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

Main text

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

Le Veve Audrey[1,2], Genete Mathieu[1], Lepers-Blassiau Christelle[1], Ponitzki Chloé[1], Céline Poux[1], Vekemans Xavier[1], Durand Eleonore[1], Castric Vincent[1*]

[1]*Univ. Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France*

[2]*Current address: Department of Botany, Faculty of Science, Charles University, Benátská 2, CZ-128 01 Prague, Czechia*

* *Correspondence : vincent.castric@univ-lille.fr*

Abstract

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

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

Introduction

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

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

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

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

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

Results

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

We first expanded the experimental approach of Llaurens et al.[13] to phenotypically evaluate the effect of *S*-allele dominance on the intensity of the sheltered load. The previous study focused on three *S*-alleles (Ah01, Ah02 and Ah15)[13]. Here we included two *S*-alleles from the same local population (Nivelle, France): Ah03 and Ah04, and included Ah01 again for comparative purposes. In the Arabidopsis genus, *S*-alleles have been shown to form a complex dominance hierarchy[15,16]. This hierarchy is largely associated with the phylogeny of *S*-alleles[15], and at least four phylogenetic classes (I, II, III and IV) have been described, from the most recessive (class I) to the most dominant of *S*-alleles (class IV). Dominance interactions also exist among *S*-alleles within classes, such that these five *S*-alleles form the following dominance hierarchy[15,16]: Ah01<Ah03<Ah02<Ah04<Ah15, from the most recessive (Ah01) to the most dominant (Ah15). To reveal the linked load, we enforced homozygosity at the *S*-locus using controlled crosses between parental individuals sharing a given *S*-allele that was masked by different dominant *S*-alleles (*e.g.*, to obtain Ah_xAh_x homozygotes we deposited pollen from a Ah_xAh_y plant, where $Ah_y>Ah_x$, on pistils of a Ah_xAh_z plant where $z\neq y$, or on a Ah_xAh_x pistil when available; table S1). We obtained 399 offspring from a total of six such crosses. Note that our experimental procedure differs slightly from that of Llaurens et al.[13] in that their procedure required a CO_2 treatment to bypass the SI system and obtain selfed offspring, while here we took advantage of the dominance interactions to obtain outcrossed *S*-locus homozygous

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

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

Table 1: Proportion of S-locus homozygous offspring having reached the reproductive stage for three different S-alleles. *The test is performed relative to the expected proportion of homozygous genotypes in the offspring (25% when both parents are heterozygous; 50% when one of the parents 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[a])	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	**0.074**	**0.3 (0.04; 0.02)**	17 (1; 0.69)
Ah04	III	96	0.479	0.96 (0.75; 0.39)	50

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

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

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

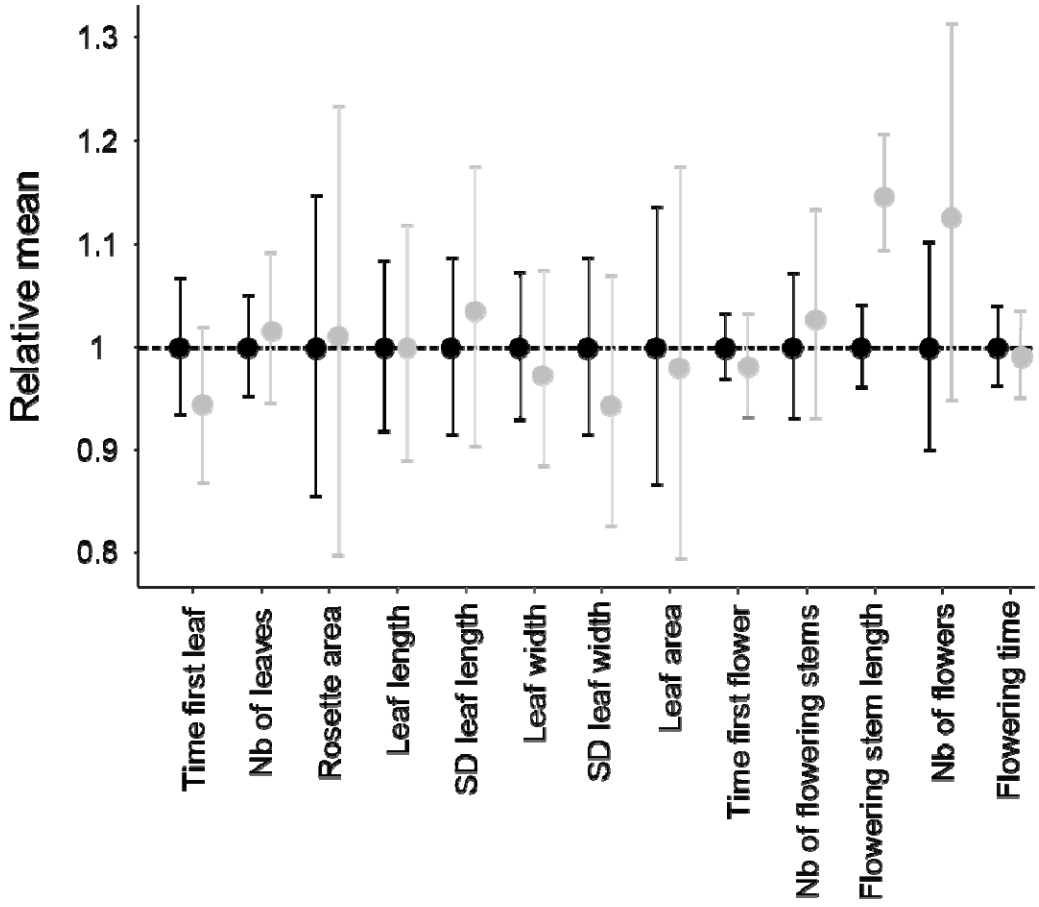


Figure 1: Effect of homozygosity at the S-locus on 13 phenotypic traits compared to heterozygotes. *For each trait, the phenotypic values in homozygotes (in grey) were normalised relative to the mean phenotypic values in heterozygotes (in black). The distributions were obtained by 10,000 random permutations.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

