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, we forced self-pollinated plants from self-incompatible populations of *Leavenworthia alabamica* to produce families of progeny that were genotyped at hundreds of single nucleotide polymorphism (SNP) loci.
- Bayesian analysis of segregation data for each SNP was used to explore support for different dominance and selection coefficients at linked viability loci.
- There was strong support for overdominance (or pseudo-overdomiance) at many viability loci, and 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 reduction of fitness 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 the expression of 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 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. The first was from the Hatton population (Latitude, Longitude: 34.509931, −87.442041), while the second was from the M32 population (Latitude, Longitude 34.489356, −87.63559) (Baldwin & Schoen, 2018). 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)

108 Young leaf tissue samples from the progeny were collected from flowering stage plants
109 and stored at -80°C. DNA was extracted from the leaf tissue with Qiagen plant mini-DNA kits.
110 Extractions were repeated if concentrations were lower than 20 ng/ul and were diluted to 30-50
111 ng/ul if over 50 ng/ul.
112 The genotyping-by-sequencing (GBS) method was used to produce fragments for
113 genotyping with restriction-site associated polymorphisms (Elshire *et al*., 2011) using the protocol
114 of Wallace & Mitchell (2017). The standard protocol was used to generate libraries with 1.44 ng
115 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 =
117 91 and Family 2, N = 90). FASTQC was used to evaluate the basic quality of the sequencing
118 runs. Sequence reads begin with a barcode on the 5' end identifying the plant of origin, then the
119 restriction site, the sequence to be used for genotyping, followed by another restriction site, and
120 finally the Illumina universal adapter. Reads were demultiplexed using Process Radtags in
121 STACKS v. 2.59. The *Leavenworthia alabamica* reference assembly version 2.5 (Haudry *et al.*,
122 2013, CoGe platform: Genome ID 20247) was indexed and sorted, and aligned to the reference
123 genome with the Burrows-Wheeler Aligner. The average read depth across the loci was 276x for
124 Family 1, 277x for Family 2. SNP locus segregation ratios derived from the two families were
125 obtained by producing a catalog of segregating loci using the GSTACKS and POPULATIONS
126 routines in STACKS v. 2.59 (Catchen *et al*., 2013) running in R (version 4.1.1). The in-house
127 scripts for the processing steps and associated information are available in the Supplementary
128 Materials section (Table S1).
129
130 <u>Treatment of potential genotyping errors</u>
131 Genomic analysis of *Leavenworthia alabamica* indicates that this species underwent
132 whole genome triplication during its evolutionary history, and chromosome painting with
133 *Arabidopsis thaliana* BAC probes suggests that that some triplicated genome segments were
134 retained as duplicates or triplicates (Haudry *et al*., 2013). The reference genome assembly in this
135 species is only partially complete (N50 = 309 kb); while the haploid chromosome number of *L.*
136 *alabamica* is 11, the genome assembly consists of 3,321 scaffolds. It comprises *ca.* 174 million
137 bp in total of the expected 315 million bp genome size, the latter estimated from flow cytometry
138 (Brian Husband, pers. comm.). Over half of the scaffolds are short, with lengths 5,000 bp or

139 smaller, though larger ones are > 1M bp. Due to sequence divergence of the duplicated (or
140 triplicated) regions, longer scaffolds in the reference genome likely represent separate genomic
141 regions (Haudry *et al*., 2013). But many shorter duplicated (or triplicated) regions of the genome
142 with extended identity have likely been collapsed into single (artifactual) contigs. GBS products
143 from separate genomic regions that map to such short, artifactual contigs will often yield
144 heterozygote genotype calls. To minimize these artifacts we: (1) restricted our analysis to SNPs
145 that map to scaffolds > 100 kbp in length (these account for 73% of the *L. alabamica* assembly);
146 (2) rejected SNP loci that exhibited only heterozygotes if adjacent SNP loci on the same scaffold
147 exhibited segregation ratios that were not significantly different from 1:2:1 by Chi-square tests,
148 on the basis that evidence of strong selection should be shared by adjacent SNP loci; and (3)
149 restricted our analysis to loci where data were available from > 90% of the progeny of a cross
150 since genotyping errors are more likely to have occurred when few progeny per family are
151 genotyped successfully.
152
153 Linkage map construction and tests for departure from normal segregation
154 Linkage groups (each containing a minimum of three SNP loci) were constructed with
155 the software package 'qtl' version 1.50 (Broman *et al*., 2003) running in R (version 4.1.1).
156 Separate linkage maps were created from the Family 1 and 2 data. These maps allowed us to
157 examine segregation ratios of each SNP in relation to that of other SNPs. To arrive at a
158 conservative estimate of the number of viability loci, we limited our analysis to those SNPs that
159 exhibited departures from the expected (null) 1:2:1 segregation ratio at a false discovery rate
160 (Benjamini & Hochberg, 1995) of $q = 0.05$. Moreover, a viability locus located adjacent to a
161 given SNP locus will likely affect not only the segregation ratio at that SNP but also at other
162 nearby SNPs, and so within each 50 cM window if several such SNPs were found, we selected
163 the one exhibiting the most extreme departure from 1:2:1 for analysis of the selection and
164 dominance coefficients at the linked viability locus.
165
166 Selection and dominance coefficients at linked viability loci
167 The analysis of marker locus segregation ratios to estimate selection and dominance at
168 viability loci linked to marker loci assumes that departure from the expected 1:2:1 Mendelian
169 segregation ratio is due to viability selection, rather than to meiotic drive and/or gametic selection.

