



**HAL**  
open science

# The genetic architecture of the load linked to dominant and recessive self-incompatibility alleles in *Arabidopsis halleri* and *A. lyrata*

Le Veve Audrey, Genete Mathieu, Lepers-Blassiau Christelle, Ponitzki Chloé, Céline Poux, Vekemans Xavier, Durand Eleonore, Castric Vincent

## ► To cite this version:

Le Veve Audrey, Genete Mathieu, Lepers-Blassiau Christelle, Ponitzki Chloé, Céline Poux, et al.. The genetic architecture of the load linked to dominant and recessive self-incompatibility alleles in *Arabidopsis halleri* and *A. lyrata*. 2023. hal-04306051

**HAL Id: hal-04306051**

**<https://hal.science/hal-04306051v1>**

Preprint submitted on 24 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Main text**

2

3 **The genetic architecture of the load linked to dominant and recessive self-**  
4 **incompatibility alleles in *Arabidopsis halleri* and *A. lyrata***

5

6 **Le Veve Audrey<sup>1,2</sup>, Genete Mathieu<sup>1</sup>, Lepers-Blassiau Christelle<sup>1</sup>, Ponitzki Chloé<sup>1</sup>, Céline**  
7 **Poux<sup>1</sup>, Vekemans Xavier<sup>1</sup>, Durand Eleonore<sup>1</sup>, Castric Vincent<sup>1\*</sup>**

8 <sup>1</sup>*Univ. Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France*

9 <sup>2</sup>*Current address: Department of Botany, Faculty of Science, Charles University, Benátská 2,*  
10 *CZ-128 01 Prague, Czechia*

11 \* *Correspondence : [vincent.castric@univ-lille.fr](mailto:vincent.castric@univ-lille.fr)*

12 **Abstract**

13 The long-term balancing selection acting on mating-types or sex determining genes is expected to  
14 lead to the accumulation of deleterious mutations in the tightly linked chromosomal segments that  
15 are locally “sheltered” from purifying selection. However, the factors determining the extent of this  
16 accumulation are poorly understood. Here, we take advantage of the large number of alleles at the  
17 sporophytic self-incompatibility system of the Brassicaceae along a complex dominance hierarchy to  
18 evaluate how the pace at which linked deleterious mutations accumulate varies with the intensity of  
19 balancing selection. We first experimentally measured the phenotypic manifestation of the linked  
20 load at three different levels of the dominance hierarchy. We then sequenced and phased  
21 polymorphisms in the chromosomal regions linked to 126 distinct copies of *S*-alleles in two  
22 populations of *Arabidopsis halleri* and three populations of *A. lyrata*. We find that linkage to the *S*-  
23 locus locally distorts phylogenies over about 10-30kb along the chromosome. The more intense  
24 balancing selection on dominant *S*-alleles results in greater fixation of deleterious mutations, while  
25 recessive *S*-alleles accumulate more linked deleterious mutations that are segregating. Hence, the  
26 structure rather than the overall magnitude of the linked genetic load differs between dominant and  
27 recessive *S*-alleles. Our results have consequences for the long-term evolution of new *S*-alleles, the  
28 evolution of dominance modifiers between them, and raise the question of why the non-  
29 recombining regions of some sex and mating type chromosomes expand over evolutionary times  
30 while others, such as that the *S*-locus of the Brassicaceae, remain restricted to small chromosomal  
31 regions.

32 *Keywords: Supergene, sheltered load, mating types, balancing selection, genetic dominance, S-locus.*

## 33 Introduction

34 Sexual reproduction typically involves mating between individuals that belong to separate groups of  
35 reciprocal reproductive compatibility. This partition among conspecific individuals is obvious in  
36 species where the male and female reproductive functions are performed by distinct categories of  
37 individuals, but can also be present in purely hermaphroditic species, where compatibility among  
38 individuals is governed by the segregation of various numbers of mating types in the total absence of  
39 sexual specialisation. The existence of sexes or mating-types leads to one the strongest forms of  
40 long-term balancing selection, and is often associated with clusters of polymorphisms around  
41 sex/mating-type controlling regions kept together by structural rearrangements. In some cases, such  
42 rearrangements can span almost entire chromosomes (*e.g.* sex chromosomes in mammals<sup>1</sup> or  
43 mating-type chromosomes in ascomycete fungi<sup>2</sup>), while in others they remain limited to relatively  
44 small genomic regions (*e.g.* chromosomal inversions controlling male reproductive morphs in the  
45 ruff<sup>3</sup>, mating-type loci in some basidiomycete fungi, segregating indels controlling pin vs. thrum  
46 floral morphs in *Primula*<sup>4</sup>). The long-term balancing selection acting on these systems is expected to  
47 lead to the accumulation of deleterious mutations in the tightly linked chromosomal segments that  
48 are “sheltered” from purifying selection by the presence of the balanced polymorphism<sup>5,6</sup>. These  
49 deleterious mutations can have drastic short- and long-term consequences for the evolution of the  
50 species, and determining the processes by which they accumulate is crucial to understand how the  
51 rearranged regions can either expand along the chromosomes or conversely remain restricted to  
52 limited genomic tracts<sup>7,8</sup>.

53 Self-incompatibility (SI) is a genetic mechanism allowing recognition and rejection of self-pollen by  
54 hermaphrodite individuals, thereby preventing inbreeding and promoting outcrossing in  
55 hermaphroditic plant species<sup>9</sup>. In the Brassicaceae family, SI is controlled by a single non-  
56 recombining chromosomal region, the *S*-locus<sup>10,11</sup>. SI is one of the most prominent examples of long-  
57 term balancing selection<sup>12</sup>, and as such deleterious mutations are expected to accumulate in very  
58 close genetic linkage to the *S*-alleles because of the indirect effects of linked selection<sup>6</sup>. Population  
59 genetics models predict that deleterious variants should accumulate within specific *S*-allele  
60 lineages<sup>12,13</sup>, and should then be reshuffled among them by recombination. However, due to the  
61 technical difficulty of phasing polymorphisms, this process has rarely been characterised in detail<sup>14</sup>.

62 A key feature of sporophytic SI systems, also shared by sex chromosomes, is the existence of  
63 dominance interactions between *S*-alleles. While most individuals are heterozygous at the *S*-locus  
64 and thus carry two different *S*-alleles, only one of them is generally expressed at the phenotypic  
65 level, especially for the pollen specificity, following a complex genetic dominance hierarchy<sup>15,16</sup>.  
66 Evolutionary properties of *S*-alleles vary along the dominance hierarchy because balancing selection  
67 is predicted to act more strongly on dominant than on recessive *S*-alleles, as the latter are often  
68 masked at the phenotypic level<sup>17</sup>. As a result, Llaurens et al.<sup>13</sup> and Goubet et al.<sup>18</sup> suggested that the  
69 dynamics of accumulation of deleterious variation may differ in close linkage with dominant vs  
70 recessive *S*-alleles, similarly to Y vs. X chromosomes, respectively. Specifically, recessive *S*-alleles can  
71 form homozygous combinations in natural populations more often than dominant *S*-alleles<sup>17</sup>, such  
72 that recombination can occur occasionally between distinct gene copies of the same recessive *S*-  
73 allele, providing the opportunity for the linked recessive deleterious mutations to be purged from  
74 within the *S*-locus itself. In addition, because recessive *S*-alleles reach higher population

75 frequencies<sup>15,17</sup>, purifying selection on linked deleterious variants is expected to have higher efficacy  
76 among gene copies of recessive than dominant *S*-alleles. This is expected to result in a higher fixation  
77 probability of deleterious variants linked to the class of dominant *S*-alleles than to the class of  
78 recessive *S*-alleles<sup>13</sup>. Empirical support for this simple prediction has been conflicting, though. Based  
79 on phenotypic measurements in *A. halleri*, Llaurens et al.<sup>13</sup> observed a decrease of fitness associated  
80 by enforced homozygosity for one of the most dominant *S*-alleles (Ah15) but not for the most  
81 recessive *S*-allele of the allelic series (Ah01). In contrast, Stift et al.<sup>19</sup> observed no effect of  
82 dominance on the genetic load linked to three dominant vs. recessive *S*-alleles in a natural  
83 population of the closely related *A. lyrata*. Hence, the data available so far are inconclusive, but are  
84 restricted to very small numbers of *S*-alleles. They are also based on inherently limited phenotypic  
85 measurements, seriously limiting the power of the comparisons, and preventing proper  
86 generalisation of the effect of dominance on the accumulation of linked deleterious variation.

87 In this study, we combined phenotypic, genomic and theoretical approaches to finely dissect the  
88 patterns of accumulation of deleterious variation linked to the *S*-locus supergene in *A. halleri* and *A.*  
89 *lyrata*. We first extended the phenotypic approach of Llaurens et al.<sup>13</sup> to a series of additional *S*-  
90 alleles from the same local *A. halleri* population to evaluate the effect of *S*-allele dominance on the  
91 sheltered load. We then used parent-offspring trios and targeted genome re-sequencing to directly  
92 quantify the accumulation of putative deleterious mutations linked to phased dominant vs. recessive  
93 *S*-alleles in two *A. halleri* and three *A. lyrata* natural populations. Finally, we used stochastic  
94 simulations to refine the theoretical predictions about the effect of *S*-allele dominance on the  
95 dynamics of linked deleterious mutations. Overall, our results provide a more nuanced view of the  
96 effect of the intensity of balancing selection on the sheltered load, in which the structure of the  
97 sheltered load rather than its magnitude differs among *S*-alleles from different dominance classes.

## 98 **Results**

99 *The genetic load linked to the S-locus varies among S-alleles, but is not correlated with dominance.*

100 We first expanded the experimental approach of Llaurens et al.<sup>13</sup> to phenotypically evaluate the  
101 effect of *S*-allele dominance on the intensity of the sheltered load. The previous study focused on  
102 three *S*-alleles (Ah01, Ah02 and Ah15)<sup>13</sup>. Here we included two *S*-alleles from the same local  
103 population (Nivelle, France): Ah03 and Ah04, and included Ah01 again for comparative purposes. In  
104 the Arabidopsis genus, *S*-alleles have been shown to form a complex dominance hierarchy<sup>15,16</sup>. This  
105 hierarchy is largely associated with the phylogeny of *S*-alleles<sup>15</sup>, and at least four phylogenetic classes  
106 (I, II, III and IV) have been described, from the most recessive (class I) to the most dominant of *S*-  
107 alleles (class IV). Dominance interactions also exist among *S*-alleles within classes, such that these  
108 five *S*-alleles form the following dominance hierarchy<sup>15,16</sup>: Ah01<Ah03<Ah02<Ah04<Ah15, from the  
109 most recessive (Ah01) to the most dominant (Ah15). To reveal the linked load, we enforced  
110 homozygosity at the *S*-locus using controlled crosses between parental individuals sharing a given *S*-  
111 allele that was masked by different dominant *S*-alleles (e.g., to obtain Ah<sub>x</sub>Ah<sub>x</sub> homozygotes we  
112 deposited pollen from a Ah<sub>x</sub>Ah<sub>y</sub> plant, where Ah<sub>y</sub>>Ah<sub>x</sub>, on pistils of a Ah<sub>x</sub>Ah<sub>z</sub> plant where z≠y, or on a  
113 Ah<sub>x</sub>Ah<sub>x</sub> pistil when available; table S1). We obtained 399 offspring from a total of six such crosses.  
114 Note that our experimental procedure differs slightly from that of Llaurens et al.<sup>13</sup> in that their  
115 procedure required a CO<sub>2</sub> treatment to bypass the SI system and obtain selfed offspring, while here  
116 we took advantage of the dominance interactions to obtain outcrossed *S*-locus homozygous

117 individuals that we phenotypically compared to their full-sibs with *S*-locus heterozygous genotypes.  
118 Note also that the *S*-locus homozygous offspring we obtained contain distinct gene copies of a given  
119 *S*-allele lineage. Hence, they could in principle carry distinct suites of linked deleterious mutations in  
120 case these mutations segregate within *S*-allele lineages.

121 We first tested whether homozygosity at the *S*-locus affected survival by measuring for each cross  
122 the proportion of homozygotes at the *S*-locus reaching the reproductive stage for three *S*-alleles (in  
123 two replicate families per *S*-allele, Table S1). The proportion of Ah01/Ah01 and Ah04/Ah04  
124 homozygotes surviving to the reproductive stage was consistent with mendelian expectations in  
125 their respective families. However, we observed a significant decrease of Ah03/Ah03 homozygotes  
126 at the reproductive stage compared with Mendelian expectations (Table 1), whereas the observed  
127 proportion of the Ah03 *S*-allele among heterozygous individuals did not depart from expectations  
128 ( $2/3=0.67$ ; Table 1). Thus, the increased mortality is associated with Ah03 homozygosity, rather than  
129 with a lower performance of individuals carrying the Ah03 *S*-allele itself. Overall, a genetic load was  
130 thus observed linked to the Ah03 *S*-alleles, which is at an intermediate level of dominance, but  
131 neither to the most dominant (Ah04) nor to the most recessive (Ah01) *S*-allele. Hence, these  
132 observations do not support a positive correlation between *S*-allele dominance and the magnitude  
133 of the sheltered load.

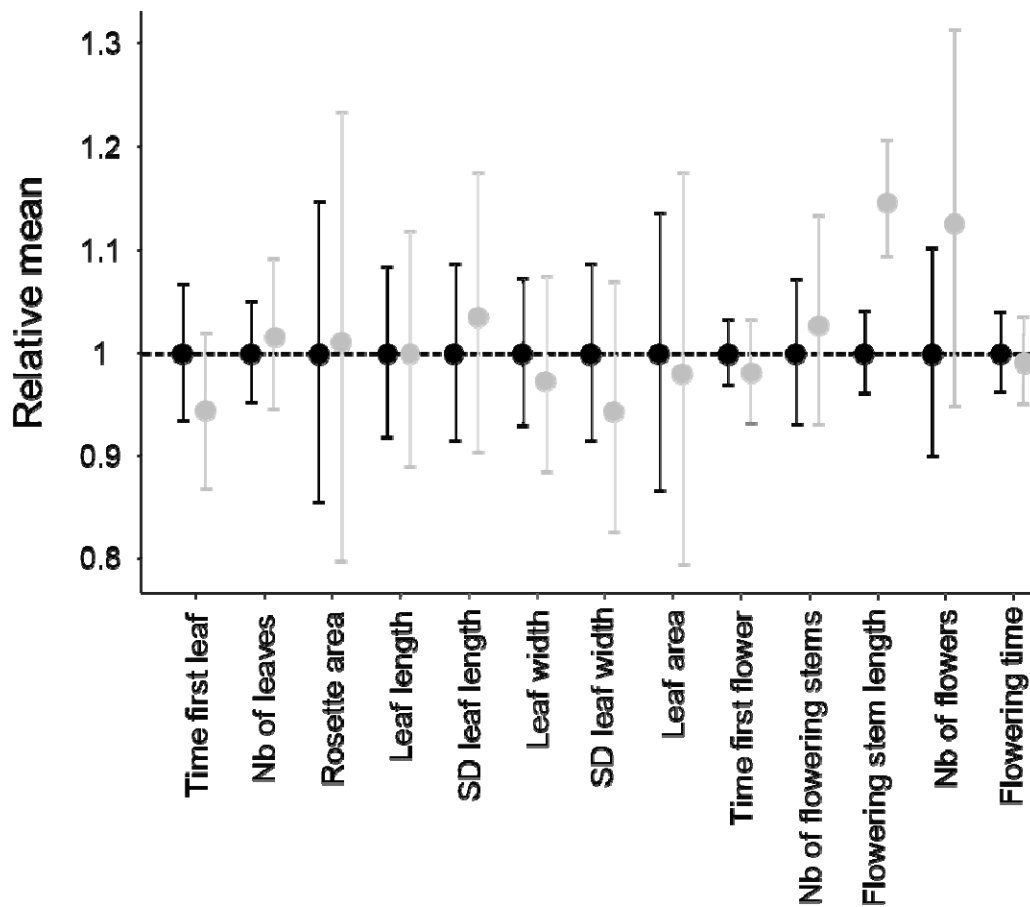
134 **Table 1: Proportion of *S*-locus homozygous offspring having reached the reproductive stage for**  
135 **three different *S*-alleles.** *The test is performed relative to the expected proportion of homozygous*  
136 *genotypes in the offspring (25% when both parents are heterozygous; 50% when one of the parents*  
137 *is homozygous and the other heterozygous).*

<i>S</i> -allele	Level of dominance	Number of seedlings having reached the reproductive stage	Observed proportion of homozygotes	Ratio of the observed to expected proportion of homozygous genotypes ( <i>p</i> -values <sup>a</sup> )	Observed proportion of heterozygotes with the <i>S</i> -allele ( <i>p</i> -values compared to expected)
Ah01	I	35	0.29	1.14 (0.70; 0.76)	19 (0.40; 0.90)
Ah03	II	27	<b>0.074</b>	<b>0.3 (0.04; 0.02)</b>	17 (1; 0.69)
Ah04	III	96	0.479	0.96 (0.75; 0.39)	50

138 <sup>a</sup> *p*-values were obtained by binomial test and after 10,000 random permutations, respectively.  
139 Values departing from Mendelian expectations are figured in bold. For Ah04, the maternal parent  
140 was Ah04/Ah04, so all heterozygous offspring carried the *S*-allele.

141 Next we measured thirteen vegetative and reproductive traits in the resulting families and compared  
142 offspring that were homozygous for their *S*-alleles with their full sibs that were heterozygous (Fig. 1).  
143 We first used permutations to test whether the mean trait value of homozygotes differed from that  
144 in heterozygotes. Overall, with a single exception, we found no effect of homozygosity at the *S*-locus  
145 on variation of the traits measured (Fig. 1; Table S2). The maximum length of flowering stems was  
146 the exception to this general pattern, with longer reproductive stems for *S*-locus homozygous than  
147 heterozygous genotypes, hence in the opposite direction from our expectation of lower fitness in

148 homozygotes. For this trait, there was significant variation among replicate families for homozygotes  
149 of the recessive allele Ah01 but not of the dominant allele Ah04 (Table S3). We then used  
150 generalised linear models (GLM) to evaluate the effect of dominance (as a fixed effect) on the mean  
151 phenotypic value of homozygotes compared to heterozygotes for each trait (Table S4; treating  
152 family of origin, attacks by phytopathogens, phytophagous and oxidative stress as random effects  
153 whenever necessary). We also observed no effect of *S*-allele dominance on the contrast between *S*-  
154 locus homozygotes and heterozygotes for any of these traits. A single of the thirteen traits was an  
155 exception to this general pattern, but again the effect was in the opposite direction from our  
156 expectation, with an earlier rather than delayed appearance of the first leaf for homozygotes of  
157 more dominant *S*-alleles; Table S4). Overall, our phenotypic results confirmed the presence of a  
158 detectable linked load on some phenotypic traits (survival; time to produce the first leaf), but we  
159 could not replicate the observation of Llaurens et al.<sup>13</sup> that dominant *S*-alleles carry a more severe  
160 deleterious load than recessive *S*-alleles, even though our samples were obtained from the same  
161 local population.



162

163 **Figure 1: Effect of homozygosity at the *S*-locus on 13 phenotypic traits compared to heterozygotes.**  
164 For each trait, the phenotypic values in homozygotes (in grey) were normalised relative to the mean  
165 phenotypic values in heterozygotes (in black). The distributions were obtained by 10,000 random  
166 permutations. SD: Standard Deviation.

167 *S*-alleles are associated with specific sets of tightly linked mutations.

168 The model of the sheltered load assumes that distinct *S*-allele lineages carry specific sets of linked  
169 deleterious mutations, but to our knowledge this prediction was never tested directly. We combined  
170 a parent-offspring trio approach with sequencing of the *S*-locus flanking regions to phase the  
171 mutations segregating in the *S*-locus flanking regions with their respective *S*-alleles. Briefly, we used  
172 a previously developed sequence capture protocol specifically targeting the nucleotide sequences  
173 over 75 kb on each side of the *S*-locus along with a series of 100 control regions from throughout the  
174 genome<sup>20</sup>, and we analysed nucleotide sequence polymorphism (including only invariant and biallelic  
175 SNPs), based on the *A. lyrata* reference genome<sup>21</sup>. We define a haplotype as a unique combination  
176 of mutations along the phased chromosome, and a *S*-allele lineage as the collection of gene copies of  
177 a given functional *S*-allele (different functional *S*-alleles are distinguished based on their strong  
178 sequence divergence at the *S*-locus pollen and pistil genes). Different gene copies within an *S*-allele  
179 lineage can thus be associated with distinct linked haplotypes in the flanking regions. The *S*-alleles  
180 were identified based on short reads sequences according to a previously published method<sup>41</sup>. We  
181 analysed two closely related *A. halleri* populations from Europe (Nivelle and Mortagne) and three  
182 allogamous *A. lyrata* populations from North America (IND, PIN and TSS<sup>22</sup>). Overall, we were able to  
183 reconstruct 34 haplotypes linked to a total of 12 distinct *S*-allele lineages in Nivelle, 38 haplotypes  
184 linked to 11 distinct *S*-allele lineages in Mortagne and 16, 22 and 16 haplotypes associated with 6, 7  
185 and 5 distinct *S*-allele lineages in populations IND, PIN and TSS, respectively (Table S5). Nine of the *S*-  
186 alleles were shared between the two *A. halleri* populations (Ah01, Ah03, Ah04, Ah05, Ah12, Ah20,  
187 Ah24, Ah25 and Ah59). In the populations of *A. lyrata*, four *S*-alleles were shared between PIN and  
188 TSS (Ah01\*, Ah03\*, Ah18\* and Ah63\*), five *S*-alleles were shared between PIN and IND (Ah01\*,  
189 Ah03\*, Ah46\* and Ah63\*), four *S*-alleles were shared between IND and TSS (Ah01\*, Ah03\*, Ah31\*  
190 and Ah63\*), and three were shared across all three (Ah01\*, Ah03\* and Ah63\*). Note that for  
191 convenience, we used *A. halleri* notations (with the addition of a \*) to refer to the trans-specifically  
192 shared *A. lyrata* *S*-alleles. Altogether, we were able to obtain the phased flanking sequences of 126  
193 *S*-locus haplotypes, comprising a total of 4,854 variable sites. This provides considerable power to  
194 evaluate the local accumulation of linked mutations across *S*-alleles of different levels of dominance  
195 and to examine their patterns of conservation between populations and between species.