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

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

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

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

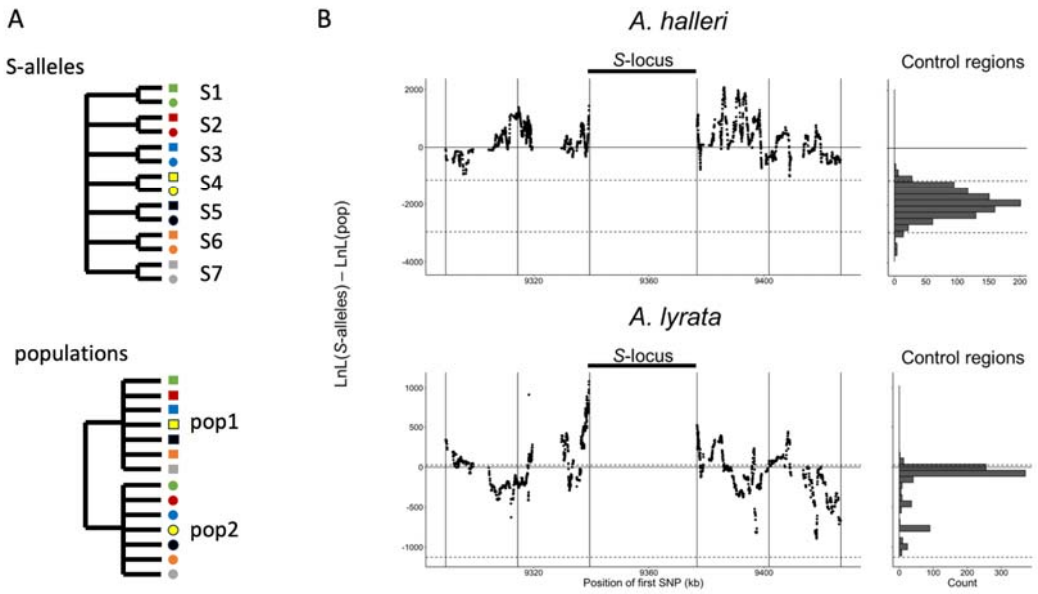


Figure 2: Linkage to the S-locus locally distorts the phylogenetic relationships. *A. The two topologies of interest cluster haplotypes either by the S-allele to which they are linked (top) or by the populations where they came (bottom). Different S-alleles are represented by symbols of different colours,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

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

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

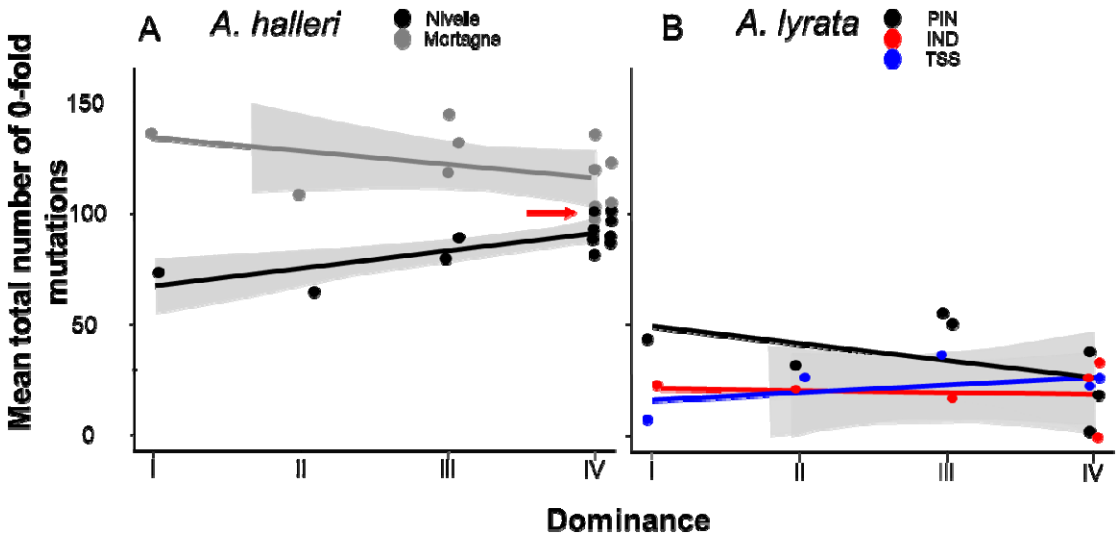


Figure 3: No overall effect of S-allele dominance on the total number of 0-fold degenerate mutations (S_{0f}) in the linked genomic regions within 25kb. *Each dot represents the mean number of mutations observed among haplotypes linked to one S-allele in one population. The correlations evaluated by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

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

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

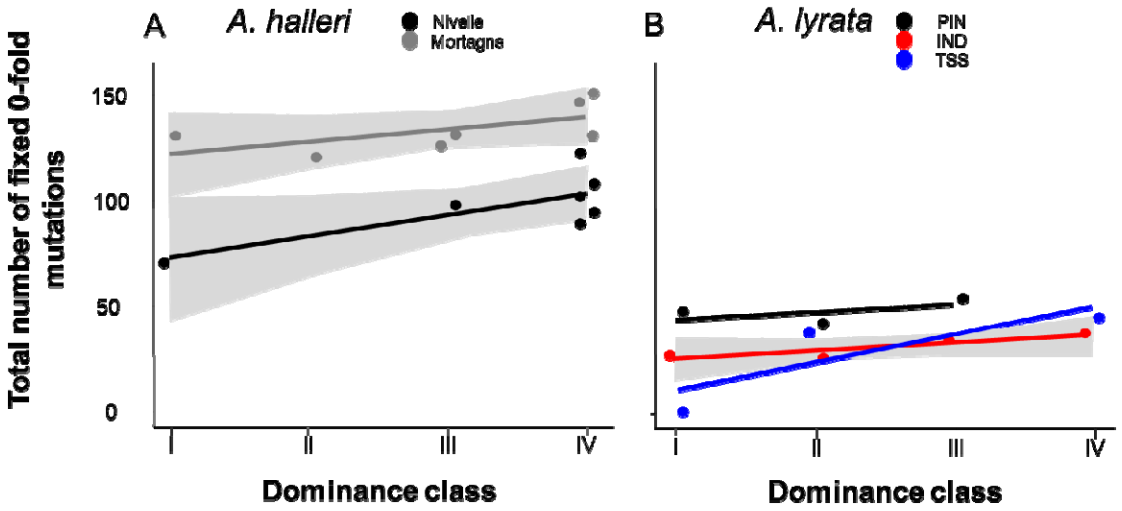


Figure 4: The number of 0-fold degenerate mutations fixed in the 25kb regions flanking the S-locus increases with dominance of the S-allele associated. *Each dot represents the value obtained for haplotypes linked to one S-allele in one population. The correlations evaluated by a GLM are represented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