The parent plant that is self-fertilized is assumed to be heterozygous at a selectively neutral marker locus (M) and a linked viability locus (V)—e.g., the parent has the genotype M_1V_1/M_2V_2, and the 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, respectively, where s is the selective coefficient or the proportional amount by which selection at this locus reduces viability in the homozygote V_2/V_2 ($0 \leq s \leq 1$), and h is the dominance coefficient. 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:

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

198 where N is the sum of N_{11}, N_{12}, and N_{22}. The joint density and posterior probability are,
199 respectively:
200
201 $$Pr(s, h, c, N_{11}, N_{12}, N_{22}) = Pr(s)\, Pr(h)\, Pr(c)\;\; Pr(N11, N12, N22 \mid s, h, c) \quad (3)$$
202
203 $$Pr(s, h, c \mid N_{11} N_{12} N_{22}) = \frac{Pr(s)\, Pr(h)\, Pr(c)\, Pr(N11, N12, N22 \mid s,h,c)}{\int_{s,h,c} Pr(s)\, Pr(h)\, Pr(c) Pr(N11, N12, N22 \mid s,h,c)\;\; ds\, dh\, dc} \quad (4)$$
204
205
206 To explore values of the parameters using a Bayesian approach, we chose uniform priors with 0
207 $< s < 1$, $-100 < h < 100$, and $0 < c < 0.2$. The broad prior for h allows for the possibility of
208 extreme overdominance and underdominance. The prior for c (less than the theoretical upper
209 limit of $c = 0.5$) is justified based on the average distance between markers of *ca.* 2 cM in this
210 study. A program written in the R statistical language (version 4.1.1) to conduct Markov Chain
211 Monte Carlo method with the Metropolis-Hastings sampler (Weir, 1996) (Supplementary
212 Materials: Note 1). This was used to "integrate out" the nuisance parameter c, allowing the
213 posterior probabilities to be estimated for different values of s and h. For the required burn-in
214 period, we ran the sampler 5.5 million times before examining the posterior probability
215 distributions. Tracking the parameter values during the runs showed that the sampler had
216 stabilized by the end of this burn-in period.
217
218 <u>The S-locus region of *Leavenworthia alabamica* and locations of viability loci</u>
219 The *Leavenworthia alabamica* S-locus region has previously been cloned and sequenced
220 from the same plant material that was used to produce the sequence information in the genome
221 assembly in the present work (GenBank accession KC981242.1) (Chantha *et al*., 2013). This
222 allowed us to locate the individual *L. alabamica* genome assembly scaffold containing the S-
223 locus. The BLAST algorithm (Altschul *et al.,* 1990) implemented in software package Geneious
224 Prime (version 2.2) was used to conduct the search of *L. alabamica* reference genome version
225 2.5 using both *L. alabamica* S-locus genes (*Lal2* and *LaSCRL*; Chantha *et al.*, 2013). Selection
226 and dominance coefficients were estimated for all potential viability loci linked to SNPs in this
227 region that showed significant departures from 1:2:1 were analysed.

228

229 Statistical power test and different modes of gene action

230 The analysis of the number and mode of action viability loci hinges on the ability to
231 detect departure from the expected 1:2:1 ratio obtained when self-fertilizing a plant with
232 heterozygous marker loci. Given that the sample size of progeny per family in our work was
233 limited to a maximum of 91 progeny (the maximum number we could obtain by forced selfing of
234 each of the SI parent plants used), we were interested in the statistical power of our study to
235 detect selection and characterize the mode of gene action. Specifically, we did this by producing
236 expected segregation ratios for a range of different values of the parameters in equation (1)
237 followed by standard methods (Cohen, 1988) for examining statistical power, as implemented in
238 the R program package 'pwr' version 1.3.