196 Mutations in the *S*-locus flanking regions can be exchanged between *S*-alleles by recombination, and  
197 between local populations by migration<sup>23</sup>. The relative time scale of these two processes  
198 (recombination vs. migration) determines the distribution of the linked mutations. To capture the  
199 chromosomal extent of this effect of linkage to *S*-alleles, we developed a new phylogenetic method  
200 comparing the likelihood of two contrasted topologies of interest in overlapping windows along the  
201 chromosome: (1) the topology clustering haplotypes by the populations where they came from vs.  
202 (2) the topology clustering them by the *S*-allele to which they are linked (Fig 2A). This allowed us to  
203 evaluate the progressive shift from a predominant topology by *S*-alleles close to the *S*-locus to a  
204 topology by populations further along the chromosome and in unlinked control regions (Fig. 2B). The  
205 difference in log likelihood between the two topologies decreased significantly with distance to the  
206 *S*-locus (Pearson coefficient = -0.015 and -0.010 for *A. halleri* and *A. lyrata* respectively;  $p$ -values  $< 2 \times 10^{-6}$   
207 <sup>16</sup>). In *A. halleri*, the topology grouping haplotypes by populations became more likely than the  
208 topology grouping them by *S*-alleles at a distance of around 30kb from the *S*-locus, but even at a  
209 distance of 50kb the phylogenetic structure was still different from that in regions unlinked to the *S*-  
210 locus used as controls for the genomic background<sup>20</sup> (Fig 2B). In *A. lyrata*, the shift was even more  
211 rapid (within 10-15kb), although we note that the phylogenetic structure of the control regions was

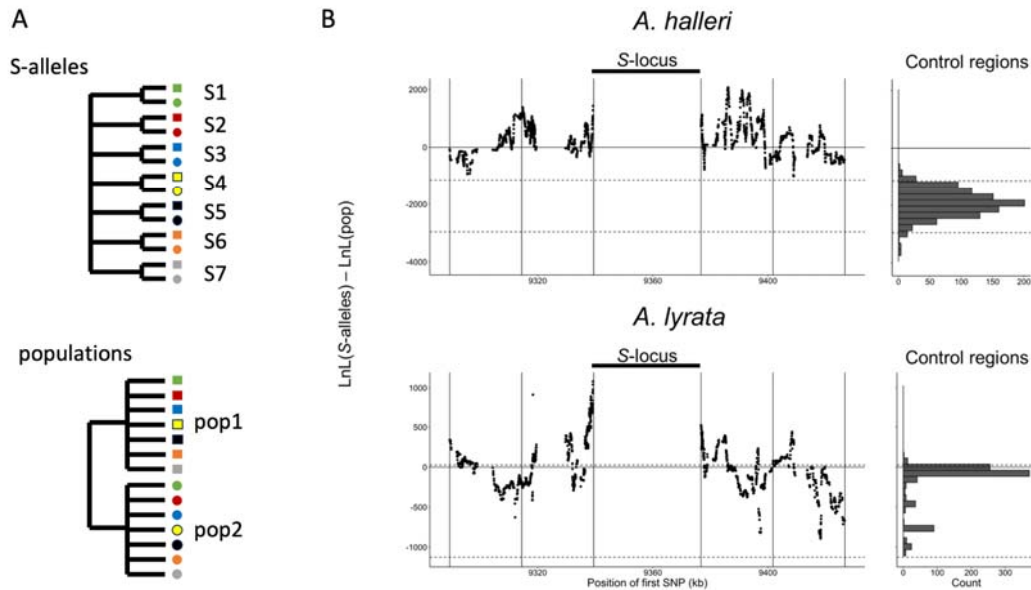
212 less resolved (Fig 2B). To evaluate these patterns more directly we first examined the data using a  
213 Major Component Analysis (MCA, a modified version of PCA adapted to binary data, Fig S1 and S2)  
214 and using simple phylogenetic reconstructions (Fig S3, S4, S5 and S6). We confirmed that haplotypes  
215 linked to a given *S*-allele tended to cluster together in the most tightly linked region, and that this  
216 grouping by *S*-alleles was progressively lost in favour of a grouping by population of origin in the  
217 most distant regions. Following Kamau et al.<sup>24</sup>, we compared the fixation index  $F_{ST}$  among local  
218 populations and among *S*-alleles in *A. lyrata* and *A. halleri*. In both *A. halleri* and *A. lyrata*,  $F_{ST}$  values  
219 among *S*-alleles were high in regions close to the *S*-locus and quickly decreased to reach the  
220 background level (Fig. S7) as the distance from the *S*-locus increased. In parallel, the differentiation  
221 among populations followed roughly the opposite pattern, i.e. it was initially low in regions close to  
222 the *S*-locus (as expected under strong balancing selection) and increased up to background level  
223 within the first few kilobases (Fig. S7). In line with our phylogenetic analysis, differentiation between  
224 populations started to exceed differentiation between *S*-alleles much closer to the *S*-locus in the *A.*  
225 *lyrata* than in the *A. halleri* populations (Fig S3, S4, S5 and S6). Finally, we explored the fine-scale  
226 patterns of association within populations between individual *S*-alleles and SNP in the linked and the  
227 control regions (Fig S8). As expected, the vast majority of significant associations were found for the  
228 most closely linked SNPs. With a single exception, all *S*-alleles were associated with unique SNPs in  
229 the 50kb region around the *S*-locus, albeit with substantial heterogeneity among *S*-alleles in the  
230 patterns and extent of associations that they show (Fig S8). Overall, our results indicate that due to  
231 limited recombination, the *S*-alleles carry a specific set of polymorphic sites in the linked region. This  
232 association fades away for more distant sites over a few kilobases, where population structure  
233 becomes predominant, as in the rest of the genome. Hence, different *S*-alleles are associated with  
234 specific sets of tightly linked mutations, but only within 10-30kb.

235

236

237





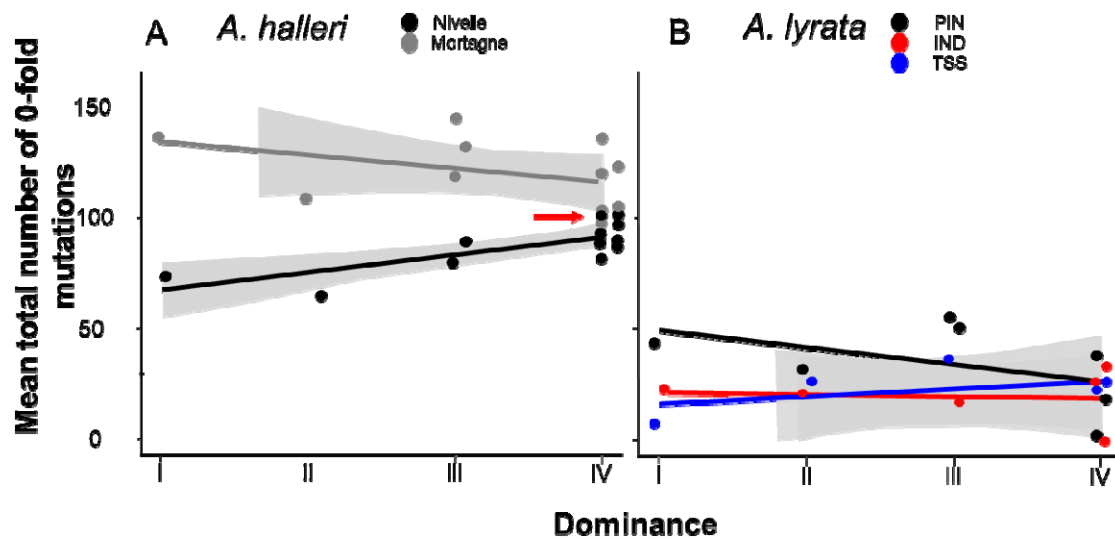
238

239 **Figure 2: Linkage to the S-locus locally distorts the phylogenetic relationships.** A. The two  
240 topologies of interest cluster haplotypes either by the S-allele to which they are linked (top) or by the  
241 populations where they came (bottom). Different S-alleles are represented by symbols of different  
242 colours, different populations of origin are represented by symbols of different shapes. B. Difference  
243 in log likelihood of the two topologies of interest. Dots correspond to the difference in log likelihood  
244 for overlapping series of 50 SNPs around the S-locus for *A. halleri* (top panel) and *A. lyrata* (bottom  
245 panel). Positive values correspond to chromosomal positions where the topology by S-alleles explains  
246 the phylogeny of haplotypes better than the topology by populations. The right panels show the  
247 difference in log likelihood in the control regions. 2.5 and 97.5 percentiles of the distribution in the  
248 control regions are indicated by dashed lines.

249 No overall evidence that dominant S-alleles accumulate more linked deleterious mutations.

250 Llaurens et al.<sup>13</sup> predicted a positive correlation between the dominance of S-alleles and their  
251 tendency to fix linked deleterious mutations. Thus, we investigated the correlation between the  
252 level of dominance of the S-alleles and their total number of 0-fold degenerate mutations ( $S_{0f}$ ) or the  
253 ratio of 0-fold to 4-fold mutations ( $S_{0f}/S_{4f}$ ) for the phased haplotypes, assuming that the vast majority  
254 of 0-fold degenerate mutations are deleterious. Based on the results presented above and the  
255 results of our previous study<sup>20</sup>, for the rest of our analyses we focused on the phased haplotypes  
256 over the first 25 kb on either side of the S-locus. We found no overall effect of dominance on  $S_{0f}$  (p-  
257 values= 0.54 and 0.07 for *A. halleri* and *A. lyrata* respectively; Fig. 3; Table S6) or  $S_{0f}/S_{4f}$  (p-values=  
258 0.54 and 0.07 for *A. halleri* and *A. lyrata* respectively; Table S6). Extending the analysis to all non-  
259 synonymous mutations led to identical conclusions (Table S6). Overall, our genomic results did not  
260 confirm the prediction that dominant S-alleles accumulate a larger number of putatively deleterious  
261 mutations in their linked regions. We note that the particular S-allele whose sheltered load was  
262 quantified in Llaurens et al.<sup>13</sup> (Ah15, red arrow on Fig 3A) appears to be one of the S-alleles  
263 associated with the highest number of 0-fold degenerate mutations among all S-alleles of the most  
264 dominant class (class IV).

265

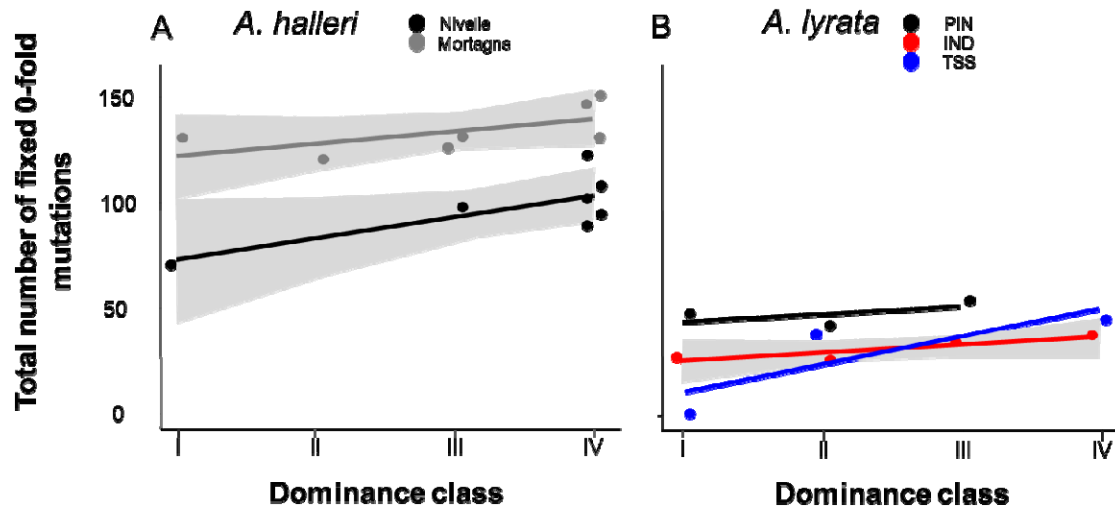


266

267 **Figure 3: No overall effect of S-allele dominance on the total number of 0-fold degenerate**  
268 **mutations ( $S_{0f}$ ) in the linked genomic regions within 25kb.** Each dot represents the mean number of  
269 mutations observed among haplotypes linked to one S-allele in one population. The correlations  
270 evaluated by a GLM are represented by lines, with confidence intervals represented in grey. A:  
271 *A. halleri*. Black dots correspond to the Nivelle population, grey dots to the Mortagne population. The  
272 red arrow points to the copy of Ah15, corresponding to the S-allele whose sheltered load was  
273 phenotypically characterised by Llaurens et al.<sup>13</sup> in Nivelle. B: *A. lyrata*. Red dots correspond to the  
274 IND population, black dots to the PIN population and blue dots to the TSS population.

275 *The structure of the linked genetic load differs between dominant and recessive S-alleles.*

276 Theory predicts that dominant S-alleles should fix linked recessive deleterious mutations with a  
277 higher probability than recessive S-alleles<sup>13</sup>, but in natural populations we observed no difference in  
278 the total number of putatively deleterious linked to dominant vs. recessive S-alleles. To clarify this  
279 discrepancy, we took advantage of our sequencing of multiple copies of S-alleles to consider  
280 separately the fixed and the segregating mutations linked to each of the S-allele lineages. For each  
281 population, we included only mutations that were segregating, and excluded those that were locally  
282 fixed. In agreement with the prediction of Llaurens et al.<sup>13</sup>, we observed that lineages of dominant S-  
283 alleles do indeed tend to fix deleterious mutations more readily (Fig. 4). The fact that they do not  
284 accumulate a larger total number of deleterious mutations is explained by the fact that the structure  
285 of the genetic load differs between dominant and recessive S-alleles: the dominant S-alleles tend to  
286 have more fixed deleterious mutations, but the recessive S-alleles compensate by accumulating a  
287 larger number of segregating mutations, resulting in similar numbers of deleterious mutations  
288 overall in most of the populations.

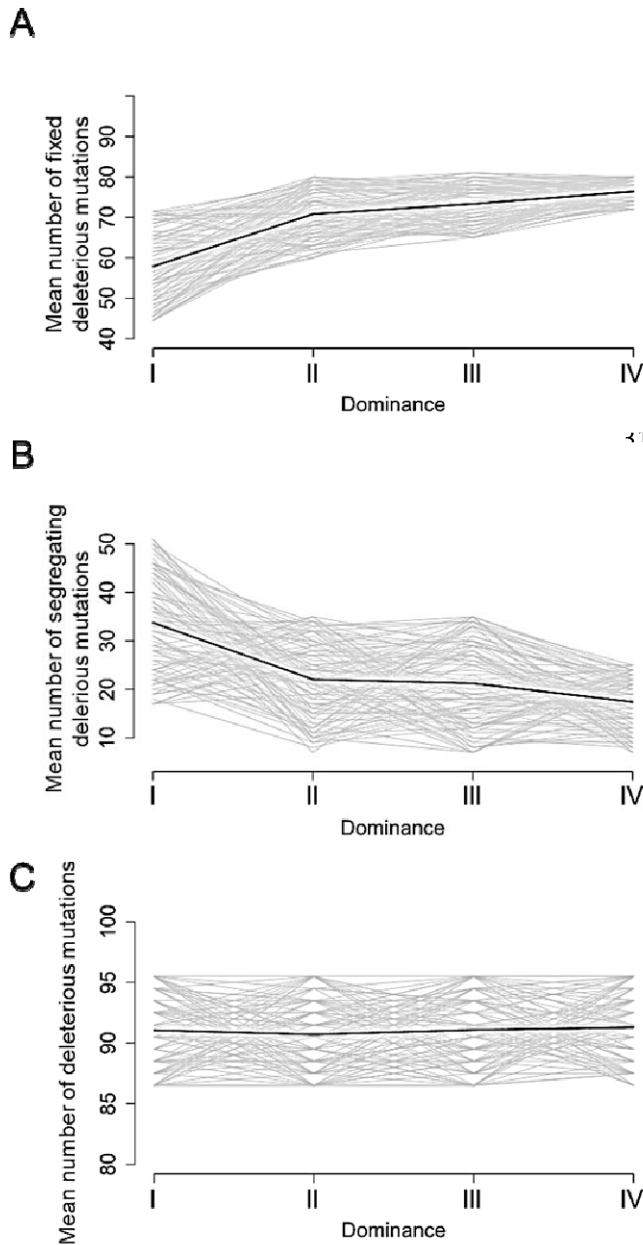


289

290 **Figure 4: The number of 0-fold degenerate mutations fixed in the 25kb regions flanking the S-locus**  
291 **increases with dominance of the S-allele associated.** Each dot represents the value obtained for  
292 haplotypes linked to one S-allele in one population. The correlations evaluated by a GLM are  
293 represented by lines, with confidence intervals represented in grey. A: *A. halleri*. Black dots  
294 correspond to the Nivelle population, grey dots to the Mortagne population. B: *A. lyrata*. Red dots  
295 correspond to the IND population, black dots to the PIN population, blue dots to the TSS population.

296 Motivated by these empirical observations, we extended the model proposed by Llaurens et al.<sup>13</sup> to  
297 predict the dynamics of accumulation of recessive deleterious mutations linked to S-alleles, focusing  
298 not only on fixed deleterious mutations but also on those that are segregating within allelic lineages.  
299 These stochastic simulations confirmed that, at equilibrium, dominant S-alleles tend to accumulate a  
300 larger number of deleterious mutations that are fixed among gene copies within S-allele lineages  
301 (Fig. 5A). In contrast, the number of segregating linked mutations was higher for recessive than for  
302 dominant S-alleles (Fig. 5B). Our model predicted that these two effects eventually compensate each  
303 other, such that in the end the total number of linked deleterious mutations was not expected to  
304 change with dominance of the S-alleles (Fig. 5C). These predictions are in line with our genomic  
305 observations and suggest that the dominance level of S-alleles modifies the structure of the genetic  
306 load they shelter: dominant S-alleles accumulate more fixed deleterious mutations, but recessive S-  
307 alleles accumulate more segregating mutations, resulting in an equivalent total load.

308



**Figure 5: Stochastic simulations confirm the contrasted architecture of the load of deleterious mutations linked to dominant vs. recessive S-alleles.** Number of fixed (A), segregating (B) and total (C) deleterious mutations linked to S-alleles at four different levels of dominance ( $I < II < III < IV$ ). The means (bold lines) were estimated per S-allele dominance classes over 100 replicate simulations after discarding an initial burn-in of 100,000 generations.  $h=0$ .  $s=0.01$ .

333

334

## 335 Discussion

336 *The genetic load linked to the S-locus is detectable and is manifested on different phenotypes*

337 Our results contribute to a still restricted but growing body of evidence confirming that the  
338 accumulation of deleterious mutations linked to strongly balanced allelic lines can be substantial,  
339 and that their effect can be detected at the phenotypic level<sup>13,19,25-28</sup>. An interesting observation is  
340 that the phenotypes on which the load was revealed varied among these studies. Here, the effect of  
341 homozygosity at the *S*-locus was apparent on juvenile survival and on the length of the longest  
342 flowering stem, but we detected no effect on any other morphological measurements, including leaf  
343 and rosette traits. In the same population of *A. halleri*, Llaurens et al.<sup>13</sup> detected an effect on juvenile  
344 survival and on leaf size. A study in North American outcrossing populations of *A. lyrata*<sup>19</sup> detected  
345 an effect on juvenile survival, but not on any other traits that they measured. In the horsenettle  
346 *Solanum carolinense*, the load was associated with reduced seed viability, flower number and  
347 germination<sup>25,28</sup>. Hence, the most consistent pattern seems to be a decrease of overall juvenile  
348 survival, possibly because it is a highly integrative measurement of fitness, whereas other  
349 morphological or life history traits can be associated with more specific components of overall  
350 fitness.