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

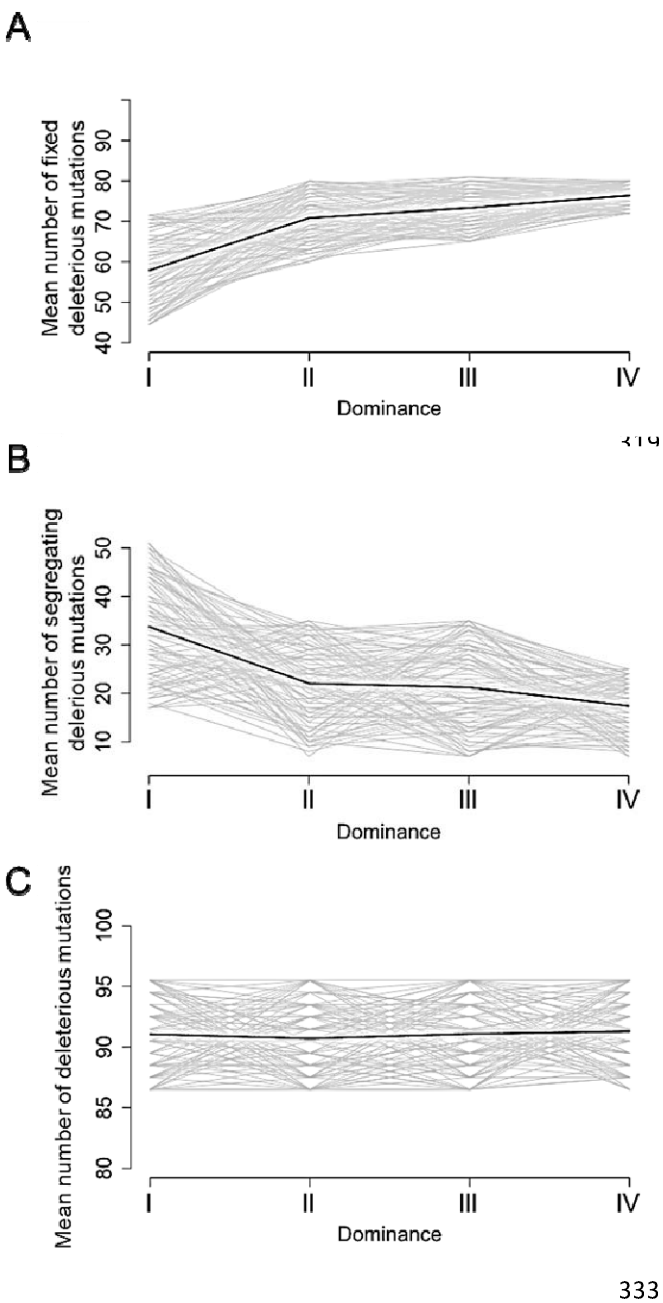


Figure 5: Stochastic simulations confirm the contrasted architecture of the load of deleterious mutations linked to dominant vs. recessive S-alleles. *Number of fixed (A), segregating (B) and total (C) deleterious mutations linked to S-alleles at four different levels of dominance (I<II<III<IV). The 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.*

Discussion

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

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

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

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

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

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

The question of whether *S*-allele dominance, which modifies the intensity of balancing selection in the *S*-locus region, could explain variation of the linked load has received conflicting support in the literature. In line with Stift et al.[19], but in contradiction with Llaurens et al.[13], we observed no overall effect of dominance on the magnitude of the load. Several technical and biological reasons could explain the contrasted results obtained in these different studies. First, phenotypic quantification of the linked load is experimentally demanding, such that these studies relied on the comparison of a limited number of alleles (three *S*-alleles in each of the studies) and therefore each of them had inherently low power. Second, the experimental procedures to reveal the load varied slightly. Llaurens et al.[13] used CO_2 treatment to by-pass the SI system and obtain homozygous progenies from crosses that would otherwise have been incompatible, whereas we used the "natural" masking by dominant *S*-alleles to enable the obtention of recessive homozygous genotypes. Our approach is experimentally simpler and avoids the possible contamination by offspring obtained by selfing, which may combine the effect of the sheltered load with that of genome-wide inbreeding depression (see Stift et al.[19] for a detailed discussion of this caveat). Third, a limitation of our approach is that it is restricted to *S*-alleles that are recessive or intermediate along the dominance hierarchy, and is thus not applicable to quantify the load associated with the most dominant *S*-alleles under more intense balancing selection. It is therefore possible that the *S*-alleles we examined did not exhibit sufficiently contrasted levels of dominance, in particular if only the most dominant ones are generating a substantial load, as suggested for fully linked deleterious mutations. In addition, since the homozygous *S*-allele genotypes we create correspond to different gene copies from the population, they may carry distinct sets of linked deleterious variants, especially for the more recessive alleles. The variation we observed in the phenotypic magnitude of the load among families confirms that linked deleterious variants are unlikely to be fixed within all allele lineages. Finally, we note that our genomic analysis of the genetic load shows that the dominant allele Ah15 previously associated with reduced fitness in homozygotes[13], is indeed unusual in terms of the number of mutations it carries. In fact, it is one of the most "loaded" alleles among all the dominant *S*-alleles present in this population, possibly explaining why Llaurens et al.[13] observed a significant effect despite the inherently limited experimental power of their analysis.

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

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

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

It is clear from our results that *S*-allele dominance can affect the linked load, through its effect on the intensity of balancing selection, but in turn the differences in structure of the linked load may affect the conditions under which dominance can evolve. The Brassicaceae *S*-locus is indeed an interesting and rather unique system, where the mechanisms by which dominance is controlled and evolves under the action of so-called "dominance modifiers" have been studied in detail[39,40] (involving the interaction between small non-coding RNAs and their target sites[16, 40]). The presence 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[40]. 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

Figures legend

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

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

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

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

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.*

Tables List

Table 1: Proportion of S-locus homozygous offspring having reached the reproductive stage for three different S-alleles. *The test is performed relative to the expected proportion of homozygous genotypes in the offspring (25% when both parents are heterozygous; 50% when one of the parents 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[a])	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	**0.074**	**0.3 (0.04; 0.02)**	17 (1; 0.69)
Ah04	III	96	0.479	0.96 (0.75; 0.39)	50

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

Methods STAR

Source plant material

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

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

Library preparation, capture and sequencing

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

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

Determination of the S-locus genotypes and dominance of S-alleles

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

Read mapping and variant calling in A. halleri and A. lyrata populations

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

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

Quantifying the sheltered load of deleterious mutations

We examined the genetic load signatures based on the accumulation of mutations on 0-fold degenerate sites, the vast majority of which are considered deleterious. The 0-fold and 4-fold degenerate sites were identified and extracted from the reference genome and the gene annotation using the script NewAnnotateRef.py[51]. We also examined the genetic load signatures based on the accumulation of all non-synonymous mutations. We obtained a total of 2,441 and 2,435 variable positions for the *A. halleri* samples from Nivelle and Mortagne respectively, and 2,360 variable positions for the *A. lyrata* samples.