239

240 **Results**

241

242 SNP segregation ratios, linkage map locations, and significance testing

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

248 SNP segregation patterns along with and their significance levels from the Chi-square test
249 of goodness-of-fit to the hypothetical 1:2:1 segregation ratio expected in the absence of selection
250 varied each linkage group are shown in Figures 1 and 2. With a false discovery rate of $q = 0.05$,
251 we detected eight putative viability loci in Family 1, and eight more in Family 2 (Table 1).

252

253 Selection and dominance of viability loci linked to SNPs

254 The 16 putative viability loci detected were further characterized for their dominance and
255 selection coefficients using the Bayesian approach outlined above. For 12 out of 16 of SNPs, the
256 posterior probability mass for inferred linked viability loci indicated overdominant gene action,
257 with Bayesian estimates of h ranging from -17.47 to -1.07, and s ranging from 0.32 to 0.96
258 (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).

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

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

Statistical power to detect different modes of gene action

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

Discussion

The genetic architecture of inbreeding depression.

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

320 with overdominant effects. Finally, Hedrick *et al.* (2016) examined segregation ratios for nearly
321 10,000 SNPs in 28 progeny descended from a self-polliinated plant of *Eucalytus grandis*, and
322 found widespread heterozygote excess, consistent with overdominance, though their simulation
323 analysis suggested that pseudo-overdominance is the more likely explanation. Overall, these
324 empirical studies suggest that modes of gene action for linked viability loci may vary from
325 species to species.
326 It is notable that a few of the above-cited studies have reported underdominant viability
327 loci linked to markers (e.g., especially Fu & Ritland, 1994b, but also Kärkkäinen *et al*. 1999).
328 Underdominance was also inferred at two marker-linked viability loci in the present
329 investigation. Underdominance at viability loci cannot contribute to inbreeding depression. We
330 note from the results of the power analyses we conducted, underdominance is detectable with
331 smaller sample sizes compared with other modes of action that require larger sample sizes to
332 detect selection of the same magnitude when $c < 0.3$ (Fig. 4). This could help account for the
333 unexpected reports of underdominance in studies that have employed few markers (e.g., Fu &
334 Ritland, 1994b).
335
336 Overdominance or pseudo-overdominance of inbreeding depression
337 For 12 of the 16 SNPs showing departures from normal segregation, our analysis
338 suggests overdominance at linked viability loci contributes to inbreeding depression.
339 Alternatively, these results could be explained as being due to pseudo-overdominance (Ohta,
340 1971). Psuedo-overdominance would be detected in a genomic region by the methods used here
341 if deleterious, recessive viability mutations are linked and in repulsion. In general, compared
342 with recessive or partly recessive action at viability loci, there is little support from past studies
343 for true overdominance as an underlying cause of inbreeding depression (Charlesworth and
344 Willis 2009). Distinguishing between true overdominance and pseudo-overdominance is difficult
345 when there is a single family of progeny, but becomes progressively easier with multiple
346 families, as in the present study. True overdominance, if present, should often be detected at the
347 same viability loci in the progeny of multiple selfed parents, since selection there should render it
348 frequent. In contrast, viability loci exhibiting pseudo-overdominance arising from deleterious
349 recessive or partly recessive alleles are less likely to be shared among parents, as the genomic
350 regions are maintained by the interaction of mutation and selection. Our study was limited to two

351 separate families, but we did, in fact, find inferred overdominant viability loci in the two contigs
352 in both Families 1 and 2 (SNPs 68_160676 and SNPs 68_16082; and SNP 72_1428232 in both
353 families). No other such cases were detected (Table S2). This leaves open the possibility that
354 both true and pseudo-overdominance are present.
355
356 Sheltered S-linked load
357 An inferred viability locus (or possibly, multiple linked loci) with recessive action was
358 detected only in one genomic region, and one family (Family 1). Interestingly, the region is
359 within 80 kb from the site of the S-locus. These results can be interpreted as evidence that there
360 is sheltering from selection of one (or more) recessive, deleterious mutations in the S-locus
361 region (Table 2).
362 In Family 2, no such effects were detected in the S-locus region. The finding is not
363 necessarily at odds with the evidence of sheltered load in the Family 1 data. There are likely
364 several levels of dominance for S-locus alleles in *L. alabamica* (Busch *et al.*, 2010), and in
365 sporophytic systems such as this one, the strongest sheltering from selection is expected for those
366 viability loci linked to S-alleles that exhibit the highest degree of dominance (Llaurens *et al.*,
367 2009).
368
369 Relationship of results to earlier findings on inbreeding depression and the evolution of the
370 mating system in *Leavenworthia alabamica*
371 Busch (2005) discovered significant levels of inbreeding depression in self-incompatible
372 populations of *Leavenworthia alabamica* for germination and survival rate, biomass, flower
373 number, and pollen viability. He estimated that inbreeding depression for total lifetime fitness
374 either as a male or female parent exceeds 0.5. Baldwin & Schoen (2019) studied inbreeding
375 depression for a similar set of fitness components in *L. alabamica* and again found inbreeding
376 depression for lifetime fitness exceeds 0.5. Baldwin & Schoen extended their study to include
377 three consecutive generations of self-pollination and found no evidence that inbreeding
378 depression was purged by this treatment. These results are consistent with the findings of the
379 present investigation, where overdominance (or more likely, pseudo-overdominance) is
380 suggested for many linked viability loci.