351 *A unique genetic load associated with each allele in each population.*

352 The model of the sheltered load posits that each *S*-allele should be associated with a specific set of  
353 linked mutations<sup>13</sup>. In line with this prediction, another consistent observation is that the magnitude  
354 of the *S*-linked load varied among *S*-alleles, as the load linked to some *S*-alleles was phenotypically  
355 detectable, while for others it was not. This variation of the genetic load is expected since  
356 deleterious mutations associated with the different alleles are likely to hit different linked genes,  
357 and affect different phenotypic traits with different effects on fitness. Also in line with the model of  
358 the sheltered load, our phasing of a large number of variants linked to *S*-haplotypes in several  
359 natural populations revealed that the same suite of linked mutations was consistently associated  
360 among different copies of a given allele when sampled from within the same population, in  
361 particular for the dominant *S*-allele lineages under more intense balancing selection. As expected for  
362 outcrossing populations with short-scale linkage disequilibrium, this association was lost when  
363 examining sites at increasing genetic distances from the *S*-locus along the chromosome (see also Le  
364 Veve et al.<sup>20</sup>). However, a proper model of sheltered load taking into account recombination among  
365 *S*-alleles is still missing. Finally, the association with linked sites was further lost when comparing  
366 gene copies of *S*-alleles sampled from different local populations, suggesting that recombination  
367 within populations decouples alleles from their linked sites faster than migration can homogenise  
368 the genetic composition among these natural populations. We note that the patterns of association  
369 and phylogenetic structure differed among populations, possibly due to their contrasted  
370 demographic histories. Indeed, the *A. lyrata* populations colonised North America from ancestral  
371 European populations about 20-30.000 years ago<sup>29,30</sup>, and are less diverse overall than the *A. halleri*  
372 populations we studied, who colonised the north of France during the last century from ancestral  
373 German populations<sup>31</sup>. The progressive decoupling between alleles and their linked sites leads to the  
374 simple prediction that *S*-locus homozygous genotypes formed by crossing individuals carrying  
375 identical alleles from distinct populations should not reveal as much load as when they are formed

376 by crossing individuals within populations. Hence, the *S*-locus region could contribute to overall  
377 hybrid vigour. Testing this prediction will be an interesting next step.

378 *Different properties of the linked load according to S-allele dominance*

379 The question of whether *S*-allele dominance, which modifies the intensity of balancing selection in  
380 the *S*-locus region, could explain variation of the linked load has received conflicting support in the  
381 literature. In line with Stift et al.<sup>19</sup>, but in contradiction with Llaurens et al.<sup>13</sup>, we observed no overall  
382 effect of dominance on the magnitude of the load. Several technical and biological reasons could  
383 explain the contrasted results obtained in these different studies. First, phenotypic quantification of  
384 the linked load is experimentally demanding, such that these studies relied on the comparison of a  
385 limited number of alleles (three *S*-alleles in each of the studies) and therefore each of them had  
386 inherently low power. Second, the experimental procedures to reveal the load varied slightly.  
387 Llaurens et al.<sup>13</sup> used CO<sub>2</sub> treatment to by-pass the SI system and obtain homozygous progenies from  
388 crosses that would otherwise have been incompatible, whereas we used the “natural” masking by  
389 dominant *S*-alleles to enable the obtention of recessive homozygous genotypes. Our approach is  
390 experimentally simpler and avoids the possible contamination by offspring obtained by selfing,  
391 which may combine the effect of the sheltered load with that of genome-wide inbreeding  
392 depression (see Stift et al.<sup>19</sup> for a detailed discussion of this caveat). Third, a limitation of our  
393 approach is that it is restricted to *S*-alleles that are recessive or intermediate along the dominance  
394 hierarchy, and is thus not applicable to quantify the load associated with the most dominant *S*-alleles  
395 under more intense balancing selection. It is therefore possible that the *S*-alleles we examined did  
396 not exhibit sufficiently contrasted levels of dominance, in particular if only the most dominant ones  
397 are generating a substantial load, as suggested for fully linked deleterious mutations. In addition,  
398 since the homozygous *S*-allele genotypes we create correspond to different gene copies from the  
399 population, they may carry distinct sets of linked deleterious variants, especially for the more  
400 recessive alleles. The variation we observed in the phenotypic magnitude of the load among families  
401 confirms that linked deleterious variants are unlikely to be fixed within all allele lineages. Finally, we  
402 note that our genomic analysis of the genetic load shows that the dominant allele Ah15 previously  
403 associated with reduced fitness in homozygotes<sup>13</sup>, is indeed unusual in terms of the number of  
404 mutations it carries. In fact, it is one of the most “loaded” alleles among all the dominant *S*-alleles  
405 present in this population, possibly explaining why Llaurens et al.<sup>13</sup> observed a significant effect  
406 despite the inherently limited experimental power of their analysis.

407 Our stochastic simulations and genomic analyses concur to the conclusion that the intensity of  
408 balancing selection applied to each allele affects the genetic architecture of the linked load: a larger  
409 proportion of putatively deleterious mutations are fixed among gene copies of the dominant as  
410 compared to the recessive *S*-alleles, while gene copies of the recessive *S*-alleles tend to accumulate  
411 more segregating deleterious variation. While these two processes eventually compensate one  
412 another, they may have distinct consequences for the evolution of *S*-alleles. Uyenoyama<sup>12</sup> showed  
413 that the existence of a sheltered load should influence the evolutionary dynamics of new *S*-alleles  
414 through self-compatible intermediates. Specifically, antagonistic interactions are expected between  
415 ancestral and derived functional specificities because they would initially share their linked  
416 deleterious mutations, slowing down the establishment of new alleles. Our observation that partially  
417 different sets of linked mutations are associated with *S*-alleles from the different populations raises

418 the question of whether the (short) time scale at which recombination decouples alleles from their  
419 sets of linked mutation is sufficiently fast to impede such antagonistic interactions to take place. In  
420 other words, the effect of the load on the diversification dynamics should be most important if the  
421 two mutational steps required for the emergence of new alleles under this model take place within  
422 local populations, rather than involving a metapopulation-scale process. As shown by Stetsenko et  
423 al.<sup>32</sup>, this is expected to occur under very low dispersal only. In addition, the observation that the  
424 architecture of the sheltered load differs between dominant and recessive *S*-alleles suggests that  
425 their diversification dynamics may also differ. Specifically, the self-compatible intermediates  
426 required for the formation of new *S*-alleles<sup>33,34</sup> are expected to be capable of selfing as well as  
427 forming homozygous genotypes that would otherwise be prevented. While the consequences of  
428 selfing may be equivalent for all alleles (because the overall number of mutations to which they are  
429 linked are equivalent), the consequences of the formation of homozygotes allowed by the crossing  
430 of separate individuals sharing a given *S*-allele are expected to be more severe for dominant *S*-  
431 alleles. The segregation of distinct deleterious variants linked to different gene copies of recessive *S*-  
432 alleles implies that linked recessive deleterious mutations are likely to remain masked when two  
433 distinct gene copies of a given recessive *S*-allele are brought together. Hence, our results lead to the  
434 prediction that in natural populations self-compatible mutants may segregate more readily for the  
435 more recessive than for the more dominant *S*-alleles, and more generally for allelic lineages under  
436 lower intensity of balancing selection. Considering that self-compatible mutants are a necessary  
437 intermediate stage in the formation of new *S*-alleles, one may predict that the diversification  
438 dynamics should be more efficient for lineages of recessive than dominant *S*-alleles. This prediction  
439 is in line with the deeper phylogenetic divergence among the most dominant *S*-alleles observed in  
440 *Arabidopsis*<sup>16</sup>. Detailed quantification of the presence of self-compatible variants in natural  
441 populations will now be necessary to test this hypothesis. At this stage, however, a proper model of  
442 allelic diversification taking into account dominance interactions among *S*-alleles is still missing.

443 Our observation that the genetic load varies across balanced allelic lines is not unprecedented. The  
444 classical case of Y or W sex chromosomes are indeed examples where one balanced line accumulates  
445 a greater genetic load than the other (X or Z, respectively), eventually leading to substantial genetic  
446 degeneration<sup>35-37</sup>. Another example is the supergene controlling variation in male plumage  
447 phenotypes of the ruff, where the genetic load on the derived “Satellite” haplotype is higher than on  
448 the ancestral “Independent” haplotype<sup>3,38</sup>. Similarly, in the butterfly *Heliconius numata*, the inverted  
449 haplotypes conferring mimetic wing patterns tend to accumulate a greater load than the non-  
450 inverted haplotypes<sup>39</sup>. Interestingly, in all these cases, the haplotypes with the greatest load also act  
451 genetically in a dominant manner, establishing a clear parallel with our observations. An interesting  
452 next step will be to determine whether similar asymmetries between the balanced allelic lines are  
453 observed in the linked region for these other systems.

454 It is clear from our results that *S*-allele dominance can affect the linked load, through its effect on  
455 the intensity of balancing selection, but in turn the differences in structure of the linked load may  
456 affect the conditions under which dominance can evolve. The Brassicaceae *S*-locus is indeed an  
457 interesting and rather unique system, where the mechanisms by which dominance is controlled and  
458 evolves under the action of so-called “dominance modifiers” have been studied in detail<sup>39,40</sup>  
459 (involving the interaction between small non-coding RNAs and their target sites<sup>16, 40</sup>). The presence  
460 of deleterious mutations linked to *S*-alleles has been shown to affect the evolution of dominance

461 modifiers, favouring evolution towards greater dominance than towards greater recessivity<sup>40</sup>. This  
462 asymmetry arises from the fact that *S*-alleles that become recessive (e.g. following acquisition of a  
463 recessivity modifier such as a small RNA target) will start forming homozygous genotypes, leading to  
464 expression of their linked load, while *S*-alleles that become dominant will not. Our observation that  
465 many deleterious mutations linked to recessive *S*-alleles are indeed segregating, suggests that  
466 expression of the load will be less severe for recessive than for dominant *S*-alleles, hence decreasing  
467 this predicted asymmetry. It will now be essential to modify models for the evolution of dominance  
468 to allow for such differential load among *S*-alleles.

## 469 **Acknowledgements**

470 This work was funded by the European Research Council (NOVEL project, grant #648321) and ANR  
471 TE-MoMa (grant ANR-18-CE02-0020-01). AL's PhD thesis was funded by the ERC and the University  
472 of Lille. The authors thank Barbara Mable for sharing seeds of *A. lyrata* and Camille Roux for  
473 discussions. This work was performed using infrastructure and technical support of the Plateforme  
474 Serre, cultures et terrains expérimentaux - Université de Lille for the greenhouse/field facilities. The  
475 authors thank the UMR 8199 LIGAN-MP Genomics platform (Lille, France) which belongs to the  
476 'Federation de Recherche' 3508 Labex EGID (European Genomics Institute for Diabetes; ANR-10-  
477 LABX-46) and was supported by the ANR Equipex 2010 session (ANR-10-EQPX-07-01; 'LIGAN-MP').  
478 The LIGAN-MP Genomics platform (Lille, France) is also supported by the FEDER and the Region des  
479 Hauts-de-France. The authors thank the GenoScreen platform (Lille, France).

480

## 481 **Author contributions**

482 AL, ED, VC and XV developed and designed the experiments for the study. AL, CBL and CP performed  
483 the experiments. AL and MG analysed and interpreted the data. AL, ED, VC and XV wrote the  
484 manuscript. All authors edited the manuscript.

485

## 486 **Declaration of interests**

487 The authors declare no competing interests

488



489 **Figures legend**

490 **Figure 1: Effect of homozygosity at the S-locus on 13 phenotypic traits compared to heterozygotes.**

491 For each trait, the phenotypic values in homozygotes (in grey) were normalised relative to the mean  
492 phenotypic values in heterozygotes (in black). The distributions were obtained by 10,000 random  
493 permutations. SD: Standard Deviation.

494 **Figure 2: Linkage to the S-locus locally distorts the phylogenetic relationships.** A. The two

495 topologies of interest cluster haplotypes either by the S-allele to which they are linked (top) or by the  
496 populations where they came (bottom). Different S-alleles are represented by symbols of different  
497 colours, different populations of origin are represented by symbols of different shapes. B. Difference  
498 in log likelihood of the two topologies of interest. Dots correspond to the difference in log likelihood  
499 for overlapping series of 50 SNPs around the S-locus for *A. halleri* (top panel) and *A. lyrata* (bottom  
500 panel). Positive values correspond to chromosomal positions where the topology by S-alleles explains  
501 the phylogeny of haplotypes better than the topology by populations. The right panels show the  
502 difference in log likelihood in the control regions. 2.5 and 97.5 percentiles of the distribution in the  
503 control regions are indicated by dashed lines.

504 **Figure 3: No overall effect of S-allele dominance on the total number of 0-fold degenerate**

505 **mutations ( $S_{0f}$ ) in the linked genomic regions within 25kb.** Each dot represents the mean number of  
506 mutations observed among haplotypes linked to one S-allele in one population. The correlations  
507 evaluated by a GLM are represented by lines, with confidence intervals represented in grey. A: *A.*  
508 *halleri*. Black dots correspond to the Nivelles population, grey dots to the Mortagne population. The  
509 red arrow points to the copy of Ah15, corresponding to the S-allele whose sheltered load was  
510 phenotypically characterised by Llaurens et al.<sup>13</sup> in Nivelles. B: *A. lyrata*. Red dots correspond to the  
511 IND population, black dots to the PIN population and blue dots to the TSS population.

512 **Figure 4: The number of 0-fold degenerate mutations fixed in the 25kb regions flanking the S-locus**

513 **increases with dominance of the S-allele associated.** Each dot represents the value obtained for  
514 haplotypes linked to one S-allele in one population. The correlations evaluated by a GLM are  
515 represented by lines, with confidence intervals represented in grey. A: *A. halleri*. Black dots  
516 correspond to the Nivelles population, grey dots to the Mortagne population. B: *A. lyrata*. Red dots  
517 correspond to the IND population, black dots to the PIN population, blue dots to the TSS population.

518 **Figure 5: Stochastic simulations confirm the contrasted architecture of the load of deleterious**

519 **mutations linked to dominant vs. recessive S-alleles.** Number of fixed (A), segregating (B) and total  
520 (C) deleterious mutations linked to S-alleles at four different levels of dominance (I<II<III<IV). The  
521 means (bold lines) were estimated per S-allele dominance classes over 100 replicate simulations after  
522 discarding an initial burn-in of 100,000 generations.  $h=0$ .  $s=0.01$ .

523

524

525 **Tables List**

526 **Table 1: Proportion of S-locus homozygous offspring having reached the reproductive stage for**  
527 **three different S-alleles. The test is performed relative to the expected proportion of homozygous**  
528 **genotypes in the offspring (25% when both parents are heterozygous; 50% when one of the parents**  
529 **is homozygous and the other heterozygous).**

S-allele	Level of dominance	Number of seedlings having reached the reproductive stage	Observed proportion of homozygotes	Ratio of the observed to expected proportion of homozygous genotypes ( $p$ -values <sup>a</sup> )	Observed proportion of heterozygotes with the S-allele ( $p$ -values compared to expected)
Ah01	I	35	0.29	1.14 (0.70; 0.76)	19 (0.40; 0.90)
Ah03	II	27	<b>0.074</b>	<b>0.3 (0.04; 0.02)</b>	17 (1; 0.69)
Ah04	III	96	0.479	0.96 (0.75; 0.39)	50

530 <sup>a</sup>  $p$ -values were obtained by binomial test and after 10,000 random permutations, respectively.  
531 Values departing from Mendelian expectations are figured in bold. For Ah04, the maternal parent  
532 was Ah04/Ah04, so all heterozygous offspring carried the S-allele.

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

## 550 **Methods STAR**

### 551 *Source plant material*

552 We worked on natural accessions from two closely related species, *A. halleri* and *A. lyrata*,  
553 represented by two population samples named Mortagne (50°47'N, 3°47'E, France,  $n=60$ ) and  
554 Nivelles (50°47'N, 3°47'E, France,  $n=61$ ) for *A. halleri*, and three highly outcrossing population  
555 samples from the North American Great Lakes, named IND (Indiana Dunes National Lakeshore in  
556 Michigan,  $n=9$ ), PIN (Pinery Provincial Park in Ontario,  $n=11$ ) and TSS (Tobermory Provincial Park in  
557 Ontario,  $n=8$ )<sup>22</sup> for *A. lyrata* (Fig. S9). The *A. lyrata* populations colonised North America from  
558 ancestral European populations about 20-30.000 years ago<sup>29,30</sup> and the *A. halleri* populations are  
559 peripheral and likely colonised the north of France during the last century from ancestral German  
560 populations<sup>31</sup>.

561 We performed 92, 91, 40, 43 and 21 controlled crosses between randomly chosen individuals within  
562 the Nivelles, Mortagne, IND, PIN and TSS populations, respectively. We successfully obtained seeds  
563 from 60, 66, 21, 21 and 10 of these crosses, respectively. Because we were not interested in  
564 estimating population frequencies of *S*-alleles, we instead tried to maximise the number of  
565 reconstructed haplotypes and avoid over representing the most recessive *S*-allele (Ah01) that tends  
566 to segregate at very high frequencies in natural populations<sup>15</sup>. To do this, we performed PCR with *S*-  
567 allele-specific primers<sup>15,18</sup> to screen the parents of the crosses and we removed from the experiment  
568 offspring with two parents carrying allele Ah01. For *A. halleri*, we selected 19 individuals from the  
569 Nivelles population and 19 individuals from the Mortagne population, based on their genotype at the  
570 *S*-locus (Fig. S9; Table S7). We also selected one offspring of 9, 11, 5, 6 and 5 pairs of selected  
571 individuals in the Nivelles, Mortagne, IND, PIN and TSS populations respectively for the phasing of *S*-  
572 haplotypes (Table S8; Fig. S9). To increase sample size for the phenotypic measurements, we  
573 included offspring from five additional crosses from the Nivelles population (Table S8).

### 574 *Library preparation, capture and sequencing*

575 We used a previously developed sequence capture approach to specifically sequence genomic  
576 regions of interest<sup>20</sup>. Briefly, indexed genomic libraries were constructed for each individual and  
577 libraries were pooled in equimolar proportions. Fragments matching a series of regions of interest  
578 (including in particular the 75kb upstream and downstream of the non-recombining *S*-locus region as  
579 well as a series of 100 unlinked 25kb regions used as genomic controls<sup>20</sup>), were then enriched using  
580 synthetic 120bp RNA probes and sequenced by Illumina MiSeq (a total of 159 million paired-end  
581 reads).

582 For six individuals (Table S7, S8), we completed the sequencing with genome-wide resequencing  
583 (WGS) in order to distinguish the homozygous and heterozygous genotypes at the *S*-locus based on  
584 read depth<sup>41</sup>, which is not possible using data from the capture protocol. The prepared libraries were  
585 sequenced by Illumina NovaSeq (2x 150pb, paired-end) from the GenoScreen platform (Lille,  
586 France).

### 587 *Determination of the S-locus genotypes and dominance of S-alleles*

588 We used a dedicated pipeline for genotyping the *S*-locus based on short reads sequencing<sup>41</sup> obtained  
589 from each individual (Table S7 and S8). The level of dominance of *S*-alleles found in our study was  
590 determined based on either previous assessment of dominance in *A. lyrata* and *A. halleri*<sup>15,19,42-44</sup> or  
591 indirectly inferred based on the observed association between the phylogeny of *S*-alleles and levels  
592 of dominance<sup>45</sup>.

### 593 *Read mapping and variant calling in A. halleri and A. lyrata populations*

594 Raw reads were mapped on the complete *A. lyrata* reference genome V1.0.23<sup>21</sup> using Bowtie2  
595 v2.4.1<sup>46</sup>, as described in Le Veve et al<sup>20</sup>. File formats were then converted to BAM using samtools  
596 v1.3.1<sup>47</sup> and duplicated reads were removed with the MarkDuplicates program of picard-tools  
597 v1.119 (<http://broadinstitute.github.io/picard>). These steps were performed by the custom Python  
598 script `sequencing_genome_vcf.py` available at [https://github.com/leveveaudrey/analysis-of-](https://github.com/leveveaudrey/analysis-of-polymorphism-S-locus)  
599 `polymorphism-S-locus`.

600 We obtained an average of 620 million properly mapped paired-end 300bp reads per population  
601 sample. For consistency, we conserved only reads which mapped to the *S*-locus flanking or control  
602 regions, even for samples sequenced by WGS, using the *targetintercept* option of bedtools v2.25.0<sup>48</sup>.  
603 We called all SNPs within the chromosomal segment comprising 50 kb upstream from the first base  
604 of the gene *Ubox* in 3' and 50 kb downstream from the last base of the gene *ARK3* in 5' of the *S*-locus  
605 using the Genome Analysis Toolkit v. 3.8 (GATK)<sup>49</sup> with the option GVCF and a quality score threshold  
606 of 60 using vcftool v0.1.15<sup>50</sup>. This region contains 20 annotated protein-coding genes. In this study  
607 we excluded the genes inside the *S*-locus itself (*SCR*, *SRK*). For each sample independently, we  
608 computed the distribution of coverage depth across control regions using samtools `depth`<sup>47</sup>. We  
609 excluded sites with either less than 15 reads aligned or coverage depth above the 97.5 % percentile,  
610 as the latter are likely to correspond to repeated sequences (e.g. transposable elements or  
611 paralogs). Finally, we removed SNPs fixed in each population using the script `1_fix_pos_vcf.py`  
612 ([https://github.com/leveveaudrey/dominance\\_and\\_sheltered\\_load](https://github.com/leveveaudrey/dominance_and_sheltered_load)), thus retaining only nucleotide  
613 sites that were variable in the population.