Phasing S-haplotypes

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

Study of the structure of S-haplotypes

We first developed a new method to evaluate the distortion of the phylogenetic patterns caused by linkage to *S*-alleles. To do this, we used phyml[52] v.3.3 to calculate the likelihood of two contrasted topologies of interest: (1) the topology clustering haplotypes by the populations where they came from vs. (2) the topology clustering them by the *S*-allele to which they are linked (Fig 2A). We used sliding windows of sequences with 50 SNPs to obtain the variation of the difference in log-likelihood 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[53], with 1,000 replicates in MEGA
642 X[54]. Third, we followed Kamau et al.'s[24] 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[55]. 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[56] V1.1 combined with Structure[57] 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[16] (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.[13] based their comparisons on outcrossed progenies obtained by enforced
675 incompatible crosses and Stift et al.[19] 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

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

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

Simulations

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

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

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

Data Availability

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

All scripts developed are available in Github (https://github.com/leveveaudrey/dominance_and_sheltered_load https://github.com/leveveaudrey/analysis-of-polymorphism-S-locus).

Supplementary data

Figures legend

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

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

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

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

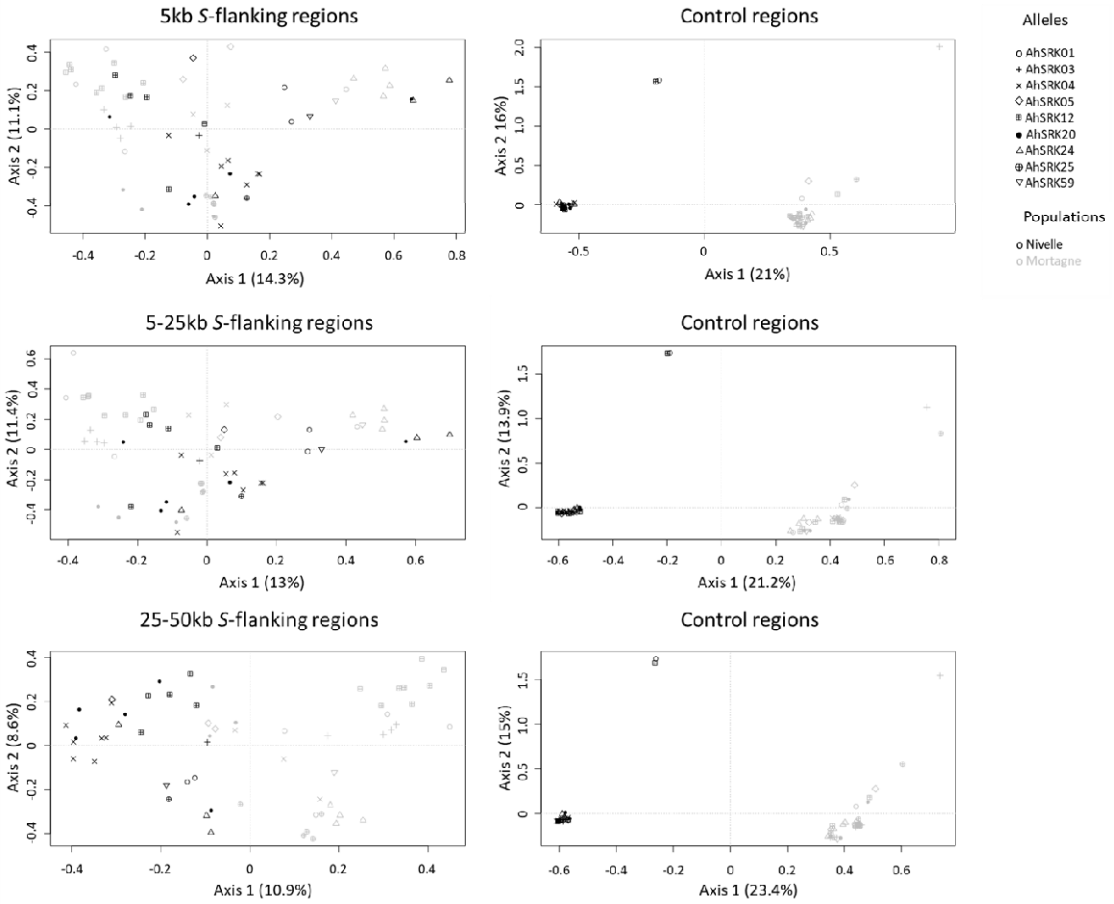
Figure S5: Phylogenetic tree obtained by maximum likelihood for haplotypes of A. lyrata (populations PIN, IND, TSS) across the first 5kb flanking the S-locus. *The Tamura-Nei model was used and the percentage of trees in which the associated haplotypes clustered together is shown next to the branches.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

