The maintenance of self-incompatibility and the genetic architecture of inbreeding depression

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: 175
Introduction: 567
Methods: 1774
Results: 684
Discussion: 1357
Total: 4557
Tables: 2
Figures: 3
Supplementary Information: Tables S1- S3, 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 mapped 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 in different genomic regions.
- There was some support for overdominance (or pseudo-overdominance) at a few viability loci, as well as for recessiveness and underdominance. One recessive viability locus mapped to the genomic region of the novel self-incompatibility locus in *Leavenworthia alabamica,* but in general there was no support for strongly recessive viability loci of major effect.
- 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, 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 depression. 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 and self-compatible 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 populations in Alabama 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. Sample sizes were 91 and 92 progeny/family, respectively.

111 Genotyping-by-sequencing (GBS)

112 Young leaf tissue samples from the progeny were collected from flowering stage plants
113 and stored at -80°C. DNA was extracted from the leaf tissue with Qiagen plant mini-DNA kits.
114 Extractions were repeated if concentrations were lower than 20 ng/ul and were diluted to 30-50
115 ng/ul if over 50 ng/ul.

116 The genotyping-by-sequencing (GBS) method was used to produce fragments for
117 genotyping with restriction-site associated polymorphisms (Elshire *et al*., 2011) using the
118 protocol of Wallace & Mitchell (2017). The standard protocol was used to generate libraries with
119 1.44 ng PstI per 100 ng DNA and 15-18 cycles of PCR. Fragments were then sequenced on one
120 lane of an Illumina High-Seq run (Family 1, N = 91 and Family 2, N = 90). FASTQC was used
121 to evaluate the basic quality of the sequencing runs. Sequence reads begin with a barcode on the
122 5' end identifying the plant of origin, then the restriction site, the sequence to be used for
123 genotyping, followed by another restriction site, and finally the Illumina universal adapter. Reads
124 were demultiplexed using Process Radtags in STACKS v. 2.59. The *Leavenworthia alabamica*
125 reference assembly version 2.5 (Haudry *et al.*, 2013, CoGe platform: Genome ID 20247) was
126 indexed and sorted with Burrows-Wheeler Aligner Tool, BWA-0.7.17 (Li & Durbin, 2010), and
127 reads were aligned to the reference genome with BWA-0.7.17. SNP locus segregation ratios
128 derived from the two families were obtained by producing a catalog of segregating loci using the
129 GSTACKS and POPULATIONS routines in STACKS v. 2.59 (Catchen *et al*., 2013) running in
130 R (version 4.1.1). The in-house scripts and associated information used for processing the data
131 are available in the Supplementary Materials section (Table S1).

132

133 SNP filtering and treatment of potential genotyping errors

134 Previous genomic analysis of *Leavenworthia alabamica* indicates that this species
135 underwent whole genome triplication during its evolutionary history, and chromosome painting
136 with *Arabidopsis thaliana* BAC probes suggests that that some triplicated genome segments
137 were retained as duplicates or triplicates (Haudry *et al*., 2013). The reference genome assembly
138 in this species is only partially complete (N50 = 309 kb); while the haploid chromosome number
139 of *L. alabamica* is 11, the genome assembly consists of 3,321 scaffolds. It comprises *ca.* 174
140 million bp in total of the expected 315 million bp genome size, the latter estimated from flow
141 cytometry (Brian Husband, pers. comm.). Over half of the scaffolds are short, with lengths 5,000

142 bp or smaller, though larger ones are > 1M bp. While there is sequence divergence of the
143 duplicated (or triplicated) regions, longer scaffolds in the reference genome likely represent
144 separate genomic regions (Haudry *et al*., 2013). But many shorter duplicated (or triplicated)
145 regions of the genome with extended identity have likely been collapsed into single (artifactual)
146 contigs. GBS products from separate genomic regions that map to such short, artifactual contigs
147 would be expected to yield heterozygote genotype calls. To minimize these artifacts we: (1)
148 filtered out heterozygotes when allele balance fell below 0.333 or above 0.667 using the R-
149 package SNPfiltr v. 1.00 (DeRaad, 2022), (2) restricted our analysis to SNPs that map to
150 scaffolds > 100 kbp in length (these account for 73% of the *L. alabamica* assembly); (3) filtered
151 out SNP loci that exhibited only heterozygotes when adjacent SNP loci on the same scaffold
152 exhibited segregation ratios that were not significantly different from 1:2:1 by Chi-square tests,
153 on the basis that evidence of strong selection should be shared by adjacent SNP loci; and (4)
154 restricted our analysis to loci where data were available from > 90% of the progeny of a cross
155 since genotyping errors are more likely to have occurred when few progeny per family are
156 genotyped successfully.
157
158 <u>Selection and dominance coefficients at linked viability loci</u>
159 Linkage groups (each containing a minimum of three SNP loci) were constructed
160 separately for Families 1 and 2 with the software package 'qtl' version 1.50 (Broman *et al*., 2003)
161 running in R (version 4.1.1). A viability locus located adjacent to a given SNP locus will likely
162 affect not only the segregation ratio at that SNP but also at other nearby SNPs. The linkage maps
163 allowed us to examine segregation ratios of each SNP in relation to that of other nearby SNPs. To
164 arrive at a conservative estimate of the number of viability loci, we limited our analysis to those
165 SNPs that exhibited departures from the expected (null) 1:2:1 segregation ratio at a false discovery
166 rate (Benjamini & Hochberg, 1995) of q = 0.05 and 0.10. For linkage groups where departures
167 from a 1:2:1 ratio were observed, we selected the marker showing the most extreme departure for
168 the analysis of selection and dominance coefficients at the linked viability locus.
169 We note that the analysis of marker locus segregation ratios to estimate selection and
170 dominance at viability loci linked to marker loci assumes departure from the expected 1:2:1
171 Mendelian segregation ratio is due to viability selection, rather than to meiotic drive and/or gametic
172 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:

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

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

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

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

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 < 5 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). The program 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.

<u>The S-locus region of *Leavenworthia alabamica* and locations of viability loci</u>

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. The BLAST algorithm (Altschul *et al.,* 1990) implemented in software package Geneious Prime (version 2.2) was used to conduct the search of *L. alabamica* reference genome version 2.5 using both *L. alabamica* S-locus genes (*Lal2* and *LaSCRL*; Chantha *et al.*, 2013). Selection and dominance coefficients were estimated for all potential viability loci linked to SNPs in this region that showed significant departures from 1:2:1 were analysed.

230

231 Statistical power test and different modes of gene action

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

241

242 **Results**

243

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

245 We found 315 mappable SNP loci in Family 1, and 237 in Family 2. In Family 1 we
246 found 18 linkage groups with 3 or more SNPs, ranging in length from 2 to 201 cM, for a total
247 map length of 901 cM. In Family 2 we found 16 linkage groups with 3 or more SNPs, ranging in
248 length from 3 to 88 cM, for a total map length of 229 cM. SNP grouping by linkage group was
249 consistent in the two families (see Supplementary Tables S2 and S3 for linkage group, scaffold
250 positions, and segregation patterns of SNPs). SNP segregation patterns along with and their -
251 $\log_{10}$ significance levels from the Chi-square test of goodness-of-fit to the null 1:2:1 segregation
252 ratio varied within each linkage group are shown in Figures 1 and 2.

253

254 Selection and dominance of viability loci linked to SNPs

255 In Family 1, we detected four regions of the genome with putative viability loci (Fig. 1).
256 In three of these regions (linkage groups 1,6, and 9), Bayesian estimates of *h* and *s* support the
257 existence of viability loci with weakly recessive gene action and large selection coefficients
258 (Table 1, Fig. 1). In the fourth region (linkage group 6), Bayesian estimates of *h* and *s* support
259 the existence of a viability locus with overdominant gene action and a large selection coefficient,

260 though it should be noted that our analysis does not permit distinguishing between
261 overdominance and pseudo-overdominance.
262 In Family 2, we detected two regions of the genome with putative viability loci (Fig. 2)
263 and a third that we were unable to map to any linkage group. In linkage group 5, Bayesian
264 estimates of *h* and *s* support the existence of a viability locus with overdominant gene action and
265 a large selection coefficient, though again it should be noted that our analysis does not permit
266 distinguishing between overdominance and pseudo-overdominance (Table 1, Fig. 2). In linkage
267 group 11, a viability locus with underdominant gene action and a large selection coefficient was
268 inferred (Table 1, Fig. 2). The two genomic regions where these viability loci were detected are
269 in separate linkage groups from those detected in Family 1 (Table S3). The segregation ratio for
270 the unmapped viability locus suggests strong overdominance. Though it is also possible that this
271 is an artifact arising from a locus that maps to duplicated genomic regions as discussed in the
272 Methods section (Table 1).
273
274 The S-locus region of *Leavenworthia alabamica* and locations of viability loci
275 The tightly linked S-locus genes *Lal2* and *SCRL* along with their intervening non-coding
276 region are located on scaffold 46, at base pair positions 2,035,545 to 2,044,046. Several genomic
277 regions on this scaffold in Family 1 showed a strong deviation from the 1:2:1 ratio (Fig. 2 and
278 Table 2). The segregation data here support recessive gene action and strong selection $h = 0.18$ to
279 0.45, and $s = 0.80$ to 0.86 (Table 2). In Family 2, however, this region did not harbour any
280 detectable viability loci.
281
282 Statistical power to detect different modes of gene action
283 Statistical power to detect different modes of gene action in experiments of the type
284 conducted here depends on the value of the recombination fraction *c*. For close linkage, e.g., *c*
285 near 0 as is likely for most SNPs in this study (Figs. 1 and 2, Table S2), overdominant and
286 underdominant action are easier to detect than dominant and recessive action (Fig. 3). As the
287 recombination fraction between marker and inbreeding loci increases, statistical power to detect
288 underdominant selection increases, whereas statistical power to detect overdominant gene action
289 decreases (Fig. 3). For dominant and recessive gene actions, statistical power of detection is
290 reduced with increasing *c*, but compared with overdominant and underdominant action the