#### 614 *Quantifying the sheltered load of deleterious mutations*

615 We examined the genetic load signatures based on the accumulation of mutations on 0-fold  
616 degenerate sites, the vast majority of which are considered deleterious. The 0-fold and 4-fold  
617 degenerate sites were identified and extracted from the reference genome and the gene annotation  
618 using the script `NewAnnotateRef.py`<sup>51</sup>. We also examined the genetic load signatures based on the  
619 accumulation of all non-synonymous mutations. We obtained a total of 2,441 and 2,435 variable  
620 positions for the *A. halleri* samples from Nivelles and Mortagne respectively, and 2,360 variable  
621 positions for the *A. lyrata* samples.

#### 622 *Phasing S-haplotypes*

623 For each of the 9, 11, 5, 6 and 5 trios analysed in the Nivelles, Mortagne, IND, PIN and TSS  
624 populations respectively, we phased mutations in the flanking regions, resulting in 130 phased  
625 haplotypes (Fig. S9). Briefly, we used sites that were heterozygous in the offspring to resolve  
626 parental haplotypes by assuming no recombination between parent and offspring, thus attributing  
627 the allelic state that was shared between a parent and its offspring to their shared *S*-allele, and the  
628 allelic state that was not shared to the other (untransmitted) haplotype of the parent. Twelve of the  
629 parents had been used in more than one cross, and in these cases we phased their haplotypes only  
630 once (Table S8). We implemented the phasing procedure in the script `3_phase_S_allele.py` available  
631 at [https://github.com/leveveaudrey/dominance\\_and\\_sheltered\\_load](https://github.com/leveveaudrey/dominance_and_sheltered_load).

#### 632 *Study of the structure of S-haplotypes*

633 We first developed a new method to evaluate the distortion of the phylogenetic patterns caused by  
634 linkage to *S*-alleles. To do this, we used `phyml`<sup>52</sup> v.3.3 to calculate the likelihood of two contrasted  
635 topologies of interest: (1) the topology clustering haplotypes by the populations where they came  
636 from vs. (2) the topology clustering them by the *S*-allele to which they are linked (Fig 2A). We used  
637 sliding windows of sequences with 50 SNPs to obtain the variation of the difference in log-likelihood  
638 between these two topologies along the chromosome. We then compared these values to their

639 distribution throughout the genome obtained by random draws of sequences with 50 SNPs from the  
640 control regions. Second, we visualised the relationships among the phased haplotypes using  
641 maximum likelihood phylogenies based on the Tamura-Nei model<sup>53</sup>, with 1,000 replicates in MEGA  
642 X<sup>54</sup>. Third, we followed Kamau et al.'s<sup>24</sup> approach and examined the variation of  $F_{ST}$  among  
643 populations within each species (Nivelle and Mortagne for *A. halleri* and IND, PIN and TSS for *A.*  
644 *lyrata*) along the flanking region in non-overlapping windows of 5kb. We also examined the variation  
645 of  $F_{ST}$  along the flanking region obtained by grouping haplotypes by their linked *S*-allele rather than  
646 by population of origin. Then, we compared these  $F_{ST}$  values computed in the *S*-locus flanking regions  
647 with their genomic distribution as determined from the 100 control regions. The  $F_{ST}$  values were  
648 estimated with the DNAsp 6 software<sup>55</sup>. Fourth, we performed a major component analysis (MCA)  
649 based on SNPs in the first 5kb, SNPs between 5 and 25kb and SNPs between 25 and 50kb around the  
650 *S*-locus, using the R packages 'ggplot2' (version 3.4.0), 'factoextra' (version 1.0.7) and 'FactoMiner'  
651 (version 2.7). We compared the patterns obtained by these MCAs with those obtained from identical  
652 numbers of SNP (+/- 1%) from the control regions. Finally, we analysed genetic association in each  
653 population independently between each of the locally segregating variants and the *S*-alleles  
654 considered as phenotypes, using STRAT<sup>56</sup> V1.1 combined with Structure<sup>57</sup> V2.3. We examined the  
655 distribution of the top 0.1% most significant associations detected specifically for each *S*-allele in  
656 each population.

#### 657 *Estimation of the number of fixed and segregating deleterious mutations within S-allele lineages*

658 For each variable position considered in the phased haplotypes, we estimated the number of  
659 mutations on 0-fold ( $S_{0f}$ ) and 4-fold degenerate sites ( $S_{4f}$ ) compared with the reference genome. We  
660 distinguished SNPs that were fixed from those that were segregating within each of the allelic lines.  
661 We used GLM with a Poisson distribution to test whether the number of fixed and segregating  
662 mutations were associated with *S*-allele dominance, considering populations as random effects. We  
663 reiterated the GLM analysis with the number of non-synonymous ( $S_{NS}$ ) and synonymous ( $S_S$ )  
664 mutations.

#### 665 *Estimation of the phenotypic impact of homozygosity at the S-locus for three S-alleles*

666 To determine if the genetic sheltered load putatively linked to the *S*-locus has a detectable  
667 phenotypic impact, we performed 45 crosses (Table S1) between offspring of the Nivelle individuals  
668 that we chose so that they shared one *S*-allele (Fig. S9). Based on the dominance hierarchy in  
669 pollen<sup>16</sup> (Table S1), these crosses should correspond to compatible partners. The general principle of  
670 the experiment was to take advantage of the dominance hierarchy to mask recessive *S*-alleles and  
671 generate full sibs that were either homozygous (because they inherited the *S*-allele that was shared  
672 by their two parents) or heterozygous at the *S*-locus, and thus isolate the effect of homozygosity at  
673 the *S*-locus. Note that all offspring in our experiments were thus "naturally" outcrossed, whereas  
674 Llaurens et al.<sup>13</sup> based their comparisons on outcrossed progenies obtained by enforced  
675 incompatible crosses and Stift et al.<sup>19</sup> based their comparisons on enforced selfed progenies. These  
676 crosses generated 399 seeds overall, with homozygous genotypes expected for the *S*-alleles Ah01,  
677 Ah03 and Ah04 forming the following dominance relationship: Ah01<Ah03<Ah04.

678 Seedlings were grown in a greenhouse between 14.5 and 23.1°C and a photoperiod of 16 hr day/8 hr  
679 night. Offspring from the six families were placed on tables, and their position randomised every  
680 three days. After three months of growing, all the germinated plants were vernalised under a  
681 temperature between 6 and 8°C and a natural photoperiod for two months (January-February).  
682 Then, all surviving plants began reproduction in a greenhouse under temperature between 10.6 and  
683 25.3°C and a natural photoperiod. The genotypes at the *S*-locus were determined in surviving plants  
684 by a PCR approach, using *S*-allele-specific primers for the pistil-expressed *SRK* gene. We assessed the  
685 reproductive success of offspring from the different crosses on the basis of fourteen phenotypic  
686 traits (detailed below) and computed the mean difference for the trait between homozygotes and

687 heterozygotes within each family. We also tested for departures from mendelian proportions of  
688 each *S*-locus genotypic category in the family after the apparition of the first stem. Significant  
689 departures were interpreted as reflecting differences in survival between homozygous and  
690 heterozygous *S*-locus genotypes. We performed 10,000 replicate simulations of mendelian  
691 segregation based on the *S*-locus genotype of the parents. We used GLM to test whether the  
692 phenotypic impact of homozygosity at the *S*-locus increased with dominance of the *S*-alleles. The  
693 models used for GLM depended on the type of trait analysed (poisson for the counts like the number  
694 of leaves, flowers by stems or days; gaussian for continue traits like the lengths, widths and areas).

695 We measured the following fourteen phenotypic traits: the time (days) to the first leaf measured by  
696 visual control every day during seven weeks after sowing the seeds, the number of leaves, the area  
697 of the rosette (cm<sup>2</sup>), the mean length and width of leaves (cm), the standard deviation of length and  
698 width of leaves (cm) and the mean area of leaves (cm<sup>2</sup>) measured by ImageJ<sup>58</sup> based on photographs  
699 taken seven weeks (+/- five days) after the first leaf. At reproduction, we measured the time to the  
700 first flower bud for the end of vernalisation (day), scored by visual control every three days during  
701 nine weeks, the number of flower buds per flower stem produced during four week after the  
702 appearance of the first bud, the number of flower stems, the length of the highest flower stem  
703 produced four weeks after the appearance of the first bud (cm), and finally the total duration of  
704 buds production (days), scored by visual control every three days during eleven weeks after the  
705 appearance of the first bud. The last trait we measured was the proportion of homozygotes per  
706 family that survived until reproduction assuming mendelian proportions in the seeds. During the  
707 whole experiment, the presence of phytophagous insects, pathogens and stress markers were  
708 scored as binary variables. The presence of phytophagous insects and pathogen attacks were  
709 detected by the occurrence of gaps in leaves. Oxidative stress was scored qualitatively based on the  
710 occurrence of purple leaves. We also controlled the effect of the family on the phenotypic trait.  
711 These effects were controlled by redistributing 10,000 times the values observed in groups of the  
712 same size observed for each effect (for example, presence or absence of pathogen attack) and  
713 comparing the difference for the trait observed with the distribution of the differences obtained in  
714 the permutations. We considered the impact of the effect on the trait if the observed difference  
715 between groups was higher than the 95% percentile of the distribution obtained randomly (Table  
716 S9). When the test was significant, the effect was implemented as a random effect in the GLM. We  
717 used the same method to control for the family effect, which was included as a random effect in  
718 GLM if necessary (Table S10). The general experimental procedure is summarised in Fig. S9 and all  
719 data analyses were done in R ver. 3.1.2 (R Development Core Team 2014).

## 720 *Simulations*

721 Finally, we refined the model of Llaurens et al.,<sup>13</sup> in several ways. We simulated a panmictic  
722 population of *N* diploid individuals with non-overlapping generations. Each individual was defined by  
723 its genotype in a non-recombining genomic region. This region contains the *S*- locus, and a *D* locus  
724 where deleterious mutations accumulated. For the *S*-locus, we used a simple model of sporophytic  
725 SI, with four dominance classes, as observed in *A. halleri*<sup>41</sup> (only three classes were considered  
726 before<sup>13</sup>), and fourteen *S*-alleles (eight alleles in the class IV, three in the class III, two in the class II  
727 and one allele in the class I). This distribution mirrors that of the Nivelles population (Table S7), with  
728 the exception that a class II allele has been added because its presence has been reported in  
729 previous studies<sup>15</sup>. Alleles within classes were assumed to be codominant with each other, and  
730 dominant over all alleles of the more recessive classes, with the following linear hierarchy between  
731 classes : classI<classII<classIII<classIV). We also assumed that no new *S*-allele could appear by  
732 mutation during the simulations. The population size was 10,000 diploid individuals, so as to be large  
733 enough to avoid *S*-allele loss by drift during the simulations (previously it was 1,000). The “*D* locus”  
734 comprised one hundred fully linked biallelic positions (versus a single one in Llaurens et al.<sup>13</sup>). Fully  
735 recessive deleterious mutations were recurrently introduced (at a rate 10<sup>-4</sup>), and reverse mutations  
736 were possible (at a rate 10<sup>-5</sup>). We ignored partially recessive deleterious mutations because these

737 mutations were predicted to be effectively eliminated by natural selection in Llaurens et al.<sup>13</sup>. The  
738 survival probability  $p$  of a zygote depended on its genotype at the D locus:  $p = (1 - s)^n$  with  $s$  the  
739 selection coefficient and  $n$  the number of positions homozygous for the mutated allele. We explored  
740 different values of the selection coefficient (0.1, 0.05, 0.03, 0.01 and 0.005). Under strong selection  
741 ( $s=0.1$ , 0.05 and 0.03), the combined effect of multiple mutations led to low-fitness individuals,  
742 eventually causing population extinction. Under weak selection, ( $s=0.005$ ), we observed near fixation  
743 of the deleterious mutations under the influence of asymmetrical mutation. Hence, we focused on  
744 the intermediate value of the selection coefficient ( $s=0.01$ ), where deleterious mutations segregated  
745 stably in the simulations.

746 We first ran simulations without deleterious mutations until a deterministic equilibrium for  $S$ -allele  
747 frequencies was reached, which was considered to be attained when allelic frequencies changed by  
748 less than  $10^{-3}$  between generations. Recessive deleterious mutations were then allowed to  
749 accumulate at the positions within the D locus. Each simulation was performed with 100  
750 independent replicates of 100,000 generations, and the frequency of the deleterious alleles was  
751 recorded every 1,000 generations. At the end of the simulation runs, we estimated the number of  
752 deleterious mutations found in each haplotype associated with each  $S$ -allele to determine the  
753 expected patterns of association between the sheltered load and dominance at the  $S$ -locus.

## 754 **Data Availability**

755 All sequence data are available in the NCBI Short Read Archive (SRA;  
756 <https://www.ncbi.nlm.nih.gov/sra>) with accession codes: PRJNA744343, PRJNA755829.

757 All scripts developed are available in Github  
758 ([https://github.com/leveveaudrey/dominance\\_and\\_sheltered\\_load](https://github.com/leveveaudrey/dominance_and_sheltered_load)  
759 <https://github.com/leveveaudrey/analysis-of-polymorphism-S-locus> ).

760

## 761 **Supplementary data**

### 762 **Figures legend**

763 **Figure S1: Analysis of Major Components obtained for haplotypes of *A. halleri* of the Nivelles (black**  
764 **dots) and Mortagne (grey dots) populations) based on SNPs in the first 5kb, between 5 and 25kb,**  
765 **and between 25kb and 50kb away from the S-locus. The S-allele to which the SNPs are linked are**  
766 **represented by different symbols. The right panels show the analysis on control regions, each time**  
767 **matching the number of SNPs with that of the corresponding linked regions (left panels).**

768 **Figure S2: Analysis of major components (AMC) obtained for haplotypes of *A. lyrata* (of the PIN:**  
769 **grey dots, IND: red dots and TSS: blue dots) populations based on the SNPs in the first 5kb,**  
770 **between 5 and 25kb, and between 25kb and 50kb away from the S-locus. The S-allele to which the**  
771 **SNPs are linked are represented by different symbols. The right panels show the analysis on control**  
772 **regions, each time matching the number of SNPs with that of the corresponding linked regions (left**  
773 **panels).**

774 **Figure S3: Phylogenetic tree obtained by Maximum Likelihood for haplotypes of *A. halleri***  
775 **(populations Nivelles and Mortagne) across the first 25kb flanking the S-locus. The Tamura-Nei**  
776 **model was used and the percentage of trees in which the associated haplotypes clustered together is**  
777 **shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number**  
778 **of substitutions per site. The black braces indicate haplotypes clustering by populations. The overall**  
779 **pattern shows a structure by S-alleles, with exceptions highlighted in yellow.**

780 **Figure S4: Phylogenetic tree obtained by maximum likelihood for haplotypes of *A. halleri***  
781 **(populations Nivelles and Mortagne) based on the nucleotide positions between 25kb and 50kb**  
782 **away from the S-locus. The Tamura-Nei model was used and the percentage of trees in which the**  
783 **associated haplotypes clustered together is shown next to the branches. The tree is drawn to scale,**  
784 **with branch lengths measured in the number of substitutions per site. The black braces indicate**  
785 **haplotypes clustering by populations. The overall pattern shows a structure by S-alleles, with**  
786 **exceptions highlighted in yellow.**

787 **Figure S5: Phylogenetic tree obtained by maximum likelihood for haplotypes of *A. lyrata***  
788 **(populations PIN, IND, TSS) across the first 5kb flanking the S-locus. The Tamura-Nei model was**  
789 **used and the percentage of trees in which the associated haplotypes clustered together is shown next**  
790 **to the branches. The tree is drawn to scale, with branch lengths measured in the number of**  
791 **substitutions per site. The black braces indicate haplotypes clustering by populations. The overall**  
792 **pattern shows a structure by S-alleles, with exceptions highlighted in yellow.**

793 **Figure S6: Phylogenetic tree obtained by maximum likelihood for haplotypes of *A. lyrata***  
794 **(populations PIN, IND, TSS) based on the nucleotide positions between 5kb and 10kb away from**  
795 **the S-locus. The Tamura-Nei model was used and the percentage of trees in which the associated**  
796 **haplotypes clustered together is shown next to the branches. The tree is drawn to scale, with branch**  
797 **lengths measured in the number of substitutions per site. The black braces indicate haplotypes**  
798 **clustering by populations. The overall pattern shows a structure by S-alleles, with exceptions**  
799 **highlighted in yellow.**



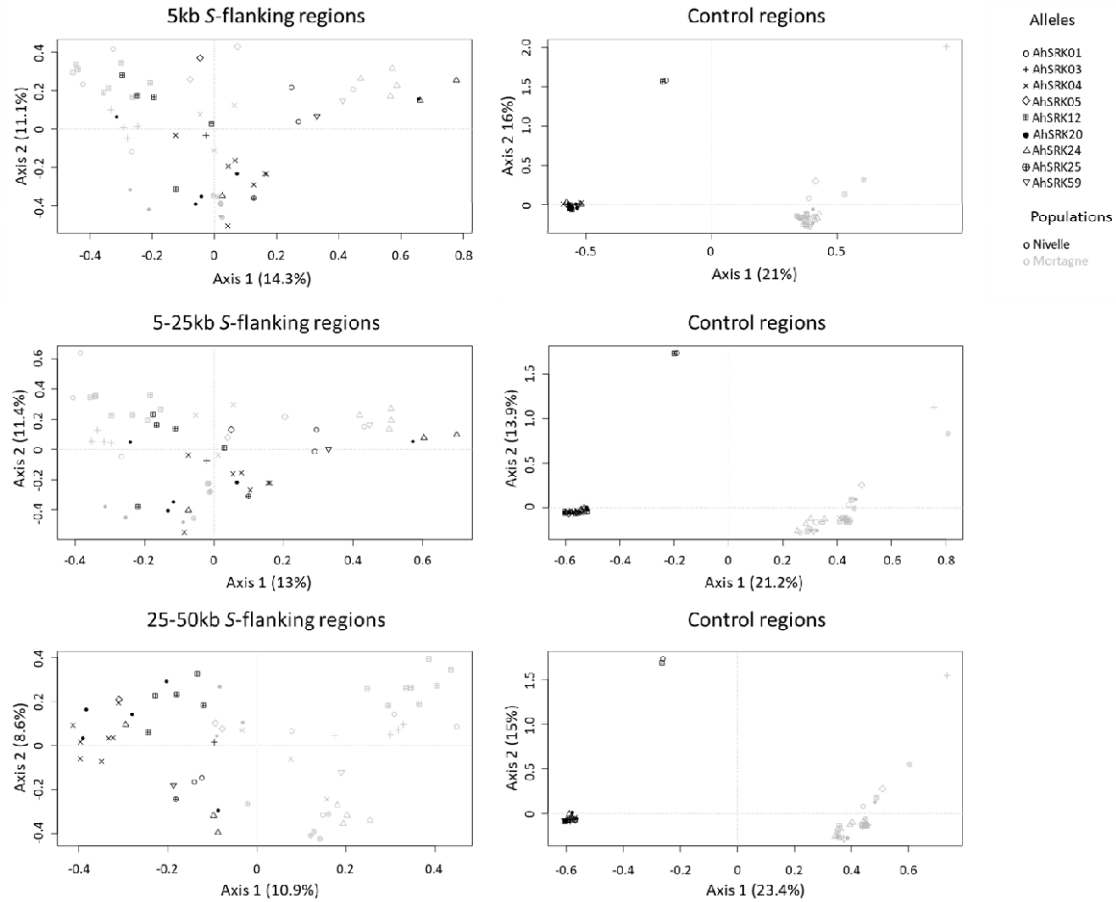
800 **Figure S7: The genetic structure of SNPs in the S-locus flanking regions in *A. halleri* and *A. lyrata*.**  
801 *Left: Mean  $F_{ST}$  (lines) and  $F_{ST}$  by pair (point) analysed among S-alleles (grey) or among populations*  
802 *(black) in 5 kb windows in the S-locus flanking regions. Right: Distribution (count) of  $F_{ST}$  in the control*  
803 *regions analysed among S-alleles (grey bars) or among populations (black bars). The 95% percentiles*  
804 *of the distributions are represented by dotted lines and the medians by solid lines.*

805 **Figure S8: Patterns of genetic associations between S-alleles and SNPs across the genome.** *Each*  
806 *dot corresponds to a SNP showing statistically significant association (top 0.1%) with a given S-allele.*  
807 *The left panel confirms that SNPs physically linked to the S-locus (in red) are considerably more likely*  
808 *to show statistical associations with S-alleles. The chromosomes were signified by alternance of black*  
809 *and grey. The middle panel shows a zoom on the 50kb regions flanking the S-locus and shows that*  
810 *the statistical association extends over long distances, at least for some S-alleles. Each region of 25*  
811 *kb was delimited by vertical lines. The right panel shows that the observed number of associated*  
812 *SNPs in the linked regions far exceeds that in regions of identical size from control regions. The*  
813 *histogram shows the distribution across control regions of the mean number of significantly*  
814 *associated SNPs per S-allele. The vertical lines correspond to the mean number of significantly*  
815 *associated SNPs in the first 25kb (solid lines) and the 25-50kb interval (dashed lines) away from S-*  
816 *locus.*