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

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

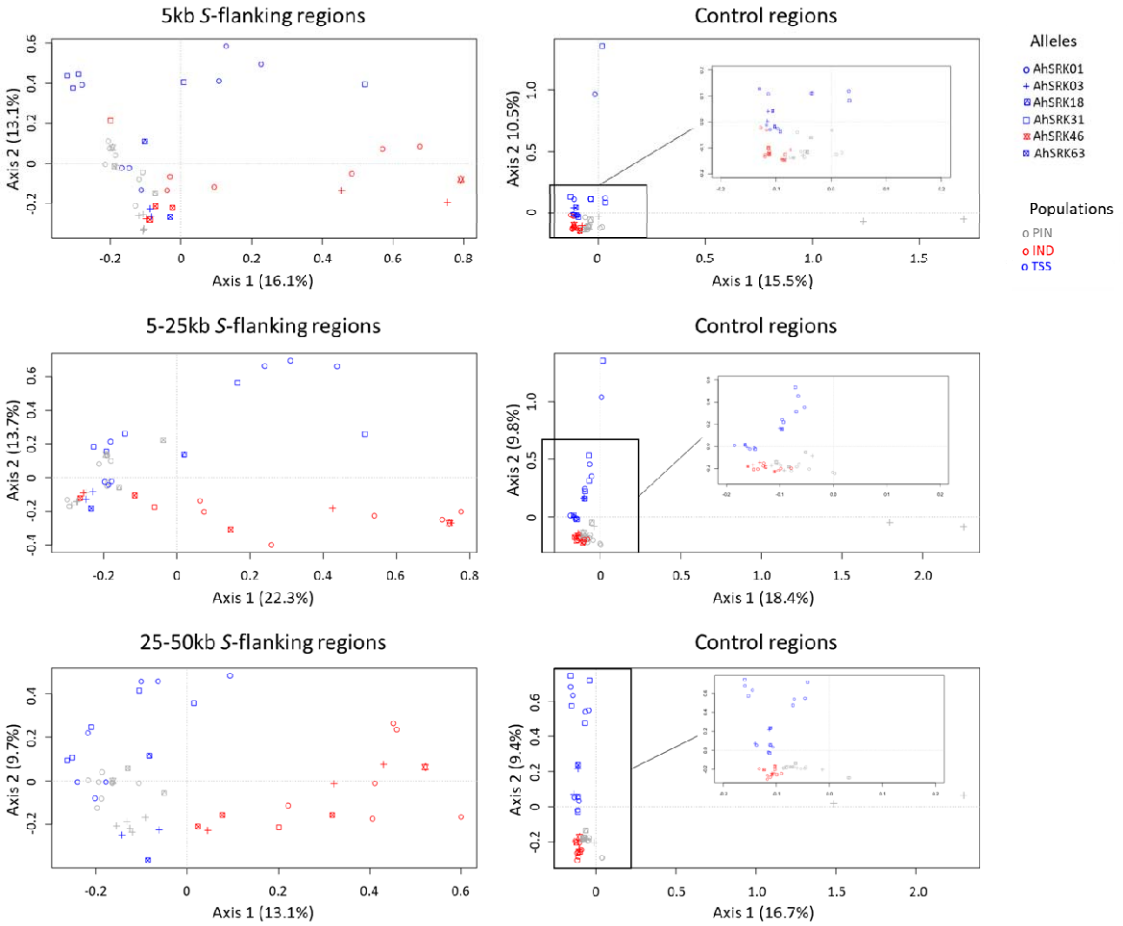
Figure S8: Patterns of genetic associations between S-alleles and SNPs across the genome. *Each dot corresponds to a SNP showing statistically significant association (top 0.1%) with a given S-allele. The left panel confirms that SNPs physically linked to the S-locus (in red) are considerably more likely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

