374 three consecutive generations of self-pollination and found no evidence that inbreeding
375 depression was purged by this treatment. These results are consistent with the findings of the
376 present investigation, where overdominance (or more likely, pseudo-overdominance) is
377 suggested for many linked viability loci.

378 Inbreeding depression is thought to be the main barrier to the evolution of self-pollination
379 (Fisher, 1941; Lloyd 1979), and yet with some self-pollination, purging of recessive load is
380 expected. In an earlier set of population surveys Baldwin & Schoen (2018) that some level of
381 selfing likely does occur in populations of *Leavenworthia alabamica* that have been formerly
382 characterized at SI. Specifically, Baldwin & Schoen (2018) discovered that between 6% and

18% of plants exhibit "leaky SI", and thus are capable of setting seed with self-pollen. They also showed that variation in the degree of this leakiness is heritable, and so at least one requirement for purging inbreeding depression (i.e., self-pollination) is likely present in populations of this species. On the other hand, theory suggests that the spread of mutations that reduce SI occurs mainly when inbreeding depression has a strong recessive basis (Porcher & Lande, 2005; Gervais *et al*., 2014). The results from the present study suggest this is not the case *Leavenworthia alabamica*. That is, apart from one viability locus (or several closely linked viability loci) that are linked to the S-locus and therefore at least partially shielded from selection, our results do not support such an underlying recessive genetic architecture for inbreeding depression but rather are consistent with an overdominant and/or pseudo-overdominant basis.

Recently it has been suggested that a pseudo-overdominance may why inbreeding depression is recalcitrant to purging (Waller, 2021). If this is true in the case of *Leavenworthia alabamica,* it may help to explain why most populations retain SI, despite the fact that SI populations exhibit variation for the strength of the trait. The results here are also consistent with the observation that the few SC populations of the *L. alabamica* that have been observed are small or on the margins of the range, conditions where drift may have played a role in reducing inbreeding depression (Glémin, 2003; Busch, 2005).

Acknowledgements

We thank Mark Romer and Claire Cooney for assistance with plant growth in the McGill Phytotron. Yong-Bi Fu provided comments on the manuscript. SJB acknowledges scholarship support from the Fonds de Recherche du Québec – Nature et technologies. This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to DJS

Author Contributions

DJS conceived the study and developed the Bayesian approach for characterizing the genetic architecture of viability loci underlying inbreeding depression, and the power analysis.
SJB and DJS initiated the study. Both authors analyzed and interpreted the data and wrote the manuscript.

414

415 **Data Availability**

416 The *Leavenworthia alabamica* reference assembly version 2.5 is available on the CoGe platform:
417 https://genomevolution.org/coge/GenomeInfo.pl?gid=20247. FASTQ files of GBS sequence data
418 for parents and progeny are available on GenBank (**https://www.ncbi.nlm.nih.gov/genbank**)
419 under SRA accession number PRJNA813545

420

421 **Supporting Information**

422

423 **File name**	**Description**
424 TableS1.xlsx	Table S1. Scripts and additional information used to process GBS data in
425	STACKS
426 TableS2.xlsx	Table S2. SNPs from Families 1 and 2, their position in linkage groups
427	and scaffolds; and segregation patterns.
428 Note 1	R code for Bayesian analysis
429 Note 2	R code for power analysis in Fig. 4

430

431

432 **References**

433

434 **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search
435 tool. *Journal of Molecular Biology* **215**: 403-410.
436 **Baldwin SJ, Schoen DJ. 2017.** Genetic variation for pseudo-self-compatibility in self-
437 incompatible populations of *Leavenworthia alabamica* (Brassicaceae). *New Phytologist* **213**:
438 430-429.
439 **Baldwin SJ, Schoen DJ. 2017.** Genetic variation for pseudo-self-compatibility in self-
440 incompatible populations of *Leavenworthia alabamica* (Brassicaceae). *New Phytologist* **213**:
441 430-429.
442 **Baldwin SJ, Schoen DJ. 2019.** Inbreeding depression is difficult to purge in self-incompatible
443 populations of *Leavenworthia alabamica*. *New Phytologist* **224**:1330-1338.
444 **Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful
445 approach to multiple testing. *Journal of the Royal Statistical Society: Series B*
446 *(Methodological)* **57**: 289-300.
447 **Broman KW, Wu H, Sen Ś, Churchill GA. 2003.** R/qtl: QTL mapping in experimental crosses.
448 *Bioinformatics* **19**: 889-90.
449 **Busch JW. 2005**. Inbreeding depression in self-incompatible and self-compatible populations of
450 *Leavenworthia alabamica*. *Heredity* **94**: 159-165.
451 **Busch JW, Joly S, Schoen DJ. 2010.** Does mate limitation in self-incompatible species promote
452 the evolution of selfing? The case of *Leavenworthia alabamica. Evolution* 64:1657-1670.
453 **Busch JW, Joly S, Schoen DJ. 2011.** Demographic signatures accompanying the evolution of
454 selfing in *Leavenworthia alabamica*. *Molecular Biology and Evolution* **28**: 1717-1729.
455 **Chantha SC, Herman AC, Platts AE, Vekemans X, Schoen DJ. 2013.** Secondary evolution of
456 a self-incompatibility locus in the Brassicaceae genus *Leavenworthia*. *PLoS Biology* **14**:
457 e1001560.
458 **Charlesworth D, Lyons EE, Litchfield LB.** 1994 Inbreeding depression in two highly
459 inbreeding populations of *Leavenworthia*. *Proceedings of the Royal Society of London,*
460 *Series B* **258**: 209-214.

Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**: 237-268.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* **10**: 783-796

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences, 2nd ed*. Hillsdale, NJ (USA): Lawrence Erlbaum Associates.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* **6**: e19379.

Frankham R. 2010. Where are we in conservation genetics and where do we need to go? *Conservation Genetics* **11**: 661-3.

Fu YB, Ritland K. 1994a. On estimating the linkage of marker genes to viability genes controlling inbreeding depression. *Theoretical and Applied Genetics* **88**: 925-932.

Fu YB, Ritland K. 1994b. Evidence for the partial dominance of viability genes contributing to inbreeding depression in *Mimulus guttatus*. *Genetics* **136**: 323-331.

Gervais C, Awad DA, Roze D, Castric V, Billiard S. 2014. Genetic architecture of inbreeding depression and the maintenance of gametophytic self-incompatibility. *Evolution* **68**: 3317-3324.

Glémin S. 2003. How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* **57**: 2678-2687.

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, Dewar K. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics* **45**:891-898.

Hedrick PW, Muona O. 1990. Linkage of viability genes to marker loci in selfing organisms. *Heredity* **64**: 67-72.

Hedrick PW, Hellsten U, Grattapaglia D. 2016. Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus grandis. *New Phytologist* **209**: 600-611.

Herman AC, Busch JW, Schoen DJ. 2012. Phylogeny of Leavenworthia S-alleles suggests unidirectional mating system evolution and enhanced positive selection following an ancient population bottleneck. *Evolution* **66**: 1849-1861.

Kärkkäinen K, Kuittinen H, van Treuren R, Vogl C, Oikarinen S, Savolainen O. 1999. Genetic basis of inbreeding depression in *Arabis petraea*. *Evolution* **53**:1354-1365.

Lande R, Schemske DW. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* 39: 24-40.

Llaurens V, Gonthier L, Billiard S. 2009. The sheltered genetic load linked to the S locus in plants: new insights from theoretical and empirical approaches in sporophytic self-incompatibility. *Genetics* **183**: 1105–1118.

Lloyd DG. 1965. Evolution of self-compatibility and racial differentiation in Leavenworthia (Cruciferae). Contributions of the Gray Herbarium **195**:3-134.

Lloyd, DG. 1979. Some reproductive factors affecting the selection of self-fertilization in plants. *The American Naturalist* **113:** 67-79.

Mitchell-Olds T. 1995. Interval mapping of viability loci causing heterosis in Arabidopsis. *Genetics* **140**: 1105-1109.

Porcher E, Lande R. 2005. The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation. *Journal of Evolutionary Biology* **18**: 497-508.

Remington DL, O'Malley DM. 2000. Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. *Genetics* **155**: 337-348.

Schierup MH, Vekemans X, Christiansen FB. 1997. Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**: 835-846.

Seymour DK, Chae E, Arioz BI, Koenig D, Weigel D. 2019. Transmission ratio distortion is frequent in *Arabidopsis thaliana* controlled crosses. *Heredity* **122**: 294-304.

Uyenoyama MK. 1997. Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147**: 1389–1400.

Wallace J, Mitchell S. 2017. Genotyping-by-sequencing. *Current Protocols in Plant Biology* **2**: 64-77.

Waller DM. 2021. Addressing Darwin's dilemma: can pseudo-overdominance explain persistent inbreeding depression and load? *Evolution* **75**: 779-793.

Weir BS. 1996. *Methods for discrete population genetic data. Genetic Data Analysis II*. Sunderland, MA (USA): Sinauer Associates, Inc.

Wright S. 1984. *Evolution and the genetics of populations, vol. 3: Experimental results and evolutionary deductions*. Chicago, IL (USA): University of Chicago Press.

525
526 **Figure Legends**
527
528 **Figure 1.** (Top) -Log_{10} significance levels from Chi-square goodness-of-fit tests to the 1:2:1
529 segregation ratio for SNPs in linkage groups of Family 1. Red dashed line show significance
530 levels corresponding to the false discovery rate of q = 0.05. (Bottom) Segregation ratios for
531 SNPs. Blue denotes heterozygote, while red and black denote homozygotes.
532
533 **Figure 2.** (Top) -Log_{10} significance levels from Chi-square goodness-of-fit tests to the 1:2:1
534 segregation ratio for SNPs in linkage groups of Family 2. Red dashed line show significance
535 levels corresponding to the false discovery rate of q = 0.05. (Bottom) Segregation ratios for
536 SNPs. Blue denotes heterozygote, while red and black denote homozygotes.
537
538 **Figure 3.** Bayesian analysis of SNP segregation ratio data illustrating posterior probability
539 density for values of the parameters h and s (left) with heatmaps showing joint posterior
540 probability densities for h and s (right). Vertical coloured lines define ranges of different nodes
541 of action at linked viability loci: overdominant (left of red line), recessive (between red and blue
542 lines), dominant (between blue and red lines), and underdominant (right of green line). These are
543 sample plots for SNP loci in Table 1 as follows: (Top) 1055_15,414 in Family 1; (Middle)
544 46_1,956,910 in Family 1; (Bottom) 8_24,357 in Family 2.
545
546 **Figure 4.** Statistical power to reject the 1:2:1 segregation ratio for different modes of action of
547 viability loci, selection coefficients, and recombination fractions. Shown are four cases:
548 overdominant with h = -3.0 (red), recessive with h = 0.25 (black), dominant (blue) with h = 0.75,
549 and underdominant (h = 1.5) (orange). From top to bottom graphs, the values of the
550 recombination fraction are c = 0, 0.1, 0.2, and 0.3. The graphs are based on a sample size of 90
551 progeny.
552

Fig. 1

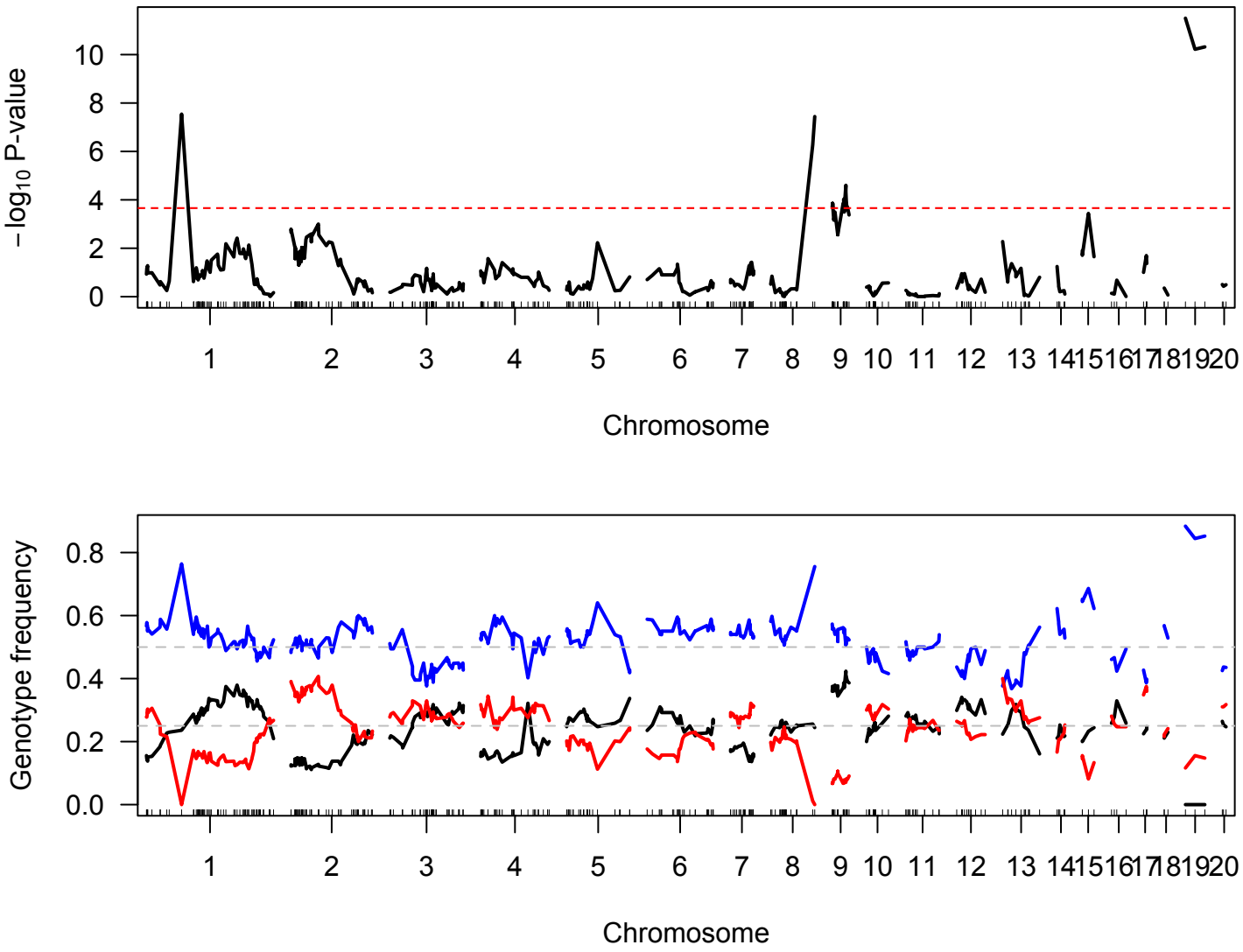

$-\log_{10}$ P-value
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome
Genotype frequency
0.0
0.2
0.4
0.6
0.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome

Fig. 2

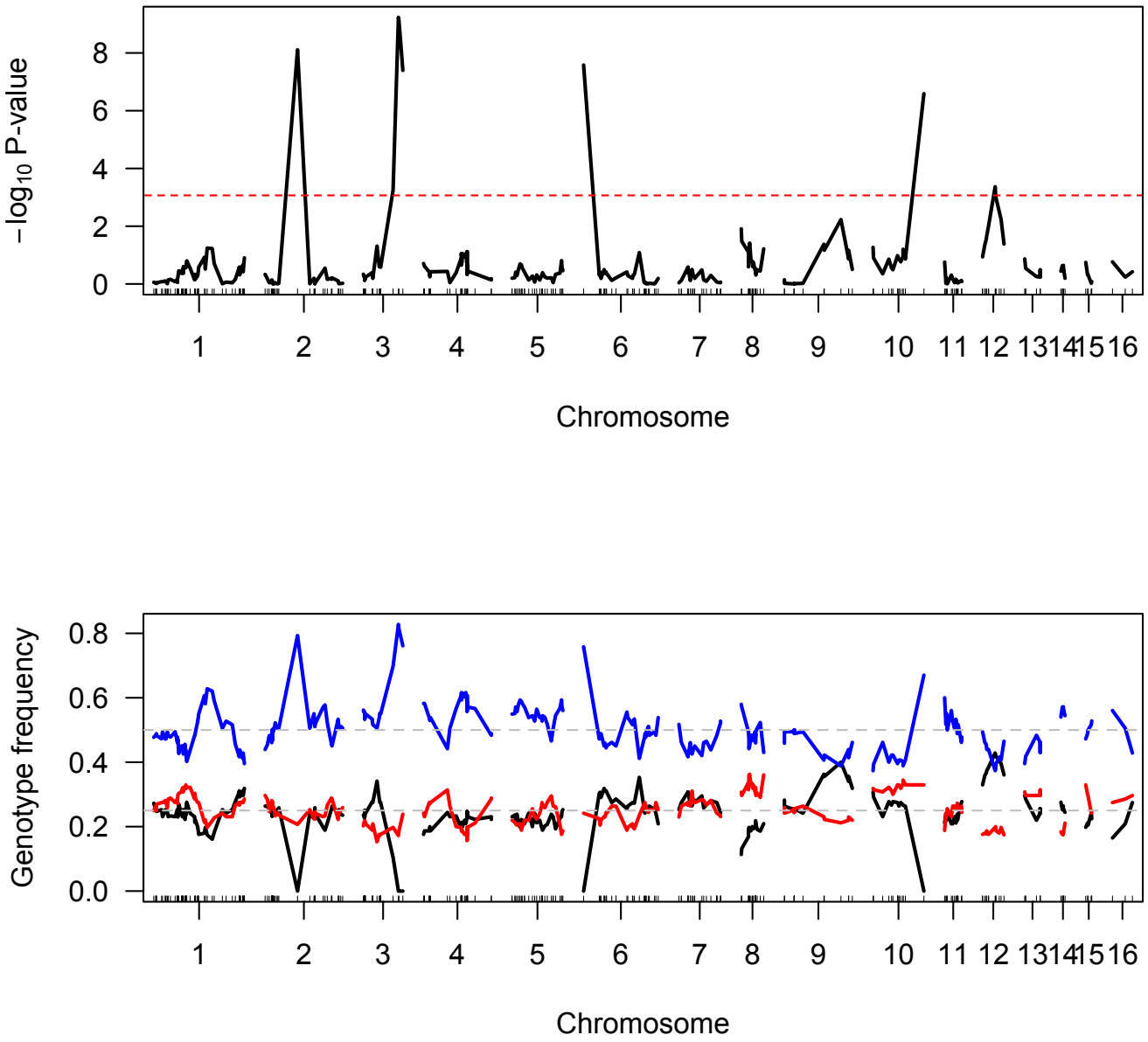
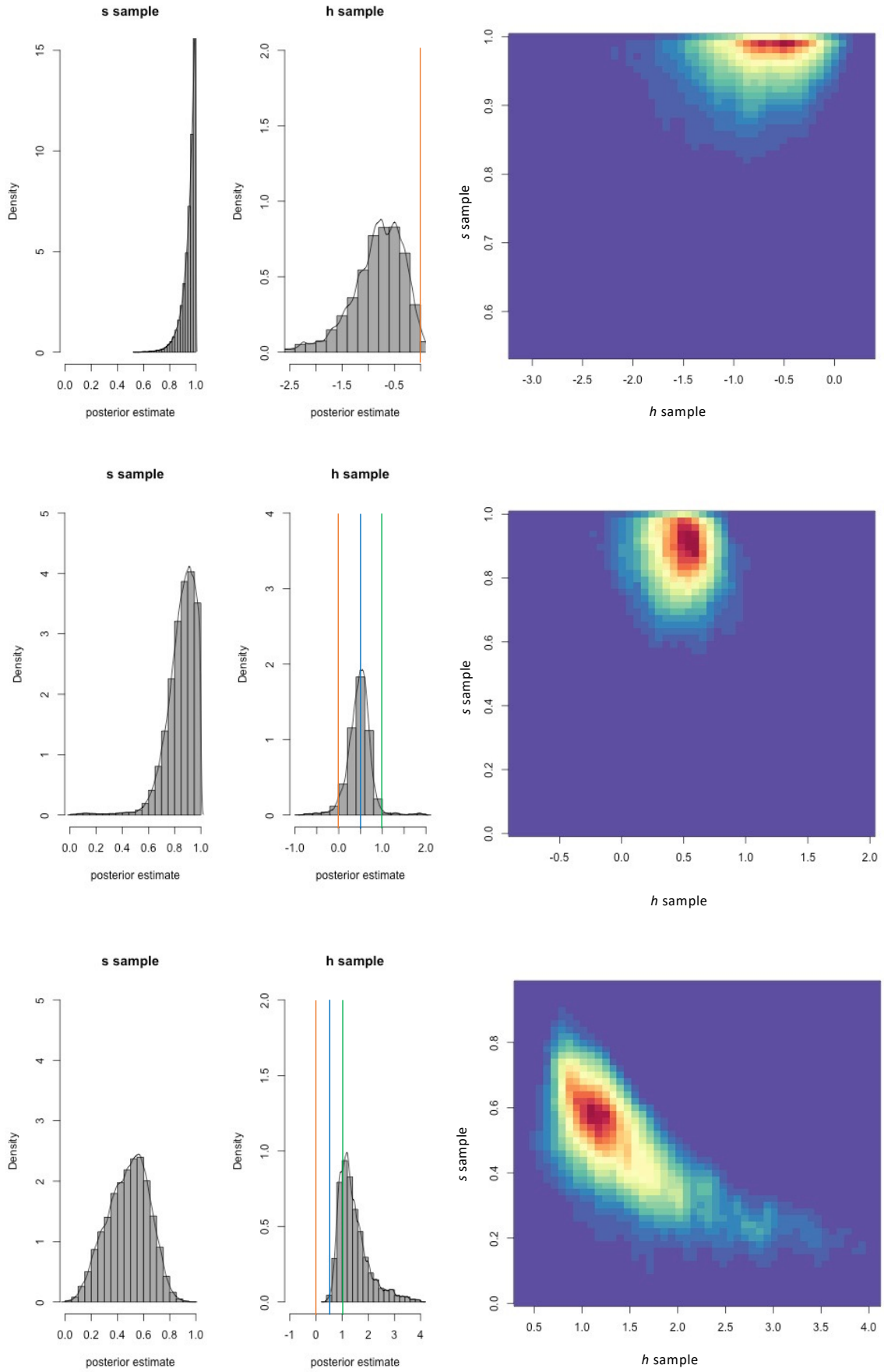