817 **Figure S9: Experimental protocol.** *A) We randomly crossed *A. lyrata* individuals from the PIN, TSS*  
818 *and IND populations in North America (left) and *A. halleri* from the Nivelle (middle) and Mortagne*  
819 *(right) populations. Individuals were sequenced by a capture protocol. Numbers between*  
820 *parentheses represent the number of individuals per dataset. B) One offspring from each cross was*  
821 *sequenced along with its two parents for trio haplotyping. Offspring from the Nivelle population*  
822 *(black circle) were conserved for the study of the impact of homozygosity at the S-locus on fitness (G1*  
823 *population). C) Individuals sequenced in A and B were used to reconstruct haplotypes linked to each*  
824 *copy of S-allele, assuming no recombination between the S-locus and its flanking regions between*  
825 *parents and offspring. D) We used the dominance hierarchy between S-alleles expressed in pollen<sup>22,23</sup>*  
826 *to cross G1 individuals from the Nivelle population and obtained six G2 families constituted of*  
827 *heterozygous and homozygous individuals for the alleles Ah01, Ah03 and Ah04. E) Description of the*  
828 *traits measured and the methods used to estimate the impact of homozygosity at the S-locus in*  
829 *homozygotes. Traits 1-8 are related to biomass and traits 10-14 are related to reproductive success.*

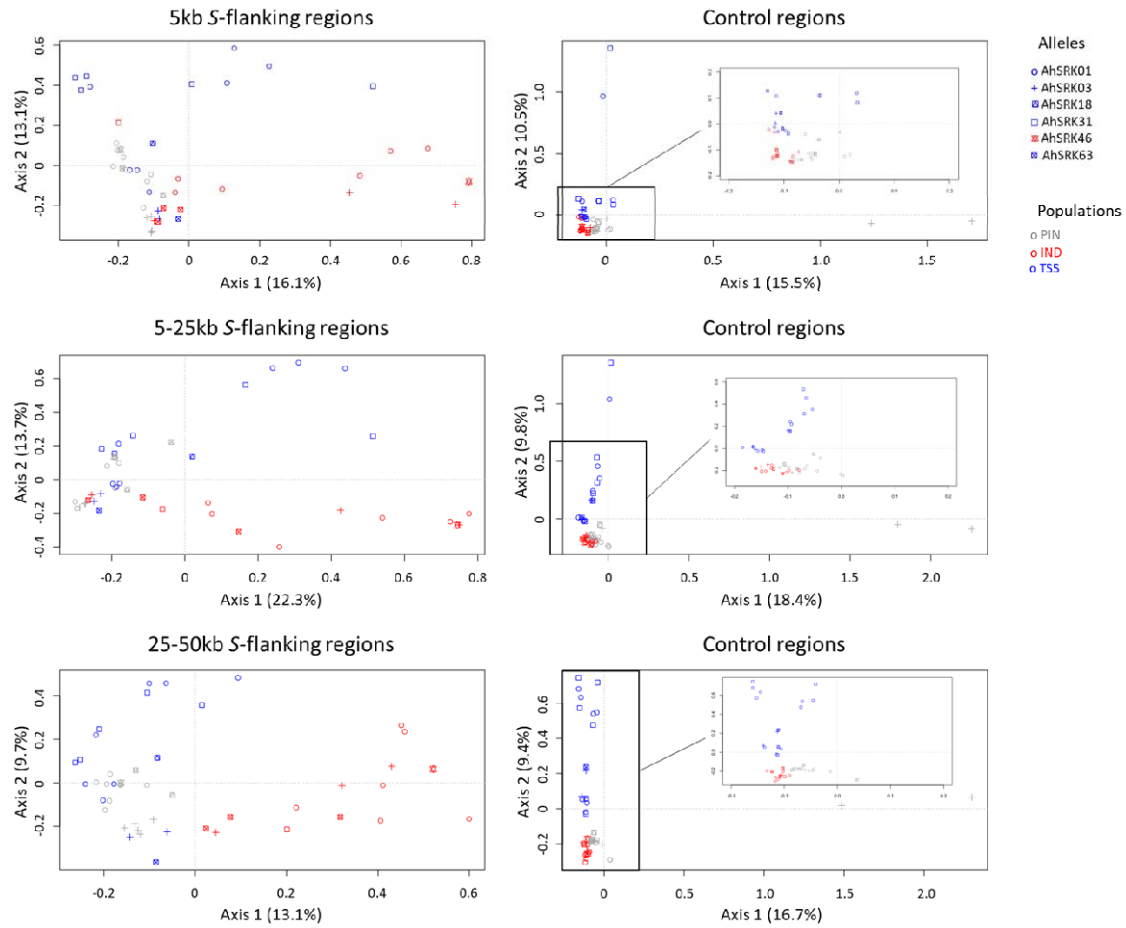
830

831



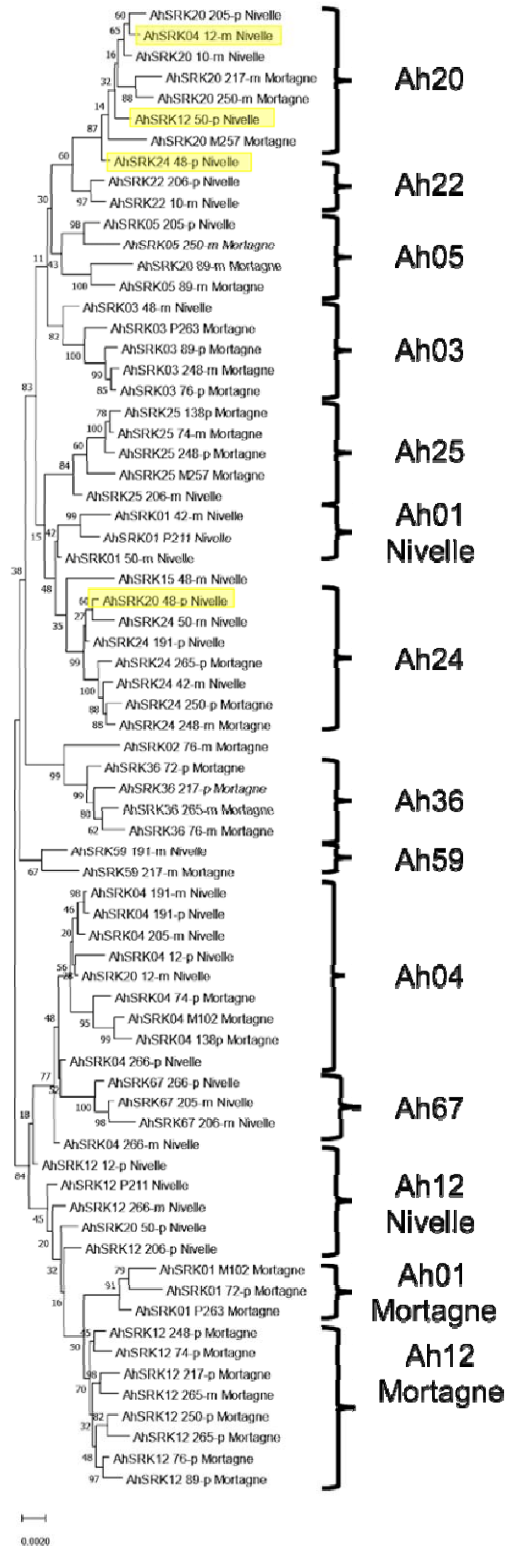
832

833 **Figure S1: Analysis of Major Components obtained for haplotypes of *A. halleri* of the Nivelle ( black**  
834 **dots) and Mortagne (grey dots) populations) based on SNPs in the first 5kb, between 5 and 25kb,**  
835 **and between 25kb and 50kb away from the S-locus.**



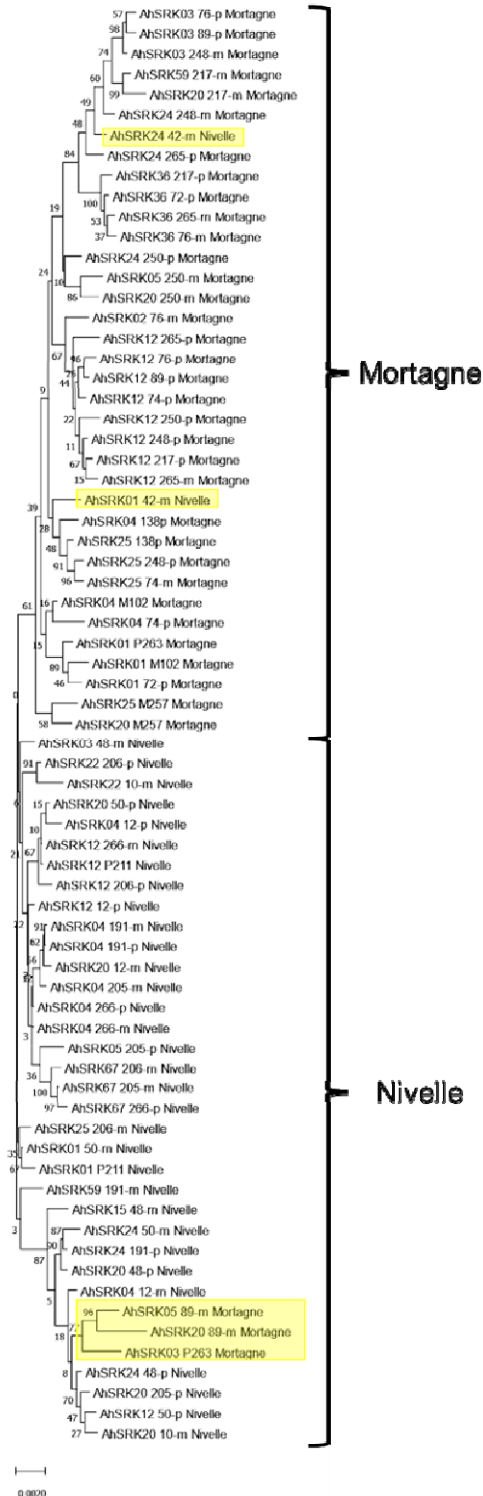
836

837 **Figure S2: Analysis of major components (AMC) obtained for haplotypes of *A. lyrata* (of the PIN:**  
838 **grey dots, IND: red dots and TSS: blue dots) populations based on the SNPs in the first 5kb,**  
839 **between 5 and 25kb, and between 25kb and 50kb away from the S-locus.**



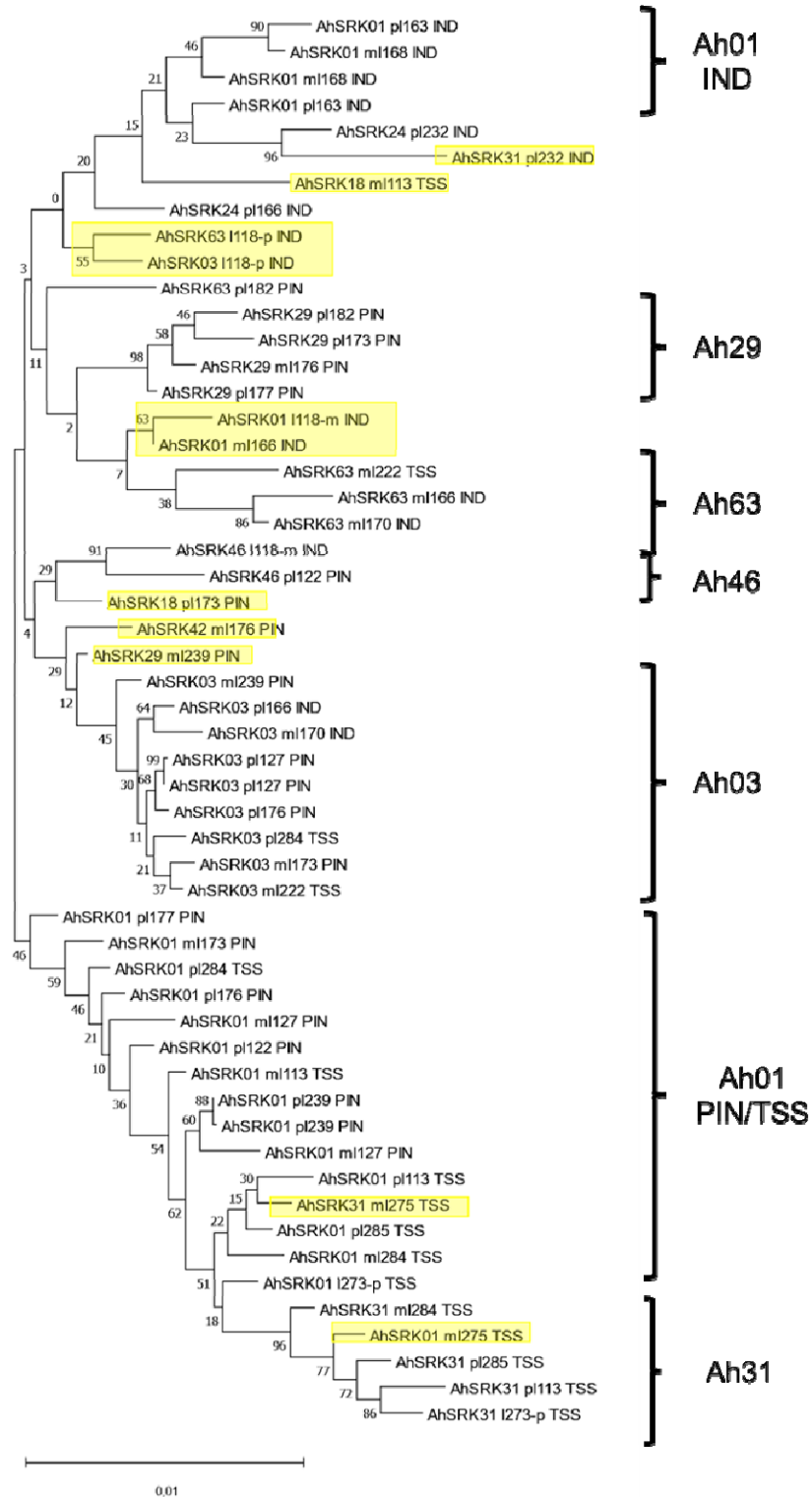
840

841 **Figure S3: Phylogenetic tree obtained by Maximum Likelihood for haplotypes of *A. halleri***  
842 **(populations Nivelles and Mortagne) across the first 25kb flanking the *S*-locus.**



843

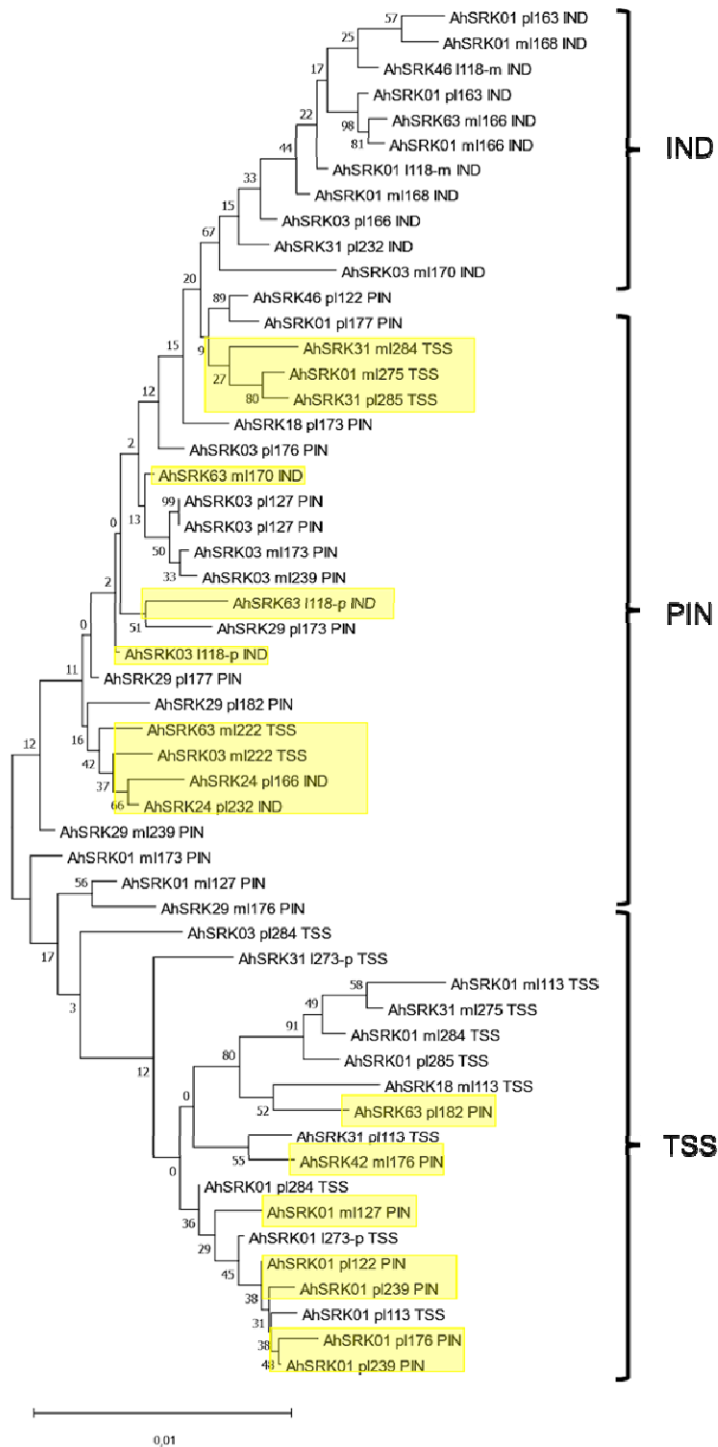
844 **Figure S4: Phylogenetic tree obtained by maximum likelihood for haplotypes of *A. halleri***  
845 **(populations Nivelles and Mortagne) based on the nucleotide positions between 25kb and 50kb**  
846 **away from the *S*-locus.**



847

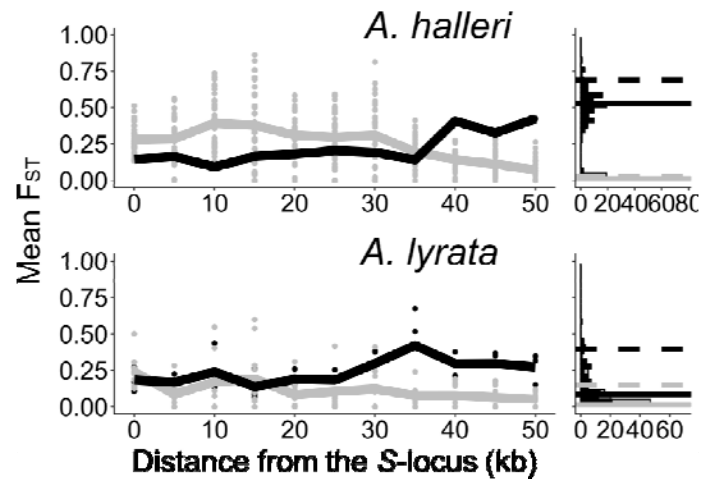
848 **Figure S5: Phylogenetic tree obtained by maximum likelihood for haplotypes of *A. lyrata***  
 849 **(populations PIN, IND, TSS) across the first 5kb flanking the S-locus.**

850



851

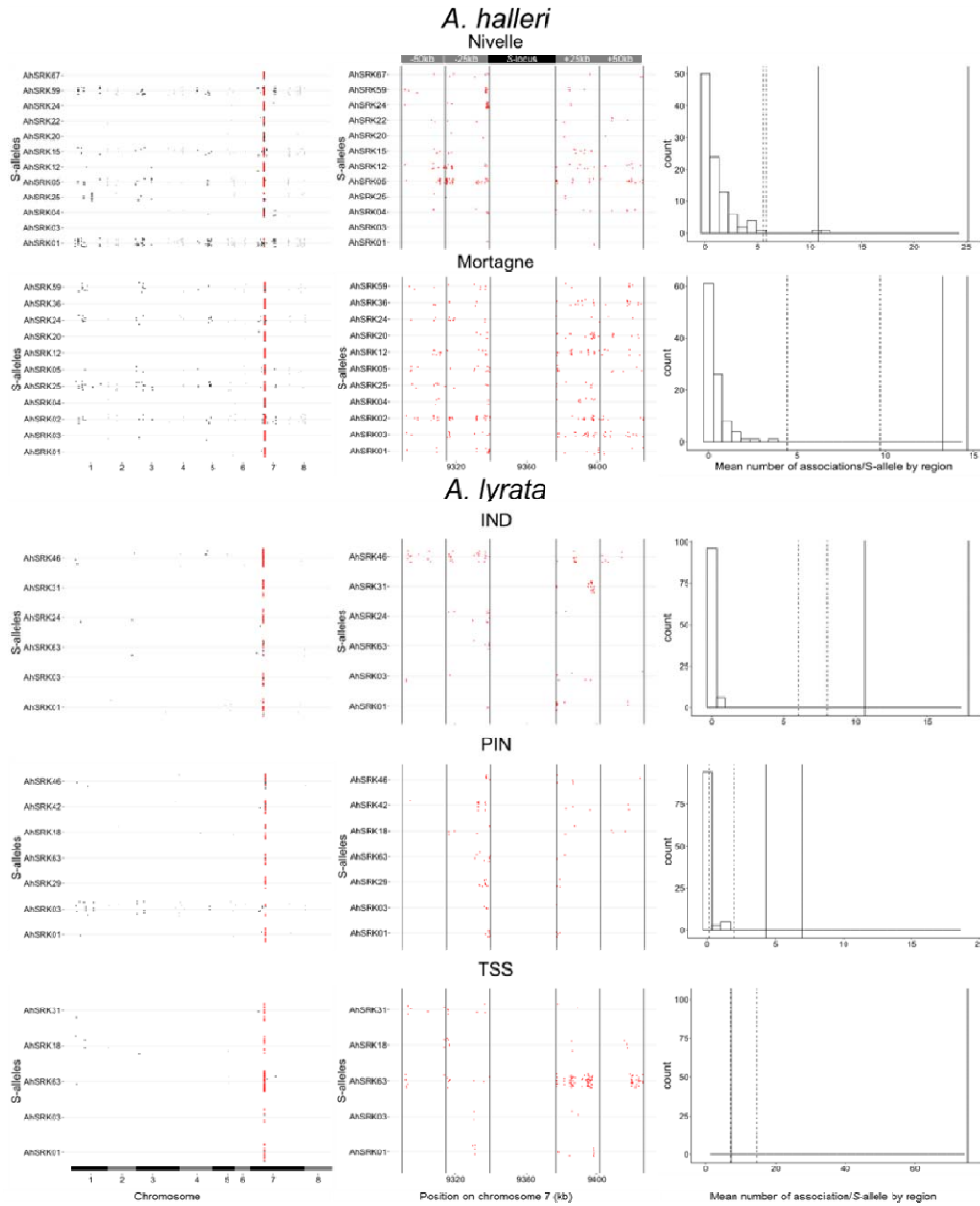
852 **Figure S6: Phylogenetic tree obtained by maximum likelihood for haplotypes of *A. lyrata***  
853 **(populations PIN, IND, TSS) based on the nucleotide positions between 5kb and 10kb away from**  
854 **the S-locus.**