291 changes are less affected by recombination fraction. Given a sample size of 90 progeny, close
292 linkage, and overdominant or underdominant genes actions, selection coefficients of *ca*. $s = 0.3$
293 at viability loci are detectable with statistical power of 80%. For dominant and recessive actions,
294 viability loci with selection coefficients of *ca*. $s = 0.6$ are detectable with statistical power of
295 80%.
296
297 **Discussion**
298
299 The genetic architecture of inbreeding depression.
300 We detected a number of significant departures from the 1:2:1 segregation ratio at SNPs
301 located in different regions of the *Leavenworthia alabamica* genome in the progeny of parental
302 plants from two SI populations. The segregation data were used to explore support for different
303 values for selection and dominance coefficients associated with linked viability loci. The
304 approach we used assumes that meiotic drive, gametic selection, and epistasis do not contribute
305 to departures from normal segregation. The inability to distinguish the effect of these factors
306 from viability selection is a limitation of this approach. Nevertheless, there are several reasons to
307 believe that most departures from the 1:2:1 segregation ratio observed in this study are associated
308 with inbreeding depression rather than with meiotic drive or gametic selection. First, where
309 evidence of meiotic drive has been studied in detail in other species (e.g., in *Arabidopsis*
310 *thaliana*), the effects were generally found in specific genomic regions, rather than genome wide
311 as observed in the present study (Salomé *et al.*, 2012; Seymour *et al.*, 2019). Second, gametic
312 selection favouring a specific allele linked to markers is not expected to produce a heterozygote
313 excess at linked marker loci, as observed in this study for some of the marker loci in this study.
314 Several past studies have used marker data to explore the genetic architecture of
315 inbreeding depression in plants. Hedrick & Muona (1990) used the approach in *Pinus sylvestris*
316 and detected a single viability locus with a strong contribution to the fitness of progeny. In
317 another study, using progeny from self-pollinated plants collected from a wild *Mimulus guttatus*
318 population, Fu & Ritland (1994b) applied a graphical approach that divides the space of possible
319 segregation ratios into different modes of gene action. They detected 24 viability loci, with the
320 majority showing underdominance or dominance effects. Mitchell-Olds (1995), studying
321 inbreeding depression in the progeny of parent of an *Arabidopsis thaliana* F1 hybrid between

322 different accessions, reported a single viability locus with overdominance that mapped to
323 chromosome 1. Remington and O'Malley (2000) examined progeny segregation ratios at AFLP
324 loci in selfed *Pinus taeda* plants and detected both recessive and overdominant action at 19
325 embryonic viability loci. In the only other study to apply a Bayesian approach to estimating
326 selection and dominance at loci associated with inbreeding depression, albeit with a different
327 parameterisation that the one employed here, Kärkkäinen *et al*. (1999) in *Arabidopsis lyrata* ssp.
328 *petraea* reported linkages of viability loci to seven separate allozyme loci, mostly with
329 overdominant effects. Finally, Hedrick *et al.* (2016) examined segregation ratios for nearly
330 10,000 SNPs in 28 progeny from a self-pollinated plant of *Eucalytus grandis*, and found
331 widespread heterozygote excess, consistent with overdominance, though their simulation
332 analysis suggested that pseudo-overdominance is the more likely explanation. Overall, these
333 empirical studies suggest that modes of gene action for linked viability loci may vary from
334 species to species.
335 It is notable that a few of the above-cited studies have reported underdominant viability
336 loci linked to markers (e.g., especially Fu & Ritland, 1994b, but also Kärkkäinen *et al*. 1999).
337 Underdominance was also inferred at two marker-linked viability loci in the present
338 investigation. Underdominance at viability loci cannot contribute to inbreeding depression. We
339 note from the results of the power analyses we conducted, underdominance is detectable with
340 smaller sample sizes compared with other modes of action that require larger sample sizes to
341 detect selection of the same magnitude when $c < 0.3$ (Fig. 3). This could help account for the
342 unexpected reports of underdominance in studies that have employed few markers (e.g., Fu &
343 Ritland, 1994b).
344
345 <u>Overdominance or pseudo-overdominance of inbreeding depression</u>
346 For a number of the marker loci showing departures from normal segregation, our
347 analysis suggests overdominance at linked viability loci contributes to inbreeding depression.
348 Alternatively, these results could be explained as being due to pseudo-overdominance (Ohta,
349 1971). Psuedo-overdominance would be detected in a genomic region by the methods used here
350 if deleterious, recessive viability mutations are linked and in repulsion. In general, compared
351 with recessive or partly recessive action at viability loci, there is little support from past studies
352 for true overdominance as an underlying cause of inbreeding depression (Charlesworth and