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

412 **Author Contributions**

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

417

418 **Data Availability**

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

423

424 **Supporting Information**

425

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

433

434

435 **References**

436

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

464 **Charlesworth D, Charlesworth B. 1987**. Inbreeding depression and its evolutionary
465 consequences. *Annual Review of Ecology and Systematics* **18**: 237-268.
466 **Charlesworth D, Willis JH. 2009.** The genetics of inbreeding depression. *Nature Reviews*
467 *Genetics* **10**: 783-796
468 **Cohen, J. 1988.** *Statistical power analysis for the behavioral sciences, 2nd ed*. Hillsdale, NJ
469 (USA): Lawrence Erlbaum Associates.
470 **Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011.**
471 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS*
472 *One* **6**: e19379.
473 **Frankham R. 2010.** Where are we in conservation genetics and where do we need to go?
474 *Conservation Genetics* **11**: 661-3.
475 **Fu YB, Ritland K. 1994a.** On estimating the linkage of marker genes to viability genes
476 controlling inbreeding depression. *Theoretical and Applied Genetics* **88**: 925-932.
477 **Fu YB, Ritland K. 1994b.** Evidence for the partial dominance of viability genes contributing to
478 inbreeding depression in *Mimulus guttatus*. *Genetics* **136**: 323-331.
479 **Gervais C, Awad DA, Roze D, Castric V, Billiard S. 2014.** Genetic architecture of inbreeding
480 depression and the maintenance of gametophytic self-incompatibility. *Evolution* **68**: 3317-
481 3324.
482 **Glémin S. 2003**. How are deleterious mutations purged? Drift versus nonrandom mating.
483 *Evolution* **57**: 2678-2687.
484 **Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-**
485 **Lopez Z, Steffen JG, Hazzouri KM, Dewar K. 2013.** An atlas of over 90,000 conserved
486 noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics*
487 **45**:891-898.
488 **Hedrick PW, Muona O. 1990.** Linkage of viability genes to marker loci in selfing organisms.
489 *Heredity* **64**: 67-72.
490 **Hedrick PW, Hellsten U, Grattapaglia D. 2016.** Examining the cause of high inbreeding
491 depression: analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus
492 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.

528
529 **Figure Legends**
530
531 **Figure 1.** (Top) -Log_{10} significance levels from Chi-square goodness-of-fit tests to the 1:2:1
532 segregation ratio for SNPs in linkage groups of Family 1. Red dashed line show significance
533 levels corresponding to the false discovery rate of q = 0.05. (Bottom) Segregation ratios for
534 SNPs. Blue denotes heterozygote, while red and black denote homozygotes.
535
536 **Figure 2.** (Top) -Log_{10} significance levels from Chi-square goodness-of-fit tests to the 1:2:1
537 segregation ratio for SNPs in linkage groups of Family 2. Red dashed line show significance
538 levels corresponding to the false discovery rate of q = 0.05. (Bottom) Segregation ratios for
539 SNPs. Blue denotes heterozygote, while red and black denote homozygotes.
540
541 **Figure 3.** Bayesian analysis of SNP segregation ratio data illustrating posterior probability
542 density for values of the parameters h and s (left) with heatmaps showing joint posterior
543 probability densities for h and s (right). Vertical coloured lines define ranges of different nodes
544 of action at linked viability loci: overdominant (left of red line), recessive (between red and blue
545 lines), dominant (between blue and red lines), and underdominant (right of green line). These are
546 sample plots for SNP loci in Table 1 as follows: (Top) 1055_15,414 in Family 1; (Middle)
547 46_1,956,910 in Family 1; (Bottom) 8_24,357 in Family 2.
548
549 **Figure 4.** Statistical power to reject the 1:2:1 segregation ratio for different modes of action of
550 viability loci, selection coefficients, and recombination fractions. Shown are four cases:
551 overdominant with h = -3.0 (red), recessive with h = 0.25 (black), dominant (blue) with h = 0.75,
552 and underdominant (h = 1.5) (orange). From top to bottom graphs, the values of the
553 recombination fraction are c = 0, 0.1, 0.2, and 0.3. The graphs are based on a sample size of 90
554 progeny.
555

Table 1. SNPs that show significant deviations from 1:2:1. SNP (scaffold_bp) and linkage group (LG) and posterior probabilities of overdominant (OD), recessive (R), dominant (D), and underdominant (UD) gene actions[1]. The values for the selection coefficient (*s*) and dominance coefficient (*h*) are the Bayesian estimators (with 95% credible intervals).