855

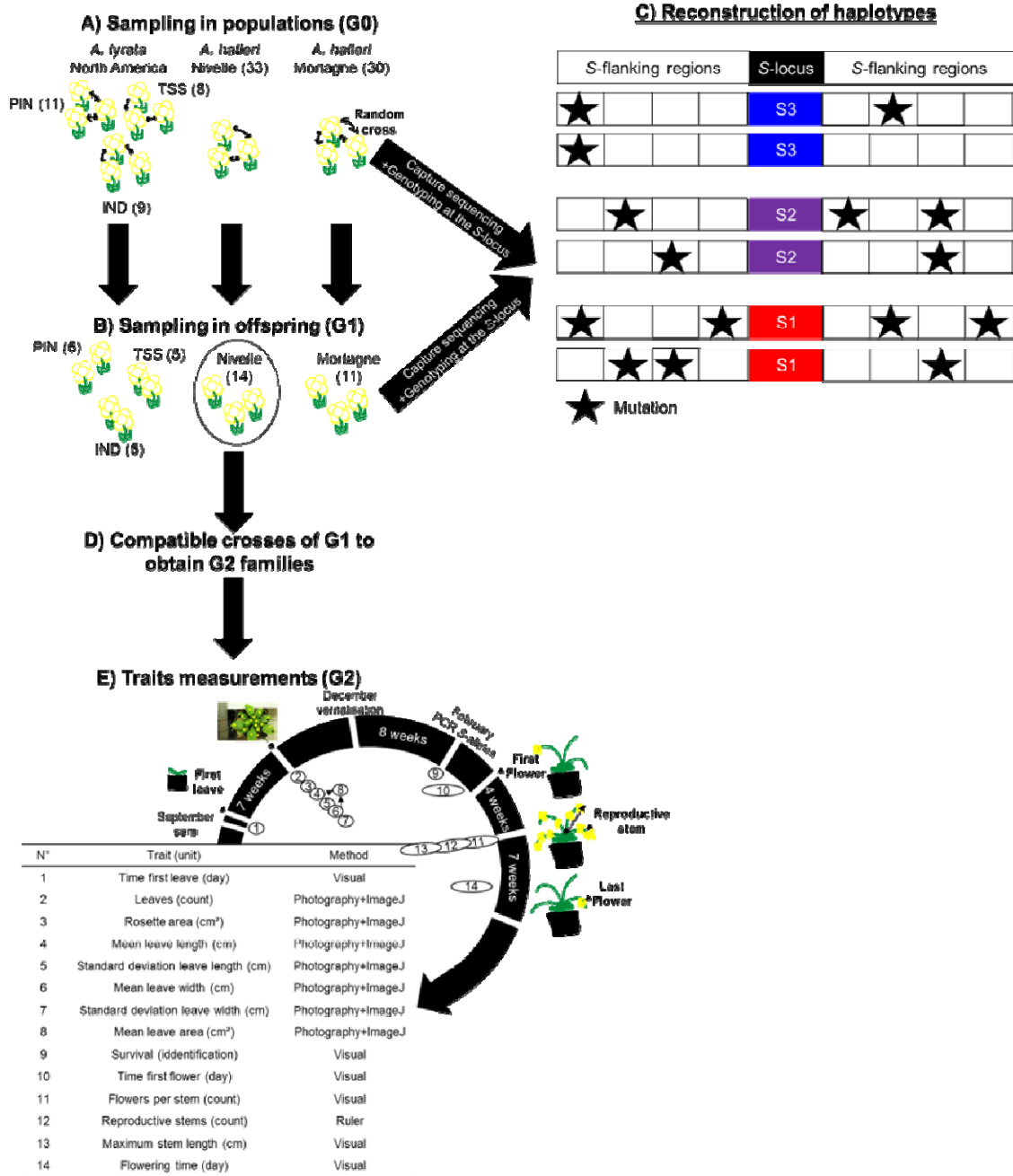
856 **Figure S7: The genetic structure of SNPs in the S-locus flanking regions in *A. halleri* and *A. lyrata*.**





857

858 **Figure S8: Patterns of genetic associations between S-alleles and SNPs across the genome.**



859

860 **Figure S9: Experimental protocol.**

861

862

863

864

865 **Supplemental items**

866 *Table S1: Crosses performed to obtain homozygotes for three S-alleles.*

867 *Table S2: Trait variation in S-locus homozygous individuals.*

868 *Table S3: Trait variation in homozygous at the S-locus for the S-alleles Ah01 and Ah04 between*  
869 *families.*

870 *Table S4: Effect of dominance on variation of phenotypic traits in S-locus homozygous individuals.*

871 *Table S5: Number of phased haplotypes linked to S-alleles in each sample.*

872 *Table S6: Effect of dominance on the accumulation of genetic load in S-flanking regions.*

873 *Table S7: S-locus genotypes of individuals sequenced using the capture protocol.*

874 *Table S8: S-locus genotypes of the offspring selected for haplotype phasing and the crosses for the*  
875 *study of phenotypic traits.*

876 *Table S9: Effect of phytopathogen, phytophagous attacks and oxidative stress on the phenotypic*  
877 *traits.*

878 *Table S10: Difference on the phenotypic traits variation between the two families for each allele*  
879 *tested.*

880

881

882

883

884

885 **Table S1: Crosses performed to obtain homozygotes for three S-alleles.** For each S-allele, we  
886 obtained homozygous and heterozygous plants within two separate families (listed as separate  
887 rows).

<b>Genotype pollen donor<sup>a</sup></b>	<b>Genotype stigma recipient</b>	<b>Homozygote studied</b>	<b>Dominance level</b>	<b>Number of fruits obtained</b>	<b>Number of seeds</b>
<b>Ah20/Ah01</b>	Ah12/Ah01	Ah01	1	5	48
<b>Ah20/Ah01</b>	Ah12/Ah01	Ah01	1	2	23
<b>Ah02/Ah03</b>	Ah03/Ah01	Ah03	2	6	69
<b>Ah24/Ah03</b>	Ah03/Ah01	Ah03	2	3	27
<b>Ah20/Ah04</b>	Ah04/Ah04	Ah04	3	21	167
<b>Ah20/Ah04</b>	Ah04/Ah04	Ah04	3	8	65

888 <sup>a</sup> The S-alleles in bold represent the dominant S-allele expressed on pollen of each donor genotype.  
889 Each line corresponds to one pair of individuals used as parents.

890

891

892

893 **Table S2: Trait variation in S-locus homozygous individuals.**

Trait (unit)	Mean heterozygotes (SD)	Mean homozygotes (SD)	P value <sup>a</sup>
Time first leaf (day)	15.7 (1.04)	14.79 (1.19)	0.14
Number of leaves	13.22 (0.65)	13.46 (0.96)	0.35
Rosette area (cm <sup>2</sup> )	34.99 (5.11)	35.5 (7.65)	0.46
Mean leaf length (cm)	1.86 (0.15)	1.86 (0.21)	0.49
St dev leaf length (cm)	0.28 (0.02)	0.29 (0.04)	0.31
Mean leaf width (cm)	1.49 (0.11)	1.46 (0.14)	0.36
St dev leaf width (cm)	0.27 (0.02)	0.25 (0.03)	0.24
Mean leaf area (cm <sup>2</sup> )	3.16 (0.43)	3.11 (0.6)	0.45
Time first flower (day)	36.28 (1.14)	34.98 (1.25)	0.07
Number of flowering stems	12.9 (0.91)	13.3 (1.32)	0.31
<b>Maximum flowering stem length (cm)</b>	<b>56.19 (2.26)</b>	<b>64.55 (3.17)</b>	<b>&lt;1e<sup>-04</sup></b>
Number of flowers by flowering stem	58.87 (5.96)	66.56 (10.71)	0.09
Flowering duration (day)	44.15 (1.7)	43.81 (1.88)	0.39

894 <sup>a</sup> The P value represents the proportions of the distribution with the expected difference of mean  
 895 between heterozygotes and homozygotes at the S-locus after 10,000 random resamples equal to or  
 896 less than the value observed. Significant values are represented in bold.

897

898

899

900

901

902 **Table S3: Trait variation in homozygous at the S-locus for the S-alleles Ah01 and Ah04 between**  
 903 **families.**

<b>Allele</b>	<b>Trait (unit)</b>	<b>Mean first family (SD)</b>	<b>Mean second family (SD)</b>	<b>p value</b>
Ah01 <sup>b</sup>	<b>Time first leaf (day)</b>	<b>17.11 ( 2.26 )</b>	<b>32 ( NA )</b>	<b>&lt;1.00<sup>e-04</sup></b>
	Number of leaves	14 ( 2.83 )	9 ( NA )	0.10
	Rosette area (cm <sup>2</sup> )	27.65 ( 18.22 )	14.21 ( NA )	0.50
	Mean leaf length (cm)	1.64 ( 0.68 )	1.52 ( NA )	0.49
	St dev leaf length (cm)	0.28 ( 0.11 )	0.17 ( NA )	0.30
	Mean leaf width (cm)	1.38 ( 0.43 )	1.38 ( NA )	0.49
	St dev leaf width (cm)	0.24 ( 0.07 )	0.16 ( NA )	0.30
	Mean leaf area (cm <sup>2</sup> )	2.52 ( 1.59 )	2.1 ( NA )	0.50
	Time first flower (day)	35.86 ( 3.13 )	36 ( NA )	0.25
	Number of flowering stems	12.29 ( 3.9 )	7 ( NA )	0.25
	<b>Maximum flowering stem length (cm)</b>	<b>43.43 ( 6.91 )</b>	<b>62.4 ( NA )</b>	<b>&lt;1.00<sup>e-04</sup></b>
	Number of flowers by flowering stem	31.73 ( 24.16 )	52.25 ( NA )	0.12
	<b>Flowering duration (day)</b>	<b>44.71 ( 6.73 )</b>	<b>52 ( NA )</b>	<b>&lt;1.00<sup>e-04</sup></b>
Ah04	Time first leaf (day)	13.15 ( 2.68 )	14.88 ( 4.36 )	0.11
	Number of leaves	13.43 ( 2.99 )	14.04 ( 3.9 )	0.24
	Rosette area (cm <sup>2</sup> )	36.63 ( 27.74 )	45.86 ( 33.48 )	0.08
	Mean leaf length (cm)	1.93 ( 0.81 )	2.13 ( 0.89 )	0.14
	St dev leaf length (cm)	0.29 ( 0.15 )	0.3 ( 0.12 )	0.33
	Mean leaf width (cm)	1.54 ( 0.57 )	1.44 ( 0.51 )	0.20
	<b>St dev leaf width (cm)</b>	<b>0.28 ( 0.13 )</b>	<b>0.22 ( 0.09 )</b>	<b>0.02</b>
	Mean leaf area (cm <sup>2</sup> )	3.4 ( 2.45 )	3.46 ( 2.27 )	0.45
	Time first flower (day)	35.57 ( 4.73 )	33.2 ( 3.66 )	0.06
	<b>Number of flowering stems</b>	<b>13.21 ( 4.25 )</b>	<b>16.4 ( 5.25 )</b>	<b>0.02</b>
	Maximum flowering stem length (cm)	64.98 ( 9.65 )	62.41 ( 9.34 )	0.13
	<b>Number of flowers by flowering stem</b>	<b>72.38 ( 33.07 )</b>	<b>45.25 ( 15.45 )</b>	<b>1.00<sup>e-04</sup></b>
	Flowering duration (day)	43.68 ( 6.96 )	41.56 ( 7.38 )	0.23

904 <sup>a</sup> The P value represents the proportions of the distribution with the expected difference of mean  
 905 between each family after 10,000 random resamples equal to or less than the observed value.  
 906 Significant values are represented in bold. The Ah03 S-allele is excluded because only two  
 907 homozygotes survived. <sup>b</sup> the standard deviation for the second family was not estimated because  
 908 only one homozygote survived.

909 **Table S4: Effect of dominance on variation of phenotypic traits in S-locus homozygous individuals.**

<b>Trait (unit)</b>	<b>Random effect implemented in GLM model</b>	<b>Distribution of trait tested</b>	<b>Linear effect</b>	<b>p value</b>
<b>Time first leaf (days)</b>	<b>Family</b>	<b>Poisson</b>	<b>-0.15</b>	<b>5.5<sup>e-04</sup></b>
Leaves (counts)	Oxydative stress	Poisson	0.006	0.91
Rosette area (cm <sup>2</sup> )	Oxydative stress + family	Gaussian	8.25	0.19
Mean leave length (cm)	Oxydative stress + family	Gaussian	0.17	0.28
St dev leave length (cm)	None	Gaussian	0.02	0.54
Mean leave width (cm)	Oxydative stress + phytophagous attack	Gaussian	0.05	0.83
St dev leave width (cm <sup>2</sup> )	Oxydative stress + phytophagous attack	Gaussian	0.01	0.91
Mean leave area (cm <sup>2</sup> )	Oxydative stress	Gaussian	0.39	0.36
Time first flower (days)	Phytopathogen + phytophagous attack + family	Poisson	-0.0002	1
Flowering stems (counts)	Family	Poisson	0.15	0.08
Maximum flowering stem length (cm)	Phytopathogen + family	Gaussian	7.6	0.06
Flowers by stem (counts)	Family	Poisson	0.15	0.44
Flowering duration (days)	Phytopathogen + family	Poisson	-0.05	0.1

910 Dominance at the S-locus was implemented as a fixed effect.

911

912

913

914 **Table S5: Number of phased haplotypes linked to S-alleles in each sample.**

Species	Population	Number of phased haplotypes	Number of S-alleles	Number of S-alleles with more than one copy	Mean number of gene copies per S-allele <sup>a</sup>
<i>A. halleri</i>	Nivelle	34	12	7	4.1
	Mortagne	38	11	9	4
<i>A. lyrata</i>	TSS	16	5	3	4.7
	IND	16	6	4	3.5
	PIN	22	7	3	6

915 <sup>a</sup> based on S-alleles with more than one copy.

916



917 **Table S6: Effect of dominance on the accumulation of genetic load in S-flanking regions.**

Type of mutations	Species	Genetic load estimated	Linear effect	p value
0fold	A. halleri	Number of fixed mutations by S-allele	0.08	2.57 <sup>e-U4</sup>
		Number of segregated mutations by S-allele	-0.49	2.21 <sup>e-U3</sup>
		Mean total number of mutation by S-allele ( $S_{of}$ )	0.01	0.54
	A. lyrata	Number of fixed mutations by S-allele	0.08	4.50 <sup>e-U3</sup>
		Number of segregated mutations by S-allele	-0.44	0.09
		Mean total number of mutation by S-allele ( $S_{of}$ )	-0.07	0.07
4fold	A. halleri	Number of fixed mutations by S-allele	0.23	8.95 <sup>e-U4</sup>
		Number of segregated mutations by S-allele	-0.48	2.02 <sup>e-U4</sup>
		Mean total number of mutation by S-allele ( $S_{4f}$ )	1.53 <sup>e-U3</sup>	0.95
	A. lyrata	Number of fixed mutations by S-allele	0.23	8.95 <sup>e-U4</sup>
		Number of segregated mutations by S-allele	-0.46	0.04
		Mean total number of mutation by S-allele ( $S_{4f}$ )	-0.04	0.32
$S_{of}/S_{4f}$	A. halleri	Ratio	0.02	0.42
	A. lyrata	Ratio	-0.11	0.11
Non-synonymous	A. halleri	Number of fixed mutations by S-allele	0.67	1.86 <sup>e-U4</sup>
		Number of segregated mutations by S-allele	-0.23	<2 <sup>e-U4</sup>
		Mean total number of mutation by S-allele ( $S_{NS}$ )	0.02	0.12
	A. lyrata	Number of fixed mutations by S-allele	0.24	1.74 <sup>e-U3</sup>
		Number of segregated mutations by S-allele	-0.28	<2 <sup>e-U3</sup>
		Mean total number of mutation by S-allele ( $S_{NS}$ )	7.99 <sup>e-U3</sup>	0.65
Synonymous	A. halleri	Number of fixed mutations by S-allele	0.08	4.55 <sup>e-U4</sup>
		Number of segregated mutations by S-allele	-0.26	<2 <sup>e-U4</sup>
		Mean total number of mutation by S-allele ( $S_s$ )	0.02	0.15
	A. lyrata	Number of fixed mutations by S-allele	0.27	2.3 <sup>e-U11</sup>
		Number of segregated mutations by S-allele	-0.34	<2 <sup>e-U10</sup>
		Mean total number of mutation by S-allele ( $S_s$ )	-0.01	0.33
$S_{NS}/S_s$	A. halleri	Ratio	-5.82 <sup>e-U3</sup>	0.81
	A. lyrata	Ratio	0.02	0.60

918

919 *Dominance at the S-locus was implemented as a fixed effect. Populations were implemented as random effects. The distribution of the genetic load*  
920 *estimated followed a Poisson distribution, except the ratios that followed normal distributions.*

921 **Table S7: S-locus genotypes of individuals sequenced using the capture protocol.**

<b>Individual</b>	<b>Allele 1<sup>a</sup></b>	<b>Allele 2<sup>a</sup></b>	<b>Population</b>	<b>Species</b>	<b>SRA</b>
<i>Mor_19_13</i>	Ah04 (III)	Ah25 (III)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088356
<i>Mor_19_14</i>	Ah04 (III)	Ah01 (I)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20844087
<i>Mor_19_19</i>	Ah12 (IV)	Ah36 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088358
<i>Mor_19_2</i>	Ah03 (II)	Ah24 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088359
<i>Mor_19_23</i>	Ah20 (IV)	Ah25 (III)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20844088
<i>Mor_19_24</i>	Ah12 (IV)	Ah24 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088362
<i>Mor_19_3</i>	Ah03 (II)	Ah01 (I)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20844098
<i>Mor_19_37</i>	Ah59 (IV)	Ah20 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088357
<i>Mor_19_38</i>	Ah36 (IV)	Ah01 (I)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088371
<i>Mor_19_4</i>	Ah25 (III)	Ah12 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088360
<i>Mor_19_41</i>	Ah25 (III)	Ah25 (III)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088372
<i>Mor_19_42</i>	Ah12 (IV)	Ah04 (III)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088373
<i>Mor_19_45</i>	Ah36 (IV)	Ah02 (III)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088374
<i>Mor_19_46</i>	Ah03 (II)	Ah12 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088375
<i>Mor_19_51</i>	Ah20 (IV)	Ah05 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088361
<i>Mor_19_53</i>	Ah36 (IV)	Ah12 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088365
<i>Mor_19_54</i>	Ah24 (IV)	Ah12 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088366
<i>Mor_19_55</i>	Ah20 (IV)	Ah05 (IV)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088377
<i>Mor_19_56</i>	Ah12 (IV)	Ah03 (II)	<i>Mortagne</i>	<i>A. halleri</i>	SAMN20088378
<i>Niv_19_18</i>	Ah15 (IV)	Ah03 (II)	<i>Nivelle</i>	<i>A. halleri</i>	SAMN20088349
<i>Niv_19_19</i>	Ah20 (IV)	Ah24 (IV)	<i>Nivelle</i>	<i>A. halleri</i>	SAMN20088350
<i>Niv_19_22</i>	Ah24 (IV)	Ah01 (I)	<i>Nivelle</i>	<i>A. halleri</i>	SAMN20088351
<i>Niv_19_23</i>	Ah20 (IV)	Ah12 (IV)	<i>Nivelle</i>	<i>A. halleri</i>	SAMN20088352
<i>Niv_19_3</i>	Ah04 (III)	Ah59 (IV)	<i>Nivelle</i>	<i>A. halleri</i>	SAMN20088336
<i>Niv_19_31</i>	Ah04 (III)	Ah24 (IV)	<i>Nivelle</i>	<i>A. halleri</i>	SAMN20088337