353 Willis 2009). Distinguishing between true overdominance and pseudo-overdominance is difficult
354 but becomes progressively easier with multiple families. True overdominance, if present, should
355 often be detected at the same viability loci in the progeny of multiple selfed parents, since
356 selection there should render it frequent. In contrast, viability loci exhibiting pseudo-
357 overdominance arising from deleterious recessive or partly recessive alleles are less likely to be
358 shared among parents, as the genomic regions are maintained by the interaction of mutation and
359 selection. This leaves open the possibility that both true and pseudo-overdominance are present.
360

361 Sheltered S-linked load

362 An inferred viability locus (or possibly, multiple linked loci) with recessive action was
363 detected only in one genomic region, and one family (Family 1). Interestingly, the region is
364 within 80 kb from the site of the S-locus. These results can be interpreted as evidence that there
365 is sheltering from selection of one (or more) recessive, deleterious mutations in the S-locus
366 region (Table 2).
367 In Family 2, no such effects were detected in the S-locus region. The finding is not
368 necessarily at odds with the evidence of sheltered load in the Family 1 data. There are likely
369 several levels of dominance for S-locus alleles in *L. alabamica* (Busch *et al.*, 2010), and in
370 sporophytic systems such as this one, the strongest sheltering from selection is expected for those
371 viability loci linked to S-alleles that exhibit the highest degree of dominance (Llaurens *et al.*,
372 2009).
373

374 Inbreeding depression, purging, and the maintenance of self-incompatibility in *Leavenworthia*
375 *alabamica*

376 Busch (2005) found significant levels of inbreeding depression for germination and
377 survival rate, biomass, flower number, and pollen viability in self-incompatible populations of
378 *Leavenworthia alabamica*, and estimated that inbreeding depression for total lifetime fitness
379 exceeds 0.5. Baldwin & Schoen (2019) studied inbreeding depression for a similar set of fitness
380 components in *L. alabamica* and again found that inbreeding depression for lifetime fitness
381 exceeds 0.5. They extended their study to include three consecutive generations of self-
382 pollination and found no evidence that inbreeding depression was purged by this treatment.

383 Inbreeding depression is thought to be the main barrier to the evolution of self-pollination
384 (Fisher, 1941; Lloyd 1979), and yet with some self-pollination, purging of recessive load is
385 expected. In an earlier set of population surveys Baldwin & Schoen (2017) found that some
386 selfing likely does occur in populations of *Leavenworthia alabamica* that have been formerly
387 characterized at SI. Between 6% and 18% of plants exhibit "leaky self-incompatibilty" in which
388 pollen tubes from self-pollen penetrates the styles. Such plants are presumably capable of setting
389 seed with self-pollen. Baldwin & Schoen (2017) also showed that variation in the degree of this
390 leakiness is heritable, and so at least one requirement for purging inbreeding depression (i.e.,
391 heritable variation for level of self-pollination) is likely present in populations of this species.
392 This raises the question of why SI is maintained. Theory suggests that the spread of mutations
393 that reduce SI occurs mainly when inbreeding depression loci are of major effect and recessive
394 (Porcher & Lande, 2005; Gervais *et al*., 2014). The nearly additive viability loci detected in this
395 study are not expected to contribute significantly to standing inbreeding depression
396 (Charlesworth & Charlesworth, 1987). Moreover, any component of inbreeding depression due
397 to loci with overdominance or pseudo-overdominance (as also detected in this study) should be
398 resistant to purging. Given our findings here, we suggest that inbreeding depression in the self-
399 incompatible populations of this species is likely due to many loci of small effect—or in other
400 words, with effects too mild to be detected in the type of study we reort here. Inbreeding
401 depression of this type would be expected to be resistant to purging (Charlesworth &
402 Charlesworth, 1987) as we observed in our earlier work (Baldwin & Schoen, 2019).
403

404 **Acknowledgements**

405 We thank Mark Romer and Claire Cooney for assistance with plant growth in the McGill
406 Phytotron. Yong-Bi Fu provided comments on the manuscript. SJB acknowledges scholarship
407 support from the Fonds de Recherche du Québec – Nature et technologies and support from a
408 Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to DJS
409

410 **Author Contributions**

411 DJS and SJB conceived the study. DJS developed the Bayesian approach for characterizing the
412 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. The authors are joint first authors.