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



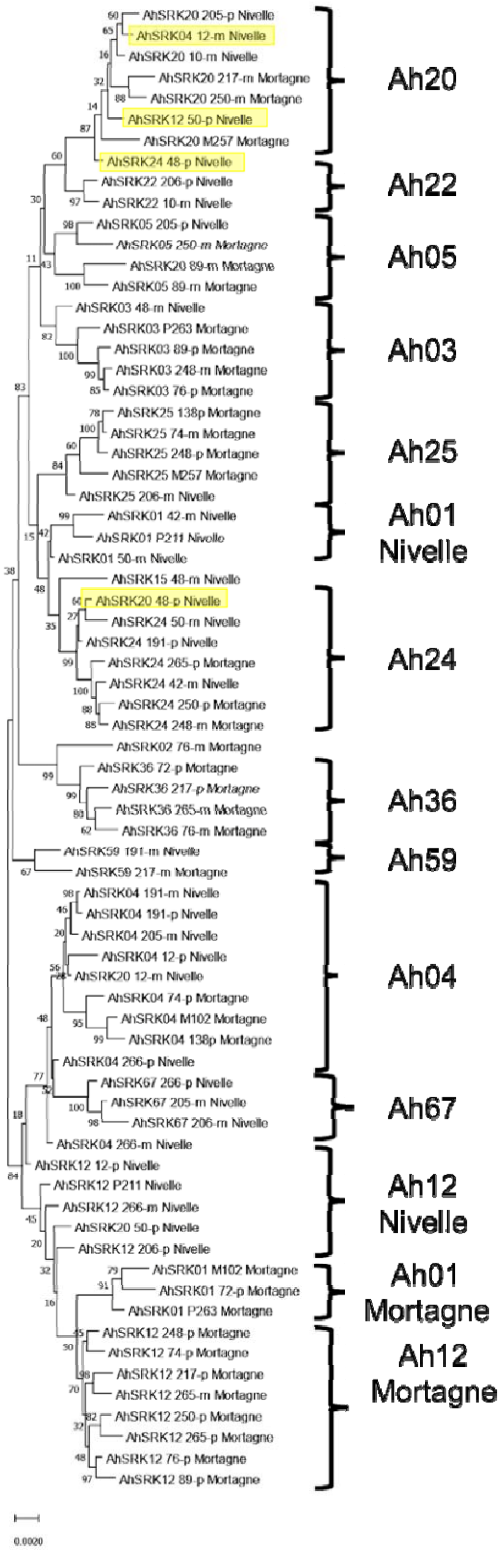
832

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

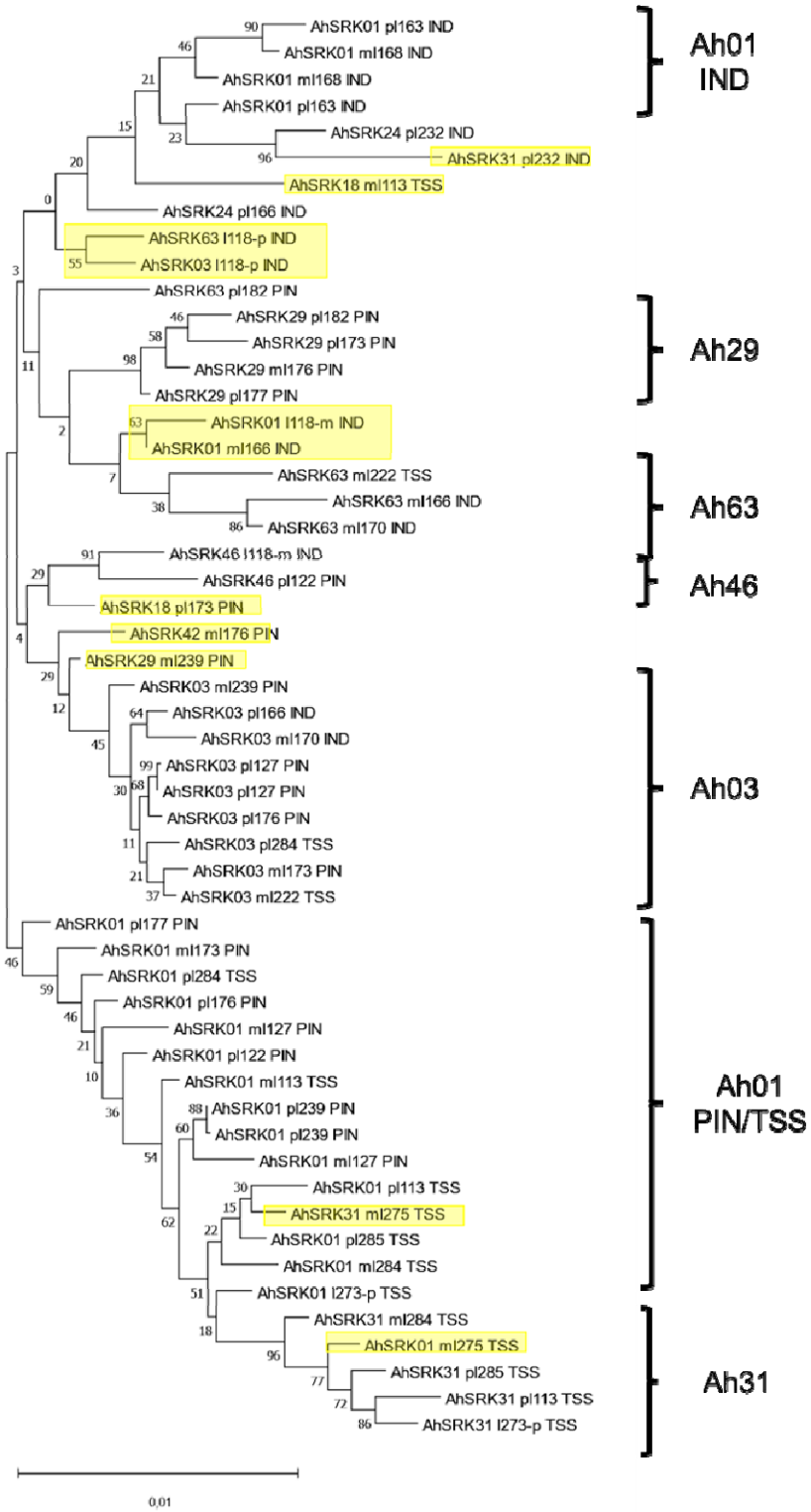
837 ***Figure S2: Analysis of major components (AMC) obtained for haplotypes of A. lyrata (of the PIN:***
838 ***grey dots, IND: red dots and TSS: blue dots) populations based on the SNPs in the first 5kb,***
839 ***between 5 and 25kb, and between 25kb and 50kb away from the S-locus.***



840

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

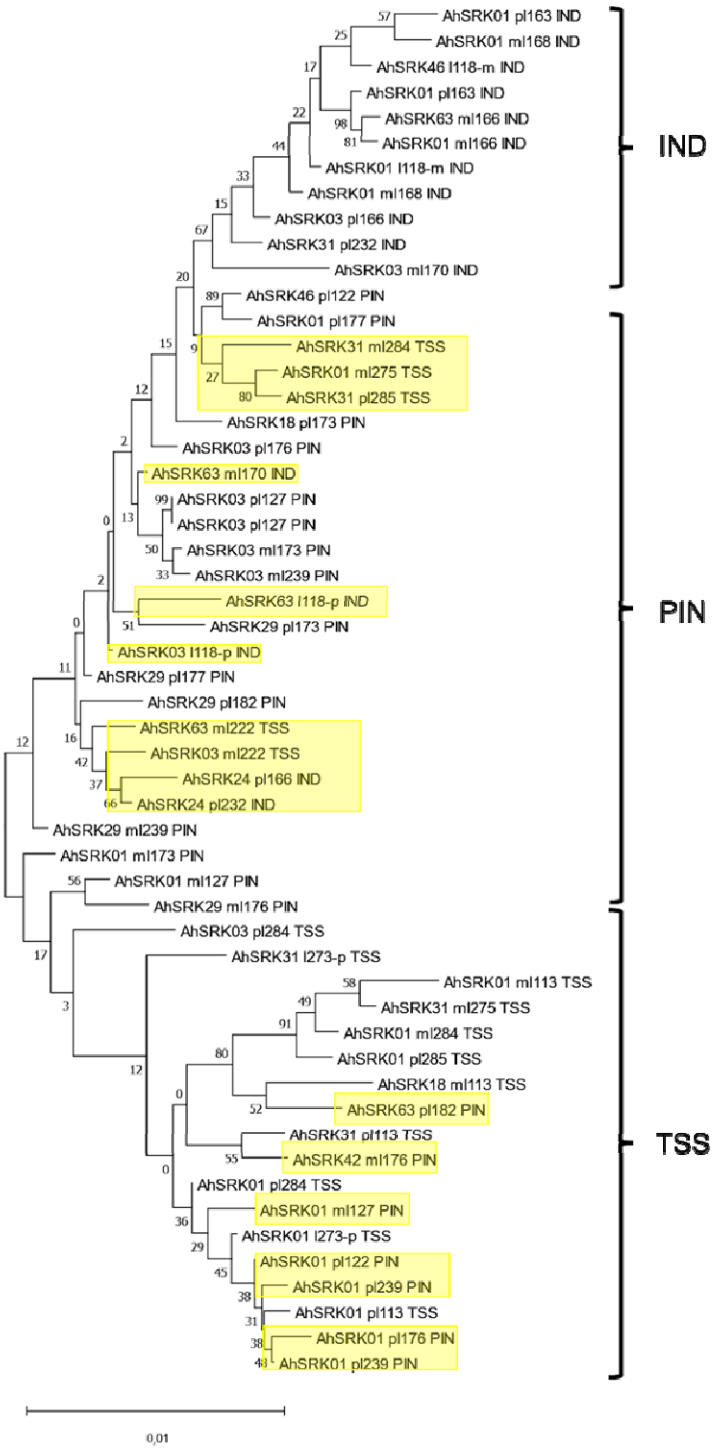
Figure S4: Phylogenetic tree obtained by maximum likelihood for haplotypes of A. halleri (populations Nivelle and Mortagne) based on the nucleotide positions between 25kb and 50kb away from the S-locus.