<i>Niv_19_4</i>	<i>Ah01 (I)</i>	<i>Ah24 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088382</i>
<i>Niv_19_42</i>	<i>Ah22 (IV)</i>	<i>Ah12 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088341</i>
<i>Niv_19_45</i>	<i>Ah12 (IV)</i>	<i>Ah04 (III)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088344</i>
<i>Niv_19_5</i>	<i>Ah25 (III)</i>	<i>Ah67 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088340</i>
<i>Niv_19_52</i>	<i>Ah20 (IV)</i>	<i>Ah22 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088331</i>
<i>Niv_19_53</i>	<i>Ah05 (IV)</i>	<i>Ah20 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088339</i>
<i>Niv_19_54</i>	<i>Ah01 (I)</i>	<i>Ah12 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20844097</i>
<i>Niv_19_58</i>	<i>Ah02 (III)</i>	<i>Ah59 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088343</i>
<i>Niv_19_59</i>	<i>Ah01 (I)</i>	<i>Ah67 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20844096</i>
<i>Niv_19_60</i>	<i>Ah04 (III)</i>	<i>Ah20 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088332</i>
<i>Niv_19_7</i>	<i>Ah04 (III)</i>	<i>Ah67 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088345</i>
<i>Niv_19_8</i>	<i>Ah04 (III)</i>	<i>Ah12 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20844090</i>
<i>Niv_19_9</i>	<i>Ah04 (III)</i>	<i>Ah67 (IV)</i>	<i>Nivelle</i>	<i>A. halleri</i>	<i>SAMN20088338</i>
<i>Pin_15_1</i>	<i>Ah03 (II)</i>	<i>Ah01 (I)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088324</i>
<i>Pin_16_1</i>	<i>Ah01 (I)</i>	<i>Ah46 (IV)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088318</i>
<i>Pin_4_24</i>	<i>Ah03 (II)</i>	<i>Ah29 (III)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088314</i>
<i>Pin_16_3</i>	<i>Ah18 (IV)</i>	<i>Ah29 (III)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088323</i>
<i>Pin_4_54</i>	<i>Ah03 (II)</i>	<i>Ah03 (II)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088319</i>
<i>Pin_5_1</i>	<i>Ah01 (I)</i>	<i>Ah01 (I)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20844102</i>
<i>Pin_5_12</i>	<i>Ah29 (III)</i>	<i>Ah01 (I)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088325</i>
<i>Pin_5_2</i>	<i>Ah03 (II)</i>	<i>Ah01 (I)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088311</i>
<i>Pin_8_15</i>	<i>Ah01 (I)</i>	<i>Ah01 (I)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088328</i>
<i>Pin_8_2</i>	<i>Ah29 (III)</i>	<i>Ah63 (III)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088326</i>
<i>Pin_9_1</i>	<i>Ah42 (IV)</i>	<i>Ah29 (III)</i>	<i>PIN</i>	<i>A. lyrata</i>	<i>SAMN20088312</i>
<i>Tss_14_3</i>	<i>Ah01 (I)</i>	<i>Ah31 (IV)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088315</i>
<i>Tss_21_10</i>	<i>Ah01 (I)</i>	<i>Ah03 (II)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088329</i>
<i>Tss_22_24</i>	<i>Ah01 (I)</i>	<i>Ah31 (IV)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088306</i>

<i>Tss_22_7</i>	<i>Ah01 (I)</i>	<i>Ah31 (IV)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088317</i>
<i>Tss_3_10</i>	<i>Ah31 (IV)</i>	<i>Ah01 (I)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088330</i>
<i>Tss_23_2</i>	<i>Ah03 (II)</i>	<i>Ah63 (III)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088313</i>
<i>Tss_3_23</i>	<i>Ah18 (IV)</i>	<i>Ah01 (I)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088307</i>
<i>Tss_5_1</i>	<i>Ah01 (I)</i>	<i>Ah31 (IV)</i>	<i>TSS</i>	<i>A. lyrata</i>	<i>SAMN20088316</i>
<i>Ind_1_1</i>	<i>Ah01 (I)</i>	<i>Ah01 (I)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088320</i>
<i>Ind_10_3</i>	<i>Ah31 (IV)</i>	<i>Ah24 (IV)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088327</i>
<i>Ind_15_1</i>	<i>Ah03 (II)</i>	<i>Ah24 (IV)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088321</i>
<i>Ind_15_2</i>	<i>Ah63 (III)</i>	<i>Ah03 (II)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088309</i>
<i>Ind_9_3</i>	<i>Ah01 (I)</i>	<i>Ah46 (IV)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088322</i>
<i>Ind_18_1</i>	<i>Ah01 (I)</i>	<i>Ah01 (I)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20844103</i>
<i>Ind_6_1</i>	<i>Ah01 (I)</i>	<i>Ah63 (III)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088308</i>
<i>Ind_8_1</i>	<i>Ah03 (II)</i>	<i>Ah63 (III)</i>	<i>IND</i>	<i>A. lyrata</i>	<i>SAMN20088305</i>

922 <sup>a</sup> *The individuals were mainly sequenced by capture approach. The genotypes of homozygotes were*  
923 *confirmed after a whole genome sequencing. The class of dominance is specified in parenthesis.*

924

925

926

927

928 **Table S8: S-locus genotypes of the offspring selected for haplotype phasing and the crosses for the**  
 929 **study of phenotypic traits.**

<i>Identity<sup>a</sup></i>	<i>Allele 1</i>	<i>Allele 2</i>	<i>Pollen donor<sup>f</sup></i>	<i>Stigma recipient<sup>c</sup></i>	<i>SRA</i>
<i>d32</i>	<i>Ah03</i>	<i>Ah12</i>	<i>Mor_19_2</i>	<i>Mor_19_19</i>	<i>SAMN20844104</i>
<i>d33</i>	<i>Ah01</i>	<i>Ah12</i>	<i>Mor_19_3</i>	<i>Mor_19_4</i>	<i>SAMN20844105</i>
<i>d38<sup>d</sup></i>	<i>Ah04</i>	<i>Ah04</i>	<i>Mor_19_13</i>	<i>Mor_19_14</i>	<i>SAMN20844106</i>
<i>d42</i>	<i>Ah20</i>	<i>Ah12</i>	<i>Mor_19_23</i>	<i>Mor_19_24</i>	<i>SAMN20844107</i>
<i>d72</i>	<i>Ah20</i>	<i>Ah36</i>	<i>Mor_19_37</i>	<i>Mor_19_38</i>	<i>SAMN20844108</i>
<i>d74</i>	<i>Ah25</i>	<i>Ah12</i>	<i>Mor_19_41</i>	<i>Mor_19_42</i>	<i>SAMN20844109</i>
<i>d76</i>	<i>Ah03</i>	<i>Ah36</i>	<i>Mor_19_46</i>	<i>Mor_19_45</i>	<i>SAMN20844110</i>
<i>d89</i>	<i>Ah03</i>	<i>Ah05</i>	<i>Mor_19_56</i>	<i>Mor_19_55</i>	<i>SAMN20844111</i>
<i>d217</i>	<i>Ah12</i>	<i>Ah59</i>	<i>Mor_19_19</i>	<i>Mor_19_37</i>	<i>SAMN20844112</i>
<i>d250</i>	<i>Ah12</i>	<i>Ah20</i>	<i>Mor_19_24</i>	<i>Mor_19_51</i>	<i>SAMN20844113</i>
<i>d265</i>	<i>Ah36</i>	<i>Ah24</i>	<i>Mor_19_53</i>	<i>Mor_19_54</i>	<i>SAMN20844114</i>
<i>d3.1</i>	<i>Ah24</i>	<i>Ah25</i>	<i>Niv_19_4</i>	<i>Niv_19_5</i>	<i>SAMN20844115</i>
<i>d10.1<sup>f</sup></i>	<i>Ah03</i>	<i>Ah24</i>	<i>Niv_19_18</i>	<i>Niv_19_19</i>	<i>SAMN20844116</i>
<i>d12.1<sup>f</sup></i>	<i>Ah01</i>	<i>Ah20</i>	<i>Niv_19_22</i>	<i>Niv_19_23</i>	<i>SAMN20844117</i>
<i>d30.1</i>	<i>Ah02</i>	<i>Ah01</i>	<i>Niv_19_58</i>	<i>Niv_19_59</i>	<i>SAMN20844118</i>
<i>d48.1</i>	<i>Ah20</i>	<i>Ah25</i>	<i>Niv_19_52</i>	<i>Niv_19_5</i>	<i>SAMN20844119</i>
<i>d50.1<sup>e</sup></i>	<i>Ah12</i>	<i>Ah04</i>	<i>Niv_19_8</i>	<i>Niv_19_60</i>	<i>SAMN20844120</i>
<i>d51.1</i>	<i>Ah12</i>	<i>Ah04</i>	<i>Niv_19_54</i>	<i>Niv_19_7</i>	<i>SAMN20844121</i>
<i>d191.1<sup>e,f</sup></i>	<i>Ah04</i>	<i>Ah04</i>	<i>Niv_19_3</i>	<i>Niv_19_31</i>	<i>SAMN20844122</i>
<i>d206.1</i>	<i>Ah22</i>	<i>Ah25</i>	<i>Niv_19_42</i>	<i>Niv_19_5</i>	<i>SAMN20844123</i>
<i>d205.1<sup>f</sup></i>	<i>Ah20</i>	<i>Ah04</i>	<i>Niv_19_53</i>	<i>Niv_19_9</i>	<i>SAMN20844124</i>
<b><i>d208.1<sup>f</sup></i></b>	<b><i>Ah03</i></b>	<b><i>Ah02</i></b>	<b><i>Niv_19_58</i></b>	<b><i>Niv_19_18</i></b>	<b><i>SAMN20844125</i></b>
<b><i>d17.1<sup>f</sup></i></b>	<b><i>Ah03</i></b>	<b><i>Ah01</i></b>	<b><i>Niv_19_33</i></b>	<b><i>Niv_19_32<sup>b</sup></i></b>	<b><i>SAMN20844126</i></b>
<b><i>d24.1<sup>f</sup></i></b>	<b><i>Ah20</i></b>	<b><i>Ah01</i></b>	<b><i>Niv_19_47<sup>b</sup></i></b>	<b><i>Niv_19_46<sup>b</sup></i></b>	<b><i>SAMN20844127</i></b>

<b>d29.1<sup>f</sup></b>	<b>Ah12</b>	<b>Ah01</b>	<b><i>Niv_19_57<sup>b</sup></i></b>	<b><i>Niv_19_56<sup>b</sup></i></b>	<b>SAMN20844128</b>
<b>d50.2<sup>f</sup></b>	<b>Ah04</b>	<b>Ah04</b>	<b><i>Niv_19_8</i></b>	<b><i>Niv_19_60</i></b>	<b>SAMN20844129</b>
d266.1	Ah04	Ah12	<i>Niv_19_7</i>	<i>Niv_19_45</i>	SAMN20844130
d122	Ah01	Ah63	<i>Pin_16_1</i>	<i>Pin_8_2</i>	SAMN20844131
d127	Ah01	Ah03	<i>Pin_5_1</i>	<i>Pin_4_54</i>	SAMN20844132
d173	Ah03	Ah18	<i>Pin_5_2</i>	<i>Pin_16_3</i>	SAMN20844133
d176	Ah03	Ah42	<i>Pin_15_1</i>	<i>Pin_9_1</i>	SAMN20844134
d177	Ah01	Ah29	<i>Pin_16_1</i>	<i>Pin_5_12</i>	SAMN20844135
d239	Ah29	Ah01	<i>Pin_4_24</i>	<i>Pin_8_15</i>	SAMN20844136
d113	Ah18	Ah31	<i>Tss_3_23</i>	<i>Tss_22_7</i>	SAMN20844137
d273	Ah03	Ah01	<i>Tss_23_2</i>	<i>Tss_22_24</i>	SAMN20844138
d275	Ah63	Ah31	<i>Tss_23_2</i>	<i>Tss_14_3</i>	SAMN20844139
d284	Ah01	Ah31	<i>Tss_21_10</i>	<i>Tss_5_1</i>	SAMN20844140
d285	Ah01	Ah31	<i>Tss_21_10</i>	<i>Tss_3_10</i>	SAMN20844141
d118	Ah03	Ah01	<i>Ind_9_3</i>	<i>Ind_8_1</i>	SAMN20844142
d163	Ah03	Ah01	<i>Ind_8_1</i>	<i>Ind_1_1</i>	SAMN20844143
d166	Ah03	Ah01	<i>Ind_6_1</i>	<i>Ind_15_1</i>	SAMN20844144
d170	Ah01	Ah63	<i>Ind_9_3</i>	<i>Ind_15_2</i>	SAMN20844145
d232	Ah01	Ah31	<i>Ind_18_1</i>	<i>Ind_10_3</i>	SAMN20844146

930 <sup>a</sup> Offspring figured in bold were not used to reconstitute haplotype of parents. <sup>b</sup> Parents were not  
 931 sequenced. <sup>c</sup> Italics indicate parental haplotypes that were suppressed from the dataset because they  
 932 had already been phased through another offspring. <sup>d</sup> The homozygous genotype of d38 was  
 933 confirmed by genome-wide sequencing. <sup>e</sup> The homozygous genotypes of d191 and d50.1 were  
 934 confirmed by the proportion of homozygous S-alleles in their offspring found by PCR after cross. <sup>f</sup> The  
 935 individuals used for crosses in the phenotypic experiments.

936

937

938

939

940 **Table S9: Effect of phytopathogen, phytophagous attacks and oxidative stress on the phenotypic**  
 941 **traits.**

<i>Effect tested</i>	<i>Trait</i>	<i>Mean with attack (sd)</i>	<i>Mean without attack (sd)</i>	<i>p value<sup>a</sup></i>
	<i>Time first leaf (day)</i>	15.37 (4.84)	15.38 (5.24)	0.49
	<i>Number of leaves</i>	13.03 (2.9)	13.46 (3.71)	0.23
	<i>Rosette area (cm<sup>2</sup>)</i>	31.93 (27.19)	37.08 (27.14)	0.13
	<i>Mean leave length (cm)</i>	1.78 (0.78)	1.91 (0.8)	0.16
	<i>St dev leave length (cm)</i>	0.29 (0.14)	0.28 (0.13)	0.38
	<i>Mean leave width (cm)</i>	1.42 (0.56)	1.51 (0.53)	0.17
	<i>St dev leave width (cm)</i>	0.25 (0.11)	0.27 (0.12)	0.23
	<i>Mean leave area (cm<sup>2</sup>)</i>	2.94 (2.33)	3.25 (2.15)	0.2
	<b><i>Time first flower (day)</i></b>	<b>37.43 (4.56)</b>	<b>31.34 (5.55)</b>	<b>2e<sup>-03</sup></b>
	<i>Number of flowering stems</i>	12.05 (4.4)	11.84 (4.62)	0.05
	<i>Maximum flowering stem length (cm)</i>	56.78 (12.04)	52.55 (11.82)	0.1
	<i>Number of flowers by flowering stem</i>	56.56 (30.99)	56.22 (34.55)	0.11
<i>Phytopathogen</i>	<b><i>Flowering duration (day)</i></b>	<b>41.44 (8.13)</b>	<b>39.51 (7.32)</b>	<b>9e<sup>-03</sup></b>
	<i>Time first leaf (day)</i>	14.48 (4.4)	15.56 (5.2)	0.16
	<i>Number of leaves</i>	12.59 (3.76)	13.44 (3.35)	0.13
	<i>Rosette area (cm<sup>2</sup>)</i>	32.15 (32.24)	35.77 (26.16)	0.27
	<i>Mean leave length (cm)</i>	1.68 (0.83)	1.9 (0.78)	0.1
	<i>St dev leave length (cm)</i>	0.29 (0.1)	0.28 (0.14)	0.36
	<b><i>Mean leave width (cm)</i></b>	<b>1.29 (0.55)</b>	<b>1.51 (0.53)</b>	<b>0.03</b>
	<b><i>St dev leave width (cm)</i></b>	<b>0.22 (0.1)</b>	<b>0.27 (0.12)</b>	<b>0.03</b>
	<i>Mean leave area (cm<sup>2</sup>)</i>	2.59 (2.3)	3.25 (2.19)	0.08
	<b><i>Time first flower (day)</i></b>	<b>34.78 (6.93)</b>	<b>33.36 (4.92)</b>	<b>0.04</b>
<i>Phytophagous</i>	<i>Number of flowering stems</i>	10.78 (4.76)	12.15 (4.55)	0.34



	<b>Maximum flowering stem length (cm)</b>	<b>53.53 (9.81)</b>	<b>54.24 (12.2)</b>	<b>0.05</b>
	Number of flowers by flowering stem	47.46 (25.45)	58.12 (34.46)	0.17
	Flowering duration (day)	37.56 (8.1)	40.76 (7.76)	0.45
	Time first leaf (day)	14.92 (3.93)	15.46 (5.28)	0.34
	<b>Number of leaves</b>	<b>10.04 (3.88)</b>	<b>13.93 (2.96)</b>	<b>1e<sup>-04</sup></b>
	<b>Rosette area (cm<sup>2</sup>)</b>	<b>11.79 (16.54)</b>	<b>39.64 (26.56)</b>	<b>1e<sup>-04</sup></b>
	<b>Mean leave length (cm)</b>	<b>1.05 (0.49)</b>	<b>2.02 (0.74)</b>	<b>1e<sup>-04</sup></b>
	St dev leave length (cm)	0.25 (0.1)	0.29 (0.14)	0.09
	<b>Mean leave width (cm)</b>	<b>0.88 (0.31)</b>	<b>1.59 (0.5)</b>	<b>1e<sup>-04</sup></b>
	<b>St dev leave width (cm)</b>	<b>0.23 (0.12)</b>	<b>0.27 (0.12)</b>	<b>0.04</b>
	<b>Mean leave area (cm<sup>2</sup>)</b>	<b>1.05 (0.98)</b>	<b>3.54 (2.17)</b>	<b>1e<sup>-04</sup></b>
	Time first flower (day)	31.77 (5.6)	33.95 (5.25)	0.06
	Number of flowering stems	10.38 (4.85)	12.21 (4.54)	0.44
	Maximum flowering stem length (cm)	50.3 (13.1)	54.85 (11.71)	0.11
	Number of flowers by flowering stem	53.97 (55.98)	56.8 (28.05)	0.21
Oxidative stress	Flowering duration (day)	37.35 (6.5)	40.78 (7.94)	0.08

942 <sup>a</sup> The P value represents the proportions of the distribution with the expected difference of mean  
 943 between individuals with and without attack after 10,000 random resamples equal to or less than the  
 944 value observed. Significant values are represented in bold.