Family	SNP	LG	OD	R	D	UD	*s*	*h*
1	68_160,676	1	0.99	0.01	0	0	0.94 (0.82, 0.99)	-1.04 (-2.29, -0.28)
1	1055_15,414	8	0.99	0.01	0	0	0.93 (0.76, 0.98)	-1.28 (-3.10, -0.25)
1	46_1,956,910	9	0.04	0.51	0.44	0.01	0.86 (0.68, 0.98)	0.45 (0.01, 0.79)
1	312_10,873	19	1.0	0	0	0	0.75 (0.42, 0.97)	-11.79 (-14.87, -9.27)
1	67_39,711	N/A[2]	1.0	0	0	0	0.58 (0.15, 0.96)	-17.47 (-21.97, -15.22)
1	72_1,428,232	N/A	0.93	0.04	0.03	0	0.32 (0.05, 0.78)	-3.87 (-6.78, 0.16)
1	32_1,491,967	N/A	0	0	0.01	0.99	0.67 (0.50, 0.81)	1.48 (1.12, 1.99)
1	197_101,568	N/A	1.0	0	0	0	0.90 (0.69, 0.98)	-3.58 (-6.30, -1.56)
2	257_36,756	2	1.0	0	0	0	0.91 (0.72, 1.00)	-2.11 (-3.52, -0.58)
2	239_37,562	3	1.0	0	0	0	0.91 (0.72, 1.00)	-2.58 (-5.08, -0.92)
2	27_153,823	6	0.99	0.01	0	0	0.93 (0.82, 1.00)	-1.07 (-2.27, -0.18)
2	194_219,515	10	0.69	0.31	0	0	0.96 (0.88, 1.00)	-0.25 (-0.62, 0.24)
2	8_24,357	12	0	0.03	0.27	0.70	0.51 (0.26, 0.74)	1.30 (0.66, 2.32)
2	36_1,952,261	N/A	0	0	0	1.0	0.68 (0.54, 0.80)	1.48 (1.14, 1.96)
2	72_1,428,232	N/A	0.99	0.01	0	0	0.94 (0.80, 1.00)	-1.10 (-2.35, -0.24)
2	175_5,265,206	N/A	0.99	0.01	0	0	0.94 (0.82, 1.00)	-1.07 (-2.28, -0.18

[1] Cumulative posterior probabilities associated with overdominant (OD), recessive (R), dominant (D), and underdominant (UD) modes of gene actions were determined by summing over the posterior probabilities associated with values of $h < 0$, $0 \leq h < 0.5$, $0.5 \leq h < 1$, and $h \geq 1$, respectively.

[2] Not mapped to a linkage group.

Table 2. SNPs on the scaffold containing the *Leavenworthia alabamica* S-locus (*Lal2*)[1] that show significant deviations from 1:2:1 segregation. SNP locus (scaffold_base pair position) and posterior probabilities of overdominant (OD), recessive (R), dominant (D), and underdominant (UD) gene actions[1]. The values for the selection coefficient (*s*) and dominance coefficient (*h*) are the Bayesian estimators (90% credible intervals).

SNP	OD	R	D	UD	*s*	*h*
46_1,497,122	0.26	0.51	0.23	0.00	0.80 (0.52, 0.98)	0.18 (-0.76, 0.71)
46_1,807,312	0.09	0.52	0. 39	0.00	0.82 (0.59 .0.98)	0.39 (-0.18, 0.80)
46_1,807,671	0.10	0.56	0. 33	0.01	0.80 (0.53, 0.95)	0.36 (-0.16,0 .79)
46_1,956,910	0.04	0.51	0.44	0.01	0.86 (0.68, 0.98)	0.45 (0.01, 0.79)
46_1,957,161	0.09	0.52	0.38	0.01	0.85 (0.64, 0.96)	0.38 (-0.18, 0.80)

[1] Base pair position of the start of exon 1 of *LalSCRL* on scaffold 46 is 2,035,545 bp