Data Availability

The *Leavenworthia alabamica* reference assembly version 2.5 is available on the CoGe platform: https://genomevolution.org/coge/GenomeInfo.pl?gid=20247. FASTQ files of GBS sequence data for parents and progeny are available on GenBank under SRA accession number PRJNA813545

Supporting Information

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.
TableS3.xlsx	Table S3. Cross reference for linkage groups in Families 1 and 2.
Note 1	R code for Bayesian analysis
Note 2	R code for power analysis in Fig. 3

432

433 **References**

434

435 **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search
436 tool. *Journal of Molecular Biology* **215**: 403-410.
437 **Baldwin SJ, Schoen DJ. 2017.** Genetic variation for pseudo-self-compatibility in self-
438 incompatible populations of *Leavenworthia alabamica* (Brassicaceae). *New Phytologist* **213**:
439 430-429.
440 **Baldwin SJ, Schoen DJ. 2019.** Inbreeding depression is difficult to purge in self-incompatible
441 populations of *Leavenworthia alabamica*. *New Phytologist* **224**:1330-1338.
442 **Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful
443 approach to multiple testing. *Journal of the Royal Statistical Society: Series B*
444 *(Methodological)* **57**: 289-300.
445 **Broman KW, Wu H, Sen Ś, Churchill GA. 2003.** R/qtl: QTL mapping in experimental crosses.
446 *Bioinformatics* **19**: 889-90.
447 **Busch JW. 2005**. Inbreeding depression in self-incompatible and self-compatible populations of
448 *Leavenworthia alabamica*. *Heredity* **94**: 159-165.
449 **Busch JW, Joly S, Schoen DJ. 2010.** Does mate limitation in self-incompatible species promote
450 the evolution of selfing? The case of *Leavenworthia alabamica. Evolution* 64:1657-1670.
451 **Busch JW, Joly S, Schoen DJ. 2011.** Demographic signatures accompanying the evolution of
452 selfing in *Leavenworthia alabamica*. *Molecular Biology and Evolution* **28**: 1717-1729.
453 **Chantha SC, Herman AC, Platts AE, Vekemans X, Schoen DJ. 2013.** Secondary evolution of
454 a self-incompatibility locus in the Brassicaceae genus *Leavenworthia*. *PLoS Biology* **14**:
455 e1001560.
456 **Charlesworth D, Lyons EE, Litchfield LB.** 1994 Inbreeding depression in two highly
457 inbreeding populations of *Leavenworthia*. *Proceedings of the Royal Society of London,*
458 *Series B* **258**: 209-214.
459 **Charlesworth D, Charlesworth B. 1987**. Inbreeding depression and its evolutionary
460 consequences. *Annual Review of Ecology and Systematics* **18**: 237-268.
461 **Charlesworth D, Willis JH. 2009.** The genetics of inbreeding depression. *Nature Reviews*
462 *Genetics* **10**: 783-796

463 **Cohen, J. 1988.** *Statistical power analysis for the behavioral sciences, 2nd ed*. Hillsdale, NJ
464 (USA): Lawrence Erlbaum Associates.
465 **DeRaad DA. 2022**. SNPfiltR: an R package for interactive and reproducible SNP filtering.
466 *Molecular Ecology Resources* (in press).
467 **Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011.**
468 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS*
469 *One* **6**: e19379.
470 **Frankham R. 2010.** Where are we in conservation genetics and where do we need to go?
471 *Conservation Genetics* **11**: 661-3.
472 **Fu YB, Ritland K. 1994a.** On estimating the linkage of marker genes to viability genes
473 controlling inbreeding depression. *Theoretical and Applied Genetics* **88**: 925-932.
474 **Fu YB, Ritland K. 1994b.** Evidence for the partial dominance of viability genes contributing to
475 inbreeding depression in *Mimulus guttatus*. *Genetics* **136**: 323-331.
476 **Gervais C, Awad DA, Roze D, Castric V, Billiard S. 2014.** Genetic architecture of inbreeding
477 depression and the maintenance of gametophytic self-incompatibility. *Evolution* **68**: 3317-
478 3324.
479 **Glémin S. 2003**. How are deleterious mutations purged? Drift versus nonrandom mating.
480 *Evolution* **57**: 2678-2687.
481 **Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-**
482 **Lopez Z, Steffen JG, Hazzouri KM, Dewar K. 2013.** An atlas of over 90,000 conserved
483 noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics*
484 **45**:891-898.
485 **Hedrick PW, Muona O. 1990.** Linkage of viability genes to marker loci in selfing organisms.
486 *Heredity* **64**: 67-72.
487 **Hedrick PW, Hellsten U, Grattapaglia D. 2016.** Examining the cause of high inbreeding
488 depression: analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus
489 grandis. *New Phytologist* **209**: 600-611.
490 **Herman AC, Busch JW, Schoen DJ. 2012.** Phylogeny of Leavenworthia S-alleles suggests
491 unidirectional mating system evolution and enhanced positive selection following an ancient
492 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.