945

946 **Table S10: Difference on the phenotypic trait variation between the two families for each allele**  
 947 **tested.**

<b>Allele</b>	<b>Family (Male/Female)</b>	<b>Trait (unit)</b>	<b>Mean (sd)</b>	<b>p value<sup>a</sup></b>
Ah01	1 (d29.1/d24.1)	<b>Time first leaf (day)</b>	<b>20.8 (5.43)</b>	<b>2e<sup>-03</sup></b>
		Number of leaves	11.9 (3.75)	0.09
		Rosette area (cm <sup>2</sup> )	24.9 (25.58)	0.11
		Mean leaf length (cm)	1.53 (0.79)	0.08
		St dev leaf length (cm)	0.21 (0.09)	0.03
		Mean leaf width (cm)	1.4 (0.58)	0.33
		St dev leaf width (cm)	0.24 (0.14)	0.32
		Mean leaf area (cm <sup>2</sup> )	2.47 (1.93)	0.17
		Time first flower (day)	28.8 (5.29)	0.44
		<b>Number of flowering stems</b>	<b>8 (4.07)</b>	<b>0.02</b>
		<b>Maximum flowering stem length (cm)</b>	<b>40.65 (6.76)</b>	<b>0.02</b>
		<b>Number of flowers by flowering stem</b>	<b>35.85 (10.72)</b>	<b>0.04</b>
		<b>Flowering duration (day)</b>	<b>40.35 (12.45)</b>	<b>8e<sup>-03</sup></b>
		Ah01	2 (d12.1/d29.1)	<b>Time first leaf (day)</b>
Number of leafs	13.45 (3.22)			0.42
Rosette area (cm <sup>2</sup> )	31.99 (22.61)			0.25
Mean leaf length (cm)	1.74 (0.71)			0.19
St dev leaf length (cm)	0.29 (0.1)			0.42
Mean leaf width (cm)	1.44 (0.51)			0.34
St dev leaf width (cm)	0.29 (0.11)			0.11
Mean leaf area (cm <sup>2</sup> )	2.84 (1.89)			0.21
<b>Time first flower (day)</b>	<b>29.86 (5.43)</b>			<b>0.04</b>
Number of flowering stems	9.59 (3.54)			0.32
<b>Maximum flowering stem length (cm)</b>	<b>34.85 (9.35)</b>			<b>1e<sup>-04</sup></b>
<b>Number of flowers by flowering stem</b>	<b>30.27 (25.45)</b>			<b>1e<sup>-04</sup></b>

		<i>Flowering duration (day)</i>	31.62 (7.17)	0.06
Ah03	3 (d208.1/d17.1)	<i>Time first leaf (day)</i>	13.33 (5.27)	0.22
		<i>Number of leaves</i>	12.05 (4.25)	0.17
		<i>Rosette area (cm<sup>2</sup>)</i>	31.23 (25.7)	0.35
		<i>Mean leaf length (cm)</i>	1.67 (0.78)	0.32
		<i>St dev leaf length (cm)</i>	0.28 (0.15)	0.42
		<i>Mean leaf width (cm)</i>	1.39 (0.62)	0.5
		<i>St dev leaf width (cm)</i>	0.24 (0.13)	0.42
		<i>Mean leaf area (cm<sup>2</sup>)</i>	2.87 (2.24)	0.45
		<b><i>Time first flower (day)</i></b>	<b>34.81 (7.87)</b>	<b>0.02</b>
		<b><i>Number of flowering stems</i></b>	<b>7.67 (3)</b>	<b>3e<sup>-04</sup></b>
		<b><i>Maximum flowering stem length (cm)</i></b>	<b>44.4 (13.87)</b>	<b>0.03</b>
		<i>Number of flowers by flowering stem</i>	61.55 (46.06)	0.06
		<b><i>Flowering duration (day)</i></b>	<b>39 (6.6)</b>	<b>5e<sup>-03</sup></b>
Ah03	4 (d10.1/d17.1)	<i>Time first leaf (day)</i>	18 (6.2)	0.12
		<i>Number of leaves</i>	12.33 (4.46)	0.22
		<b><i>Rosette area (cm<sup>2</sup>)</i></b>	<b>17.71 (11.23)</b>	<b>0.04</b>
		<i>Mean leaf length (cm)</i>	1.52 (0.53)	0.14
		<i>St dev leaf length (cm)</i>	0.23 (0.07)	0.14
		<i>Mean leaf width (cm)</i>	1.23 (0.35)	0.13
		<i>St dev leaf width (cm)</i>	0.23 (0.08)	0.23
		<i>Mean leaf area (cm<sup>2</sup>)</i>	2.01 (1.34)	0.1
		<i>Time first flower (day)</i>	33.33 (4.23)	0.09
		<i>Number of flowering stems</i>	11.67 (4.23)	0.22

		Maximum flowering stem length (cm)	57.25 (3.69)	0.34
		Number of flowers by flowering stem	75.05 (29.69)	0.12
		Flowering duration (day)	45.5 (8.67)	0.33
<i>Ah04</i>	5 (d205.1/d191.1)	<b>Time first leaf (day)</b>	<b>13.42 (3.28)</b>	<b>1e<sup>-04</sup></b>
		Number of leaves	13.45 (2.97)	0.32
		Rosette area (cm <sup>2</sup> )	36.29 (27.39)	0.31
		Mean leaf length (cm)	1.92 (0.8)	0.22
		St dev leaf length (cm)	0.29 (0.15)	0.34
		Mean leaf width (cm)	1.54 (0.56)	0.11
		St dev leaf width (cm)	0.27 (0.13)	0.15
		Mean leaf area (cm <sup>2</sup> )	3.36 (2.41)	0.14
		Time first flower (day)	35.1 (4.66)	0.33
		Number of flowering stems	12.87 (4.16)	0.32
		<b>Maximum flowering stem length (cm)</b>	<b>62.96 (9.48)</b>	<b>1e<sup>-04</sup></b>
		<b>Number of flowers by flowering stem</b>	<b>70.13 (32.5)</b>	<b>1e<sup>-04</sup></b>
		Flowering duration (day)	42.44 (7.05)	0.3
<i>Ah04</i>	6 (d205.1/d50.2)	Time first leaf (day)	14.31 (4.36)	0.30
		Number of leaves	13.5 (3.9)	0.13
		<b>Rosette area (cm<sup>2</sup>)</b>	<b>44.1 (33.48)</b>	<b>0.02</b>
		<b>Mean leaf length (cm)</b>	<b>2.04 (0.89)</b>	<b>0.04</b>
		St dev leaf length (cm)	0.29 (0.12)	0.21
		Mean leaf width (cm)	1.38 (0.51)	0.36
		<b>St dev leaf width (cm)</b>	<b>0.21 (0.09)</b>	<b>0.03</b>
		Mean leaf area (cm <sup>2</sup> )	3.33 (2.27)	0.21
		<b>Time first flower (day)</b>	<b>31.92 (3.66)</b>	<b>2e<sup>-03</sup></b>
		<b>Number of flowering stems</b>	<b>15.77 (5.25)</b>	<b>1e<sup>-04</sup></b>
		Maximum flowering stem length (cm)	60.01 (9.34)	0.07

<b>Number of flowers by flowering stem</b>	<b>43.51 (15.45)</b>	<b><math>5e^{-04}</math></b>
<b>Flowering duration (day)</b>	<b>39.96 (7.38)</b>	<b>0.04</b>

948 <sup>a</sup> The P value represents the proportions of the distribution with the expected difference of mean  
949 between each family and the other individuals after 10,000 random resamples equal to or less than  
950 the value observed. Significant values are represented in bold.

951

## 952 References

- 953 1. Katsura, Y., Iwase, M., Satta, Y. (2012) Evolution of genomic structures on mammalian sex  
954 chromosomes. *Current Genomics* 13:115–123. <https://doi.org/10.2174/138920212799860625>.
- 955 2. Hartmann, F.E., Duhamel, M., Carpentier, F., et al. (2021) Recombination suppression and  
956 evolutionary strata around mating-type loci in fungi: documenting patterns and understanding  
957 evolutionary and mechanistic causes. *New Phytologist* 229:2470–2491.  
958 <https://doi.org/10.1111/nph.17039>.
- 959 3. Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D.S., Hoepfner, M.P., Kerje, S.,  
960 Gustafson, U., Shi, C., Zhang, H., Chen, W., Liang, X., Huang, L., Wang, J., Liang, E., Wu, Q., Lee, S.M.-  
961 Y., Xu, X., Höglund, J., Liu, X., and Andersson, L. (2016). Structural genomic changes underlie  
962 alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat Genet* 48, 84–88.  
963 <https://doi.org/10.1038/ng.3430>.
- 964 4. Cocker, J.M., Wright, J., Li, J., Swarbreck, D., Dyer, S., Caccamo, M., and Gilmartin, P.M. (2018).  
965 *Primula vulgaris* (primrose) genome assembly, annotation and gene expression, with comparative  
966 genomics on the heterostyly supergene. *Sci Rep* 8, 17942. [https://doi.org/10.1038/s41598-018-](https://doi.org/10.1038/s41598-018-36304-4)  
967 [36304-4](https://doi.org/10.1038/s41598-018-36304-4).
- 968 5. Uyenoyama, M.K. (1997). Genealogical structure among alleles regulating self-incompatibility in  
969 natural populations of flowering plants. *Genetics* 147, 1389–1400.  
970 <https://doi.org/10.1093/genetics/147.3.1389>.
- 971 6. Uyenoyama, M.K. (2005). Evolution under tight linkage to mating type. *New Phytol.* 165: 63–70.  
972 <https://doi.org/10.1111/j.1469-8137.2004.01246.x>.
- 973 7. Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., and Joron, M. (2021).  
974 Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms.  
975 *Nat Genet* 53, 288–293. <https://doi.org/10.1038/s41588-020-00771-1>.
- 976 8. Jay, P., Tezenas, E., Véber, A., and Giraud, T. (2022). Sheltering of deleterious mutations explains  
977 the stepwise extension of recombination suppression on sex chromosomes and other supergenes.  
978 *PLOS Biology* 20, e3001698. <https://doi.org/10.1371/journal.pbio.3001698>
- 979 9. Nettancourt, D. (2001). Incompatibility and incongruity in wild and cultivated plants. (2nd ed.  
980 Springer-Verlag, Berlin Heidelberg).
- 981 10. Schopfer, C.R., Nasrallah, M.E., and Nasrallah, J.B. (1999). The male determinant of self-  
982 incompatibility in *Brassica*. *Science* 286, 1697–1700. <https://doi.org/10.1126/science.286.5445.1697>.
- 983 11. Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J.B., and Nasrallah, M.E. (2001).  
984 Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A.*  
985 *lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* 13, 627–643.

- 986 12. Uyenoyama, M.K. (2003). Genealogy-dependent variation in viability among self-incompatibility  
987 genotypes. *Theoretical Population Biology* 63, 281–293. [https://doi.org/10.1016/S0040-](https://doi.org/10.1016/S0040-5809(03)00020-0)  
988 [5809\(03\)00020-0](https://doi.org/10.1016/S0040-5809(03)00020-0).
- 989 13. Llaurens, V., Gonthier, L., and Billiard, S. (2009). The sheltered genetic load linked to the S locus  
990 in plants: new insights from theoretical and empirical approaches in sporophytic self-incompatibility.  
991 *Genetics* 183, 1105–1118. <https://doi.org/10.1534/genetics.109.102707>.
- 992 14. Castric, V., and Vekemans, X. (2004). Plant self-incompatibility in natural populations: a critical  
993 assessment of recent theoretical and empirical advances. *Molecular Ecology* 13, 2873–  
994 2889. <https://doi.org/10.1111/j.1365-294X.2004.02267.x>.
- 995 15. Llaurens, V., Billiard, S., Leducq, J.-B., Castric, V., Klein, E.K., and Vekemans, X. (2008). Does  
996 frequency-dependent selection with complex dominance interactions accurately predict allelic  
997 frequencies at the self-incompatibility locus in *Arabidopsis halleri*? *Evolution* 62, 2545–2557.  
998 <https://doi.org/10.1111/j.1558-5646.2008.00469.x>.
- 999 16. Durand, E., Méheust, R., Soucaze, M., Goubet, P.M., Gallina, S., Poux, C., Fobis-Loisy, I., Guillon,  
1000 E., Gaude, T., Sarazin, A., et al. (2014). Dominance hierarchy arising from the evolution of a complex  
1001 small RNA regulatory network. *Science* 346, 1200–1205. <https://doi.org/10.1111/eva.12933>.
- 1002 17. Billiard, S., Castric, V., and Vekemans, X. (2006). A general model to explore complex dominance  
1003 patterns in plant sporophytic self-incompatibility systems. *Genetics* 175, 1351–1369.  
1004 <https://doi.org/10.1016/j.tig.2011.06.005>.
- 1005 18. Goubet, P.M., Bergès, H., Bellec, A., Prat, E., Helmstetter, N., Mangenot, S., Gallina, S., Holl, A.-C.,  
1006 Fobis-Loisy, I., Vekemans, X., et al. (2012). Contrasted patterns of molecular evolution in dominant  
1007 and recessive self-incompatibility haplotypes in *Arabidopsis*. *PLoS Genetics* 8, e1002495.  
1008 <https://doi.org/10.1371/journal.pgen.1002495>.
- 1009 19. Stift, M., Hunter, B.D., Shaw, B., Adam, A., Hoebe, P.N., and Mable, B.K. (2013). Inbreeding  
1010 depression in self-incompatible North-American *Arabidopsis lyrata*: disentangling genomic and S-  
1011 locus-specific genetic load. *Heredity* 110, 19–28. <https://doi.org/10.1038/hdy.2012.49>.
- 1012 20. Le Veve, A., Burghgraeve, N., Genete, M., et al. (2023) Long-term balancing selection and the  
1013 genetic load linked to the self-incompatibility locus in *Arabidopsis halleri* and *A. lyrata*. *Molecular*  
1014 *Biology and Evolution* msad120. <https://doi.org/10.1093/molbev/msad120>.
- 1015 21. Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A.,  
1016 Grimwood, J., Gundlach, H., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of  
1017 rapid genome size change. *Nat Genet* 43, 476–481. <https://doi.org/10.1038/ng.807>.
- 1018 22. Foxe, J.P., Stift, M., Tedder, A., Haudry, A., Wright, S.I., and Mable, B.K. (2010). Reconstructing  
1019 origins of loss of self-incompatibility and selfing in North american *Arabidopsis lyrata*: a population  
1020 genetic context. *Evolution* 64, 3495–3510. <https://doi.org/10.1111/j.1558-5646.2010.01094.x>.

- 1021 23. Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome  
1022 regions. *PLOS Genetics* 2, e64. <https://doi.org/10.1371/journal.pgen.0020064>.
- 1023 24. Kamau, E., Charlesworth, B., Charlesworth, D. (2007) Linkage Disequilibrium and Recombination  
1024 Rate Estimates in the Self-Incompatibility Region of *Arabidopsis lyrata*. *Genetics* 176, 2357–2369.  
1025 <https://doi.org/10.1534/genetics.107.072231>
- 1026 25. Stone, J.L. (2004). Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* 92,  
1027 335–342. <https://doi.org/10.1038/sj.hdy.6800425>.
- 1028 26. Lane, M.D., and Lawrence, M.J. (1995). The population genetics of the self-incompatibility  
1029 polymorphism in *Papaver rhoeas*. X. An association between incompatibility genotype and seed  
1030 dormancy. *Heredity* 75, 92–97. <https://doi.org/10.1038/hdy.1995.108>.
- 1031 27. Vieira, J., Pimenta, J., Gomes, A., Laia, J., Rocha, S., Heitzler, P., and Vieira, C.P. (2021). The  
1032 identification of the *Rosa* S-locus and implications on the evolution of the Rosaceae gametophytic  
1033 self-incompatibility systems. *Sci Rep* 11, 3710. <https://doi.org/10.1038/s41598-021-83243-8>.
- 1034 28. Mena-Ali, J. I., Keser, L. H., Stephenson, A. G. (2009). The effect of sheltered load on  
1035 reproduction in *Solanum carolinense*, a species with variable self-incompatibility. *Sex. Plant Reprod.*  
1036 22 63–67. <https://doi.org/10.1007/s00497-008-0092-x>.
- 1037 29. Claus, M.J., and Mitchell-Olds, T. (2006). Population genetic structure of *Arabidopsis lyrata* in  
1038 Europe. *Molecular Ecology* 15, 2753–2766. Claus, M.J., Mitchell-Olds, T., 2006. Population genetic  
1039 structure of *Arabidopsis lyrata* in Europe. *Molecular Ecology* 15, 2753–2766.  
1040 <https://doi.org/10.1111/j.1365-294X.2006.02973.x>.
- 1041 30. Ross-Ibarra, J., Wright, S.I., Foxe, J.P., Kawabe, A., DeRose-Wilson, L., Gos, G., Charlesworth, D.,  
1042 and Gaut, B.S. (2008). Patterns of polymorphism and demographic history in natural populations of  
1043 *Arabidopsis lyrata*. *PLOS ONE* 3, e2411. <https://doi.org/10.1371/journal.pone.0002411>.
- 1044 31. Pauwels, M., Saumitou-Laprade, P., Holl, A.C., Petit, D., and Bonnin, I. (2005). Multiple origin of  
1045 metalicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central  
1046 Europe: the cpDNA testimony. *Molecular Ecology* 14, 4403–4414. <https://doi.org/10.1111/j.1365-294X.2005.02739.x>.
- 1048 32. Stetsenko, R., Brom, T., Castric, V., and Billiard, S. (2023). Balancing selection and the crossing of  
1049 fitness valleys in structured populations: diversification in the gametophytic self-incompatibility  
1050 system. *Evolution* 77:907–920. <https://doi.org/10.1093/evolut/qpac065>
- 1051 33. Gervais, C.E., Castric, V., Ressayre, A., and Billiard, S. (2011). Origin and diversification dynamics  
1052 of self-incompatibility haplotypes. *Genetics* 188, 625–636.  
1053 <https://doi.org/10.1534/genetics.111.127399>.
- 1054 34. Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex  
1055 chromosomes. *Trends in Ecology & Evolution* 24, 94–102.  
1056 <https://doi.org/10.1016/j.tree.2008.09.010>.



- 1057 35. Wright, A.E., Dean, R., Zimmer, F., and Mank, J.E. (2016). How to make a sex chromosome. *Nat*  
1058 *Commun* 7, 12087. <https://doi.org/10.1038/ncomms12087>.
- 1059 36. Ponnikas, S., Sigeman, H., Abbott, J.K., and Hansson, B. (2018). Why do sex chromosomes stop  
1060 recombining? *Trends Genet* 34, 492–503. <https://doi.org/10.1016/j.tig.2018.04.001>.
- 1061 37. Hill, J., Enbody, E., Bi, H., Lamichhaney, S., Schwochow, D., Younis, S., Widemo, F., and  
1062 Andersson, L. (2022). Low mutation load in a supergene underpinning alternative male mating  
1063 strategies in ruff. <https://doi.org/10.1101/2022.04.27.489720>.
- 1064 38. Rosser, N., Edelman, N.B., Queste, L.M., et al. (2022) Complex basis of hybrid female sterility and  
1065 Haldane’s rule in *Heliconius* butterflies: Z-linkage and epistasis. *Molecular Ecology* 31:959–977.  
1066 <https://doi.org/10.1111/mec.16272>.
- 1067 39. Durand, E., Chantreau, M., Le Veve, A., et al. (2020). Evolution of self-incompatibility in the  
1068 Brassicaceae: Lessons from a textbook example of natural selection. *Evolutionary Applications*.  
1069 <https://doi.org/10.1111/eva.12933>.
- 1070 40. Llaurens, V., Billiard, S., Castric, V., and Vekemans, X. (2009). Evolution of dominance in  
1071 sporophytic self-incompatibility systems: I. Genetic load and coevolution of levels of dominance in  
1072 pollen and pistil. *Evolution* 63, 2427–2437. <https://doi.org/10.1111/j.1558-5646.2009.00709.x>.
- 1073 41. Genete, M., Castric, V., and Vekemans, X. (2020). Genotyping and *de novo* discovery of allelic  
1074 variants at the Brassicaceae self-incompatibility locus from short read sequencing data. *Mol Biol*  
1075 *Evol.* <https://doi.org/10.1093/molbev/msz258>.
- 1076 42. Schierup, M.H., Mikkelsen, A.M., and Hein, J. (2001). Recombination, balancing selection and  
1077 phylogenies in *MHC* and self-incompatibility genes. *Genetics* 159, 1833–1844.  
1078 <https://doi.org/10.1093/genetics/159.4.1833>
- 1079 43. Mable, B.K., Schierup, M.H., and Charlesworth, D. (2003). Estimating the number, frequency, and  
1080 dominance of S -alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic  
1081 control of self-incompatibility. *Heredity* 90, 422–431. <https://doi.org/10.1038/sj.hdy.6800261>.
- 1082 44. Bechsgaard, J., Bataillon, T., and Schierup, M.H. (2004). Uneven segregation of sporophytic self-  
1083 incompatibility alleles in *Arabidopsis lyrata*. *Journal of Evolutionary Biology* 17, 554–561.  
1084 <https://doi.org/10.1111/j.1420-9101.2004.00699.x>.
- 1085 45. Prigoda, N.L., Nassuth, A., and Mable, B.K. (2005). Phenotypic and genotypic expression of self-  
1086 incompatibility haplotypes in *Arabidopsis lyrata* suggests unique origin of alleles in different  
1087 dominance classes. *Molecular Biology and Evolution* 22, 1609–1620.  
1088 <https://doi.org/10.1093/molbev/msi153>.
- 1089 46. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*  
1090 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.

- 1091 47. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin,  
1092 R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map  
1093 format and SAMtools. *Bioinformatics* 25, 2078–2079.  
1094 <https://doi.org/10.1093/bioinformatics/btp352>.
- 1095 48. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic  
1096 features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- 1097 49. DePristo, M.A., Banks, E., Poplin, R.E., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A.,  
1098 del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and  
1099 genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491–498.  
1100 <https://doi.org/10.1038/ng.806>.
- 1101 50. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E.,  
1102 Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools.  
1103 *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- 1104 51. Williamson, R.J., Josephs, E.B., Platts, A.E., Hazzouri, K.M., Haudry, A., Blanchette, M., and  
1105 Wright, S.I. (2014). Evidence for widespread positive and negative selection in coding and conserved  
1106 noncoding regions of *Capsella grandiflora*. *PLOS Genetics* 10, e1004622.  
1107 <https://doi.org/10.1371/journal.pgen.1004622>.
- 1108 52. Guindon, S., Dufayard, J-F., Lefort, V., et al (2010) New algorithms and methods to estimate  
1109 maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59, 307–321.  
1110 <https://doi.org/10.1093/sysbio/syq010>.
- 1111 53. Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the  
1112 control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*  
1113 10, 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>.
- 1114 54. Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular evolutionary  
1115 genetics analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.  
1116 <https://doi.org/10.1093/molbev/msy096>.
- 1117 55. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E.,  
1118 and Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets.  
1119 *Molecular Biology and Evolution* 34, 3299–3302. <https://doi.org/10.1093/molbev/msx248>.
- 1120 56. Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in  
1121 structured populations. *The American Journal of Human Genetics* 67:170–181.  
1122 <https://doi.org/10.1086/302959>.
- 1123 57. Pritchard, J. K., Wen, X., & Falush, D. (2010). Documentation for structure software: Version 2.3.  
1124 University of Chicago, Chicago, IL, 1-37.
- 1125 58. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image  
1126 analysis. *Nat Methods* 9, 671–675. <https://doi.org/10.1038/nmeth.2089>.