Fig. 1

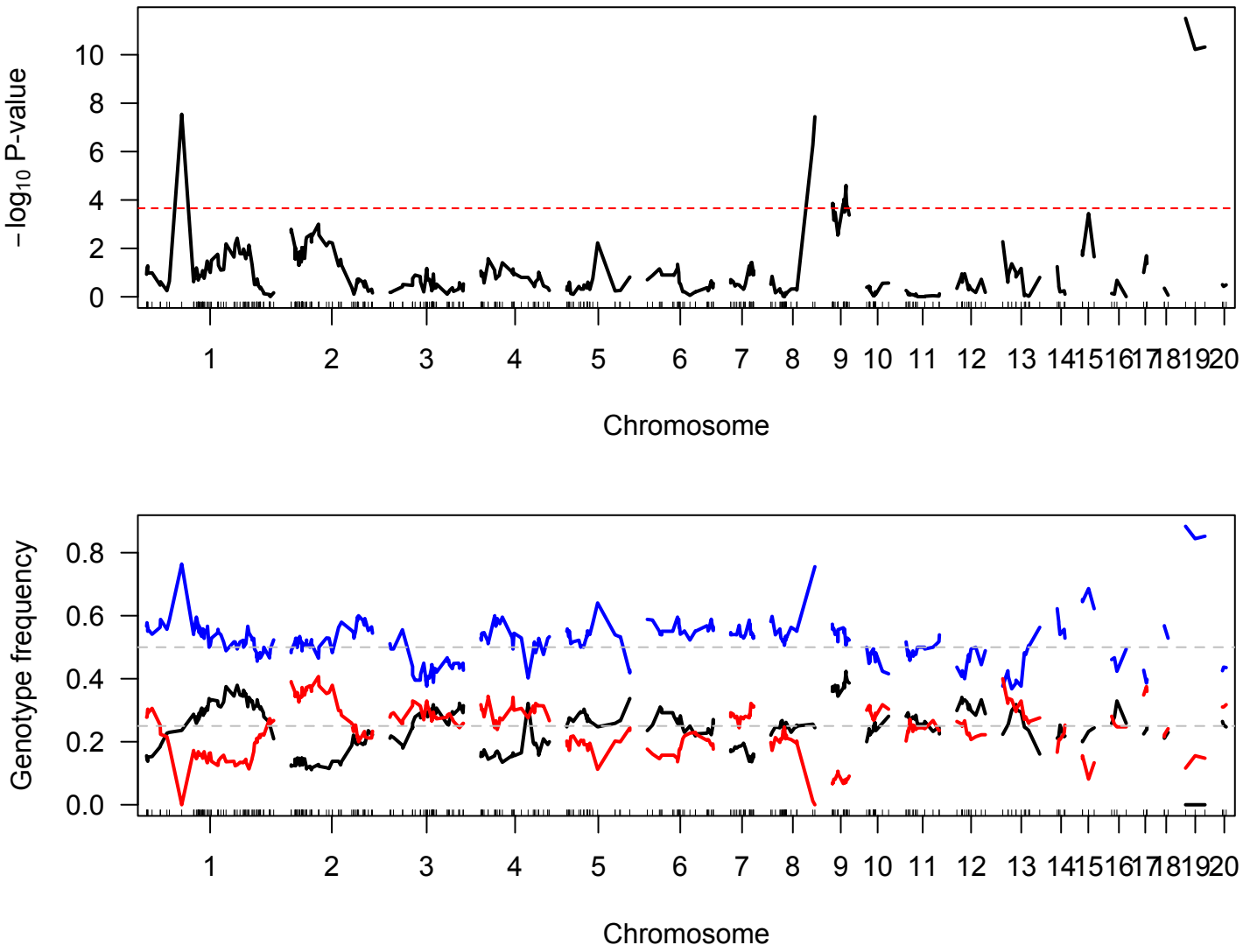

$-\log_{10}$ P-value
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome
Genotype frequency
0.0
0.2
0.4
0.6
0.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Chromosome