847

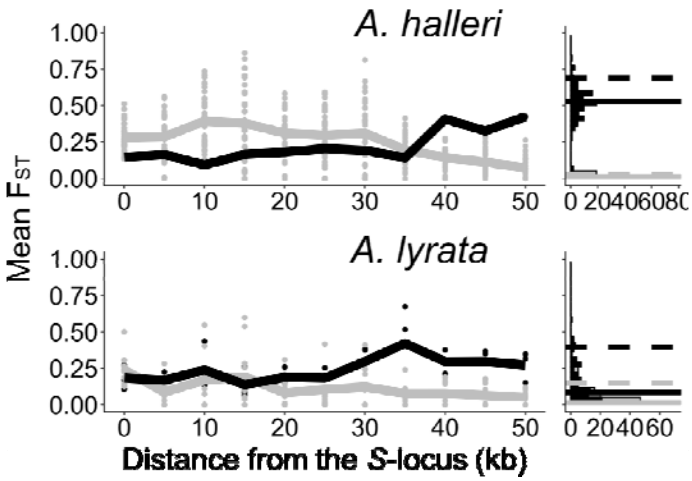
848 ***Figure S5: Phylogenetic tree obtained by maximum likelihood for haplotypes of A. lyrata***
849 ***(populations PIN, IND, TSS) across the first 5kb flanking the S-locus.***

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.***

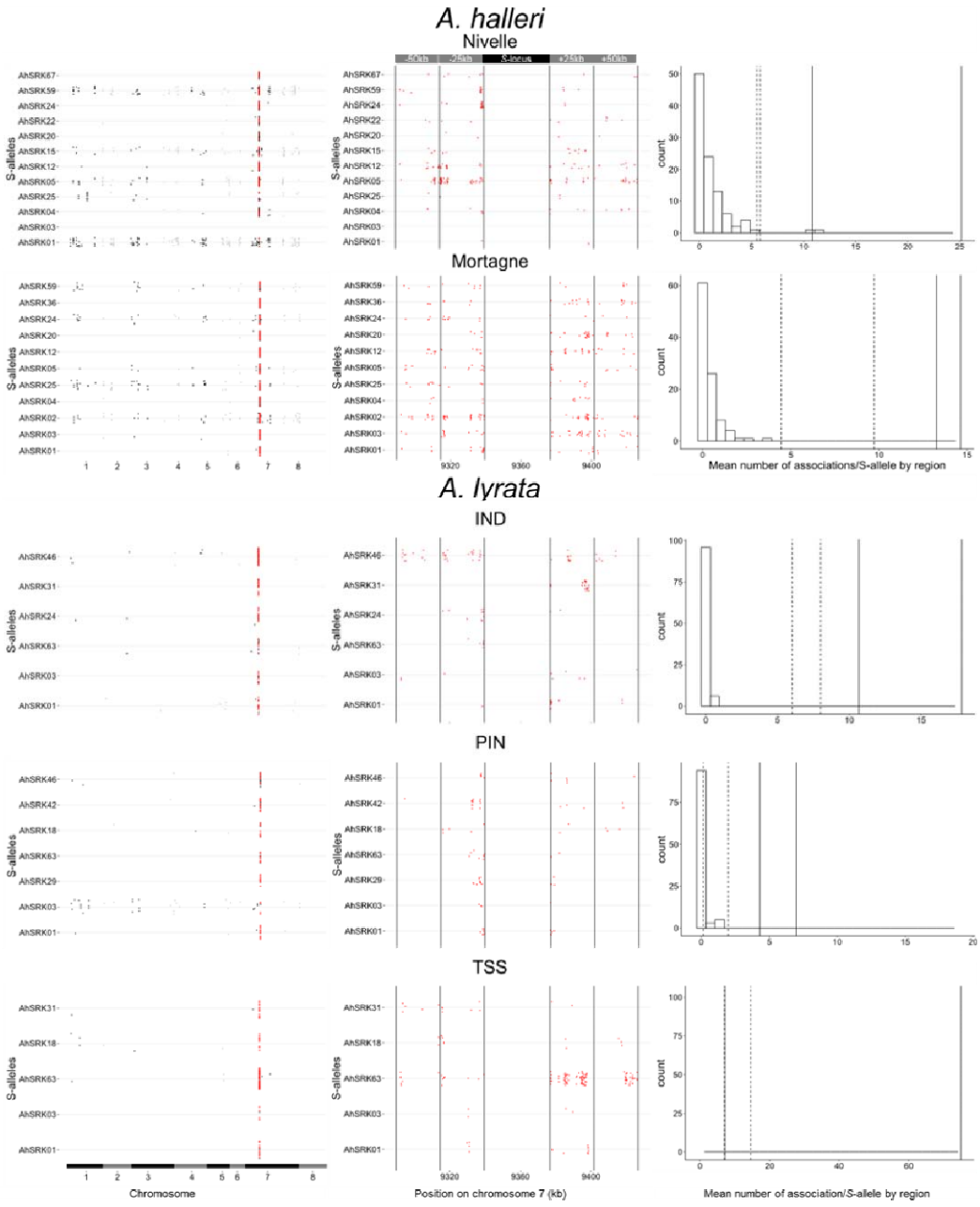
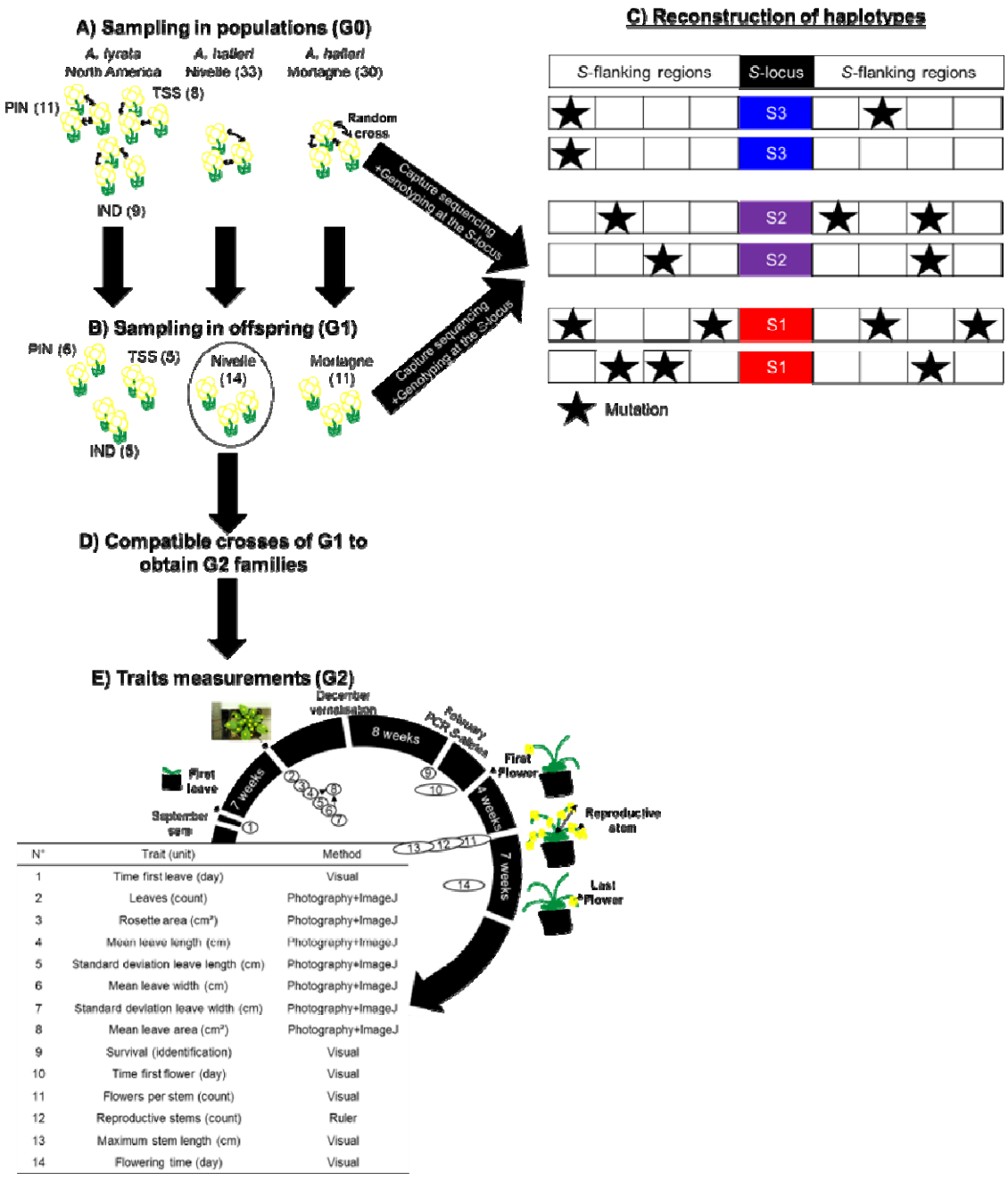