Li H. and Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589-595.

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.

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.

524

525
526 **Figure Legends**
527
528 **Figure 1.** -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 solid and dashed lines indicate
530 significance levels corresponding to the false discovery rate of q = 0.05 and 0.10, respectively.
531 Below are the segregation ratios for the SNPs. Blue denotes heterozygote, red and black denote
532 homozygotes. Above are heat maps of posterior probability densities for dominance (h) and
533 selection (s) coefficients of inferred viability loci in linkage groups (positions indicated by
534 arrows).
535
536 **Figure 2.** -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 solid and dashed lines indicate
538 significance levels corresponding to the false discovery rate of q = 0.05 and 0.10, respectively.
539 Below are the segregation ratios for the SNPs. Blue denotes heterozygote, red and black denote
540 homozygotes. Above are heat maps of posterior probability densities for dominance (h) and
541 selection (s) coefficients of inferred viability loci in different linkage groups (positions indicated
542 by arrows).
543
544 **Figure 3.** Statistical power to reject the 1:2:1 segregation ratio for different modes of action of
545 viability loci, selection coefficients, and recombination fractions. Shown are four cases:
546 overdominant with h = -3.0 (red), recessive with h = 0.25 (black), dominant (blue) with h = 0.75,
547 and underdominant (h = 1.5) (orange). From top to bottom graphs, the values of the
548 recombination fraction are c = 0, 0.1, 0.2, and 0.3. The graphs are based on a sample size of 90
549 progeny.
550

Fig. 1

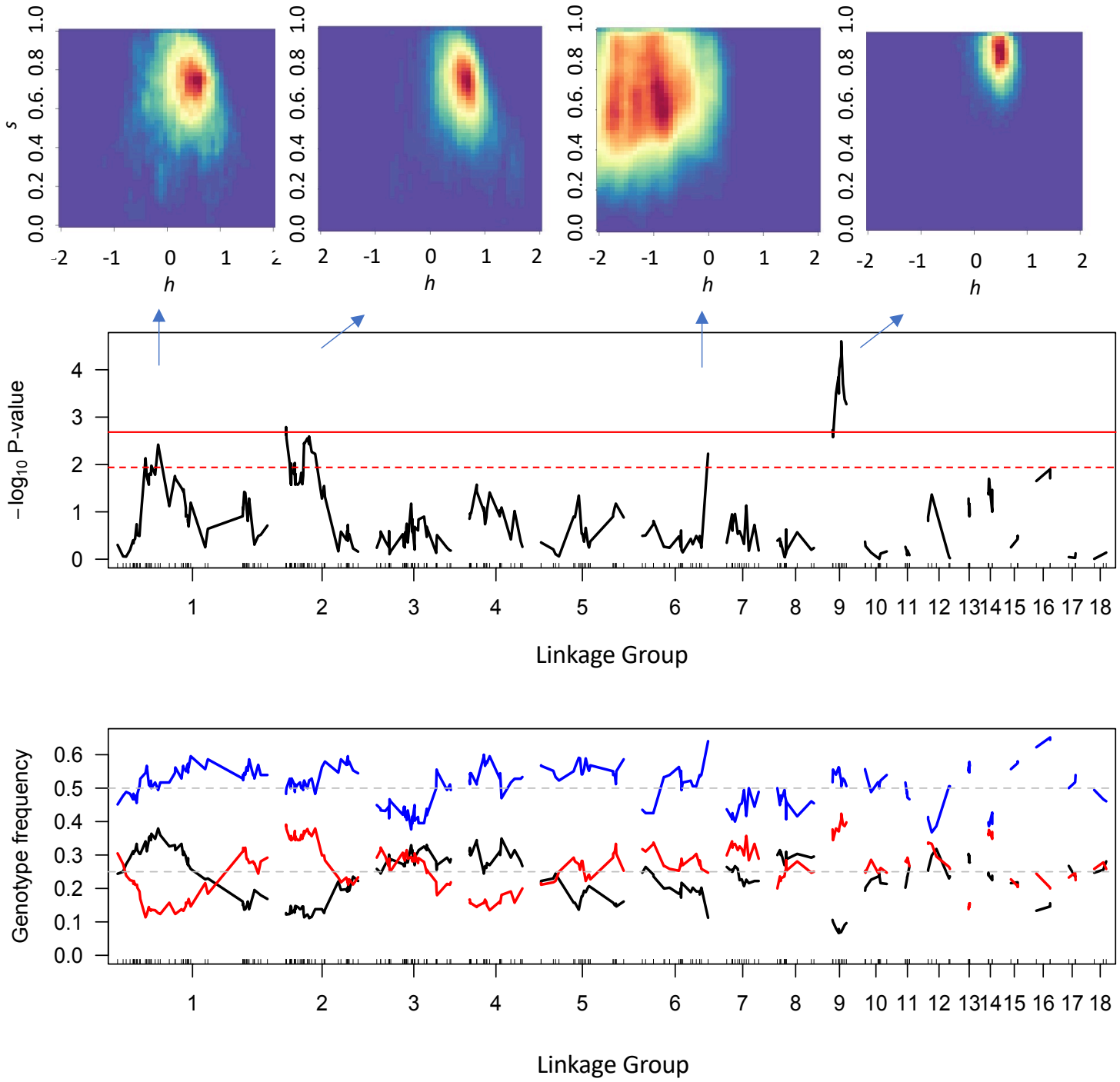
s
h
−log10 P-value
Linkage Group
Genotype frequency
Linkage Group