Fig. 2

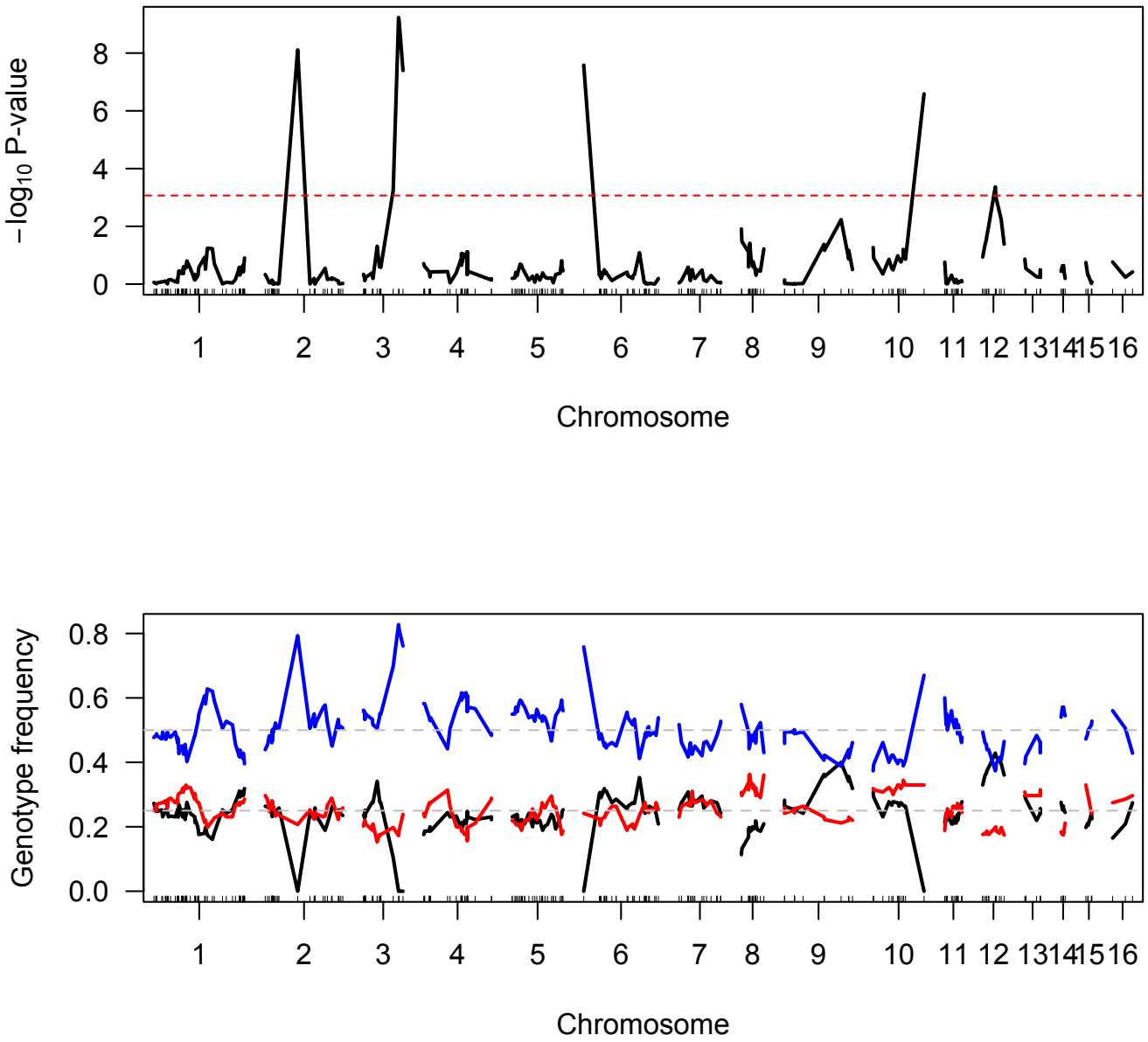
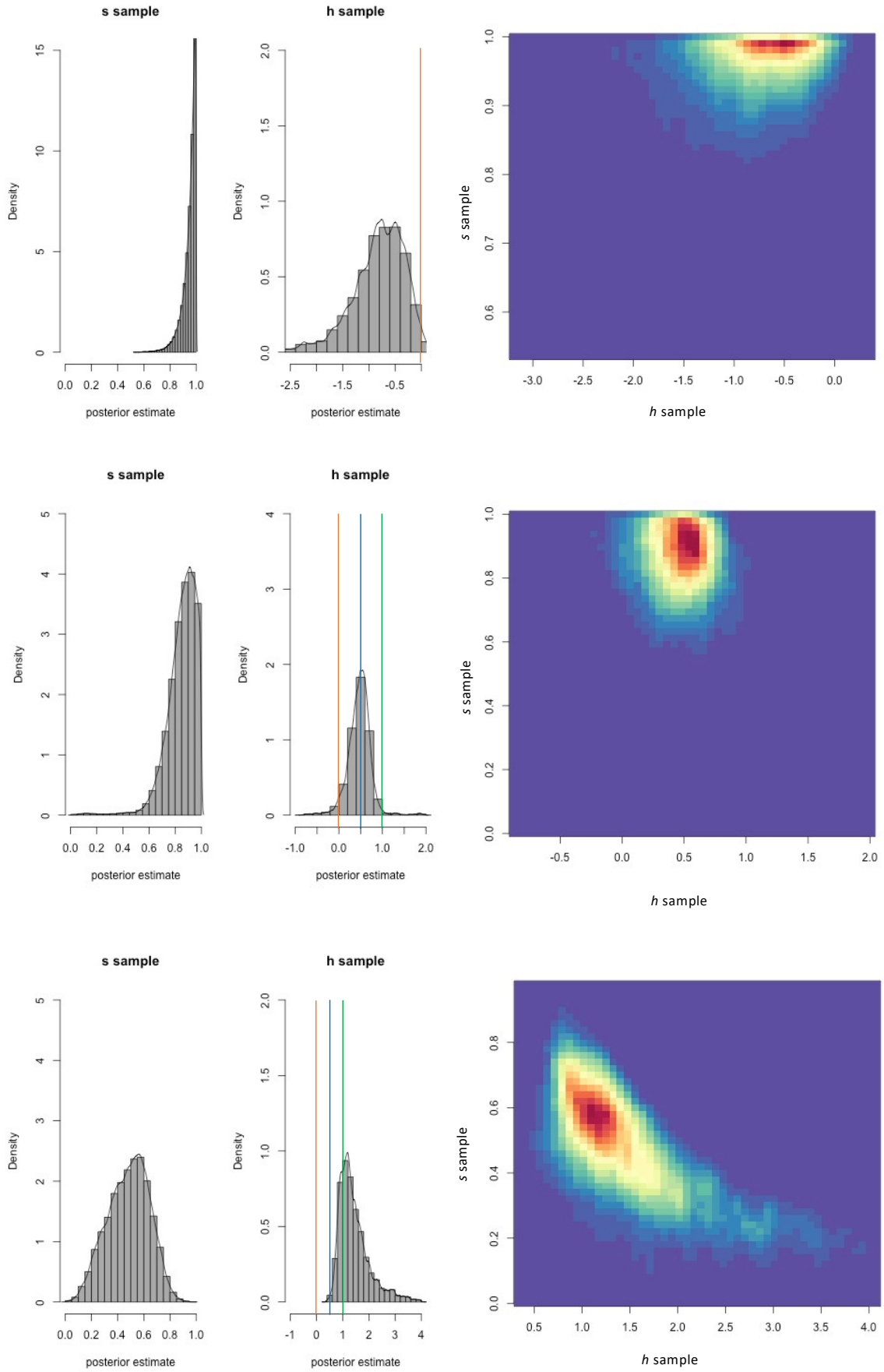