Figure S8: Patterns of genetic associations between S-alleles and SNPs across the genome.



N°	Trait (unit)	Method
1	Time first leave (day)	Visual
2	Leaves (count)	Photography+ImageJ
3	Rosette area (cm²)	Photography+ImageJ
4	Mean leave length (cm)	Photography+ImageJ
5	Standard deviation leave length (cm)	Photography+ImageJ
6	Mean leave width (cm)	Photography+ImageJ
7	Standard deviation leave width (cm)	Photography+ImageJ
8	Mean leave area (cm²)	Photography+ImageJ
9	Survival (iddentification)	Visual
10	Time first flower (day)	Visual
11	Flowers per stem (count)	Visual
12	Reproductive stems (count)	Ruler
13	Maximum stem length (cm)	Visual
14	Flowering time (day)	Visual

Figure S9: Experimental protocol.

Supplemental items

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

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

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

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

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

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

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

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

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

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

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).*

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

888 [a] *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

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

Trait (unit)	*Mean heterozygotes (SD)*	*Mean homozygotes (SD)*	*P value*[a]
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 (cm²)	*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 (cm²)	*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*
Maximum flowering stem length (cm)	***56.19 (2.26)***	***64.55 (3.17)***	***$<1e^{-04}$***
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*

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

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

Allele	*Trait (unit)*	*Mean first familly (SD)*	*Mean second familly (SD)*	*p value*
Ah01[b]	***Time first leaf (day)***	***17.11 (2.26)***	***32 (NA)***	***<1.00^{e-04}***
	Number of leaves	*14 (2.83)*	*9 (NA)*	*0.10*
	Rosette area (cm²)	*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 (cm²)	*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*
	Maximum flowering stem length (cm)	***43.43 (6.91)***	***62.4 (NA)***	***<1.00^{e-04}***
	Number of flowers by flowering stem	*31.73 (24.16)*	*52.25 (NA)*	*0.12*
	Flowering duration (day)	***44.71 (6.73)***	***52 (NA)***	***<1.00^{e-04}***
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 (cm²)	*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*
	St dev leaf width (cm)	***0.28 (0.13)***	***0.22 (0.09)***	***0.02***
	Mean leaf area (cm²)	*3.4 (2.45)*	*3.46 (2.27)*	*0.45*
	Time first flower (day)	*35.57 (4.73)*	*33.2 (3.66)*	*0.06*
	Number of flowering stems	***13.21 (4.25)***	***16.4 (5.25)***	***0.02***
	Maximum flowering stem length (cm)	*64.98 (9.65)*	*62.41 (9.34)*	*0.13*
	Number of flowers by flowering stem	***72.38 (33.07)***	***45.25 (15.45)***	***1.00^{e-04}***
	Flowering duration (day)	*43.68 (6.96)*	*41.56 (7.38)*	*0.23*

904 [a] *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.* [b] *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.***

Trait (unit)	*Random effect implemented in GLM model*	*Distribution of trait tested*	*Linear effect*	*p value*
Time first leaf (days)	***Family***	***Poisson***	***-0.15***	***5.5^{e-04}***
Leaves (counts)	*Oxydative stress*	*Poisson*	*0.006*	*0.91*
Rosette area (cm²)	*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 (cm²)	*Oxydative stress + phytophagous attack*	*Gaussian*	*0.01*	*0.91*
Mean leave area (cm²)	*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 [a]
A. halleri	Nivelle	34	12	7	4.1
	Mortagne	38	11	9	4
A. lyrata	TSS	16	5	3	4.7
	IND	16	6	4	3.5
	PIN	22	7	3	6

915 [a] *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***	***2.57e-04***
		Number of segregated mutations by S-allele	***-0.49***	***2.21e-03***
		Mean total number of mutation by S-allele (S_{0f})	*0.01*	*0.54*
	A. lyrata	***Number of fixed mutations by S-allele***	***0.08***	***4.50e-03***
		Number of segregated mutations by S-allele	*-0.44*	*0.09*
		Mean total number of mutation by S-allele (S_{0f})	*-0.07*	*0.07*
4fold	*A. halleri*	***Number of fixed mutations by S-allele***	***0.23***	***8.95e-04***
		Number of segregated mutations by S-allele	***-0.48***	***2.02e-10***
		Mean total number of mutation by S-allele (S_{4f})	*1.53e-03*	*0.95*
	A. lyrata	***Number of fixed mutations by S-allele***	***0.23***	***8.95e-04***
		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_{0f}/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***	***1.86e-04***
		Number of segregated mutations by S-allele	***-0.23***	***<2e-16***
		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***	***1.74e-08***
		Number of segregated mutations by S-allele	***-0.28***	***<2e-16***
		Mean total number of mutation by S-allele (S_{NS})	*7.99e-03*	*0.65*
Synonymous	*A. halleri*	***Number of fixed mutations by S-