Fig. 2

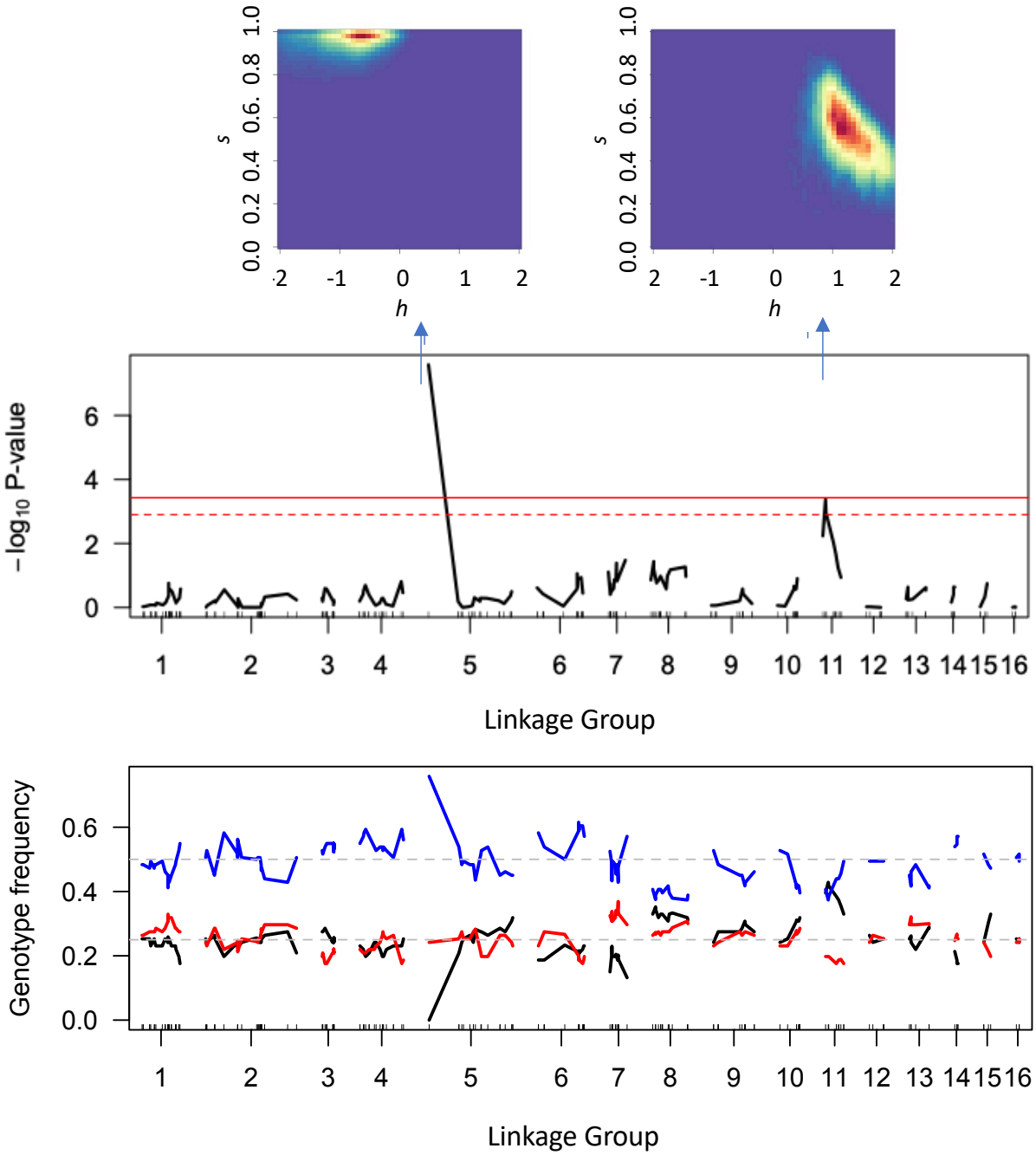
s
h
0.0 0.2 0.4 0.6 0.8 1.0
-2 -1 0 1 2
$-\log_{10}$ P-value
0 2 4 6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Linkage Group
Genotype frequency
0.0 0.2 0.4 0.6
Linkage Group