−log10 P-value
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Chromosome
Genotype frequency
0.0
0.2
0.4
0.6
0.8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Chromosome

Fig. 3


s sample
h sample
Density
posterior estimate
s sample
h sample

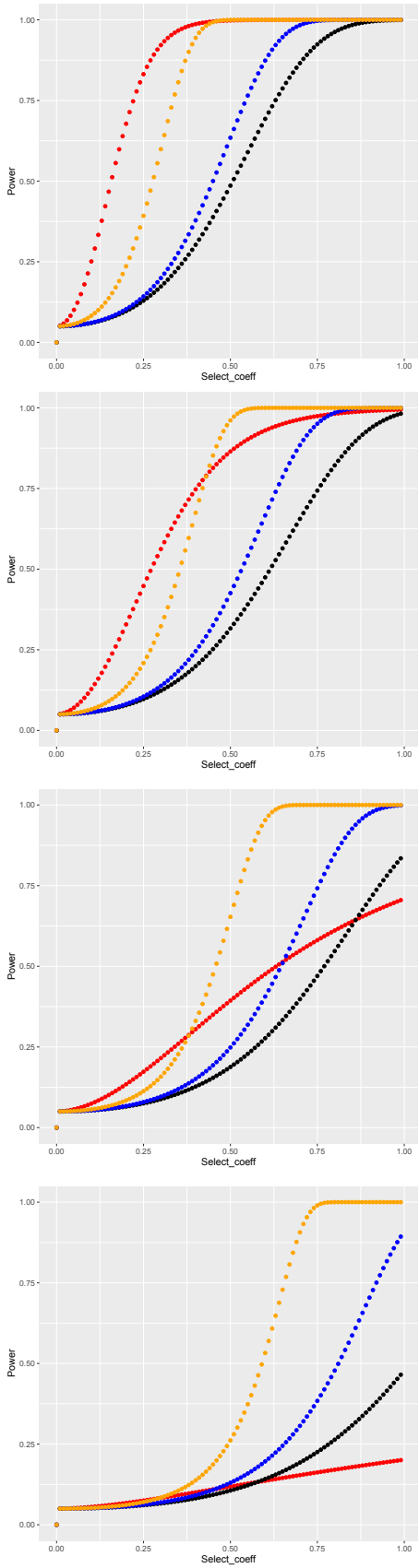



Fig. 4