−log10 P-value
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Chromosome
Genotype frequency
0.0
0.2
0.4
0.6
0.8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Chromosome

Fig. 3


s sample
h sample
Density
posterior estimate
s sample
h sample

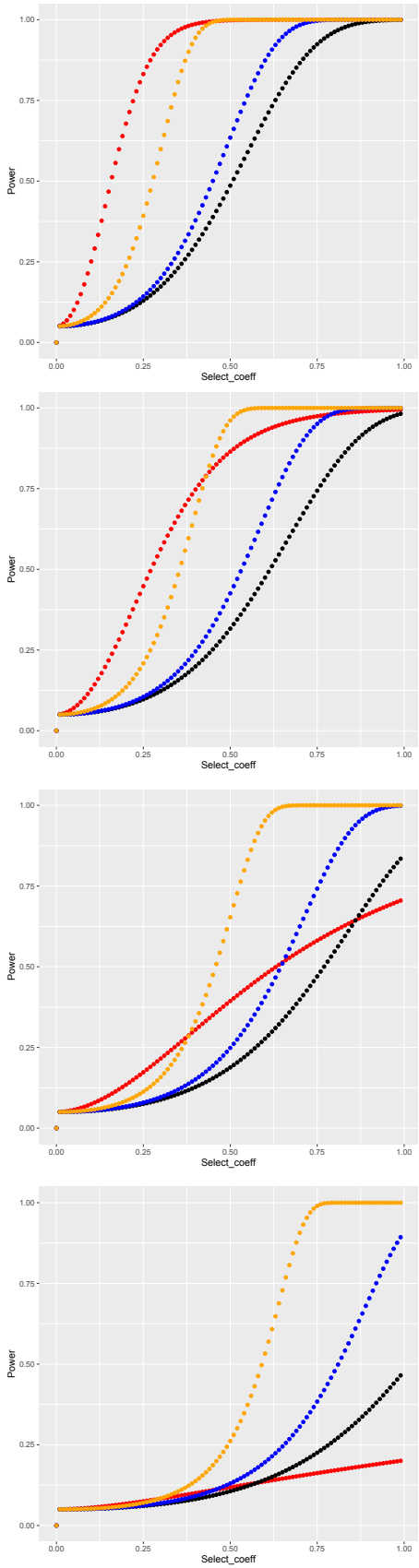



Fig. 4