Fig. 3

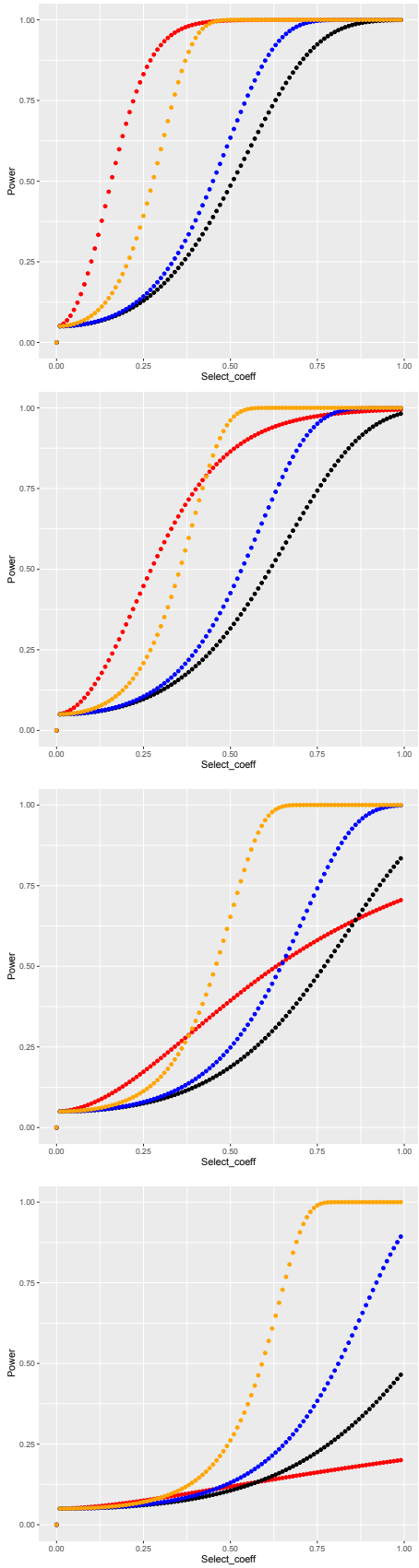

Table 1. Marker loci that show the most significant deviations from normal segregation within separate linkage groups. Marker locus (scaffold_base pair position), linkage group (LG), and posterior probabilities of overdominant (OD), recessive (R), dominant (D), and underdominant (UD) gene actions[1]. The values for the selection coefficient (*s*) and dominance coefficient (*h*) are the Bayesian estimators (95% credible intervals).

Family	Marker	LG	M_1M_1	M_1M_2	M_2M_2	OD	R	D	UD	*s*	*h*
1	61_934679	1	11	43	33	0.11	0.36	0.45	0.07	0.67 (0.34, 1.0)	0.47 (-0.33, 1.14)
1	57_279929	2	11	42	34	0.15	0.38	0.41	0.06	0.68 (0.28, 1.0)	0.42 (-0.45, 1.08)
1	28_130518	6	10	57	22	0.99	0.01	0	0	0.61 (0.23, 0.97)	-1.82 (-3.89, -0.16)
1	46_1956910	9	6	43	36	0	0.51	0.44	0.01	0.86 (0.68, 0.98)	0.45 (0.01, 0.79)
2	27_153823	5	0	69	22	0.99	0.01	0	0	0.91 (0.63, 1.0)	-1.55 (-4.28, -0.23)
2	8_243257	11	18	34	39	0	0.01	0.20	0.79	0.50 (0.20, 0.73)	1.38 (0.74, 2.31)
2	197_101568	-	0	85	0	-	-	-	-	-	-

[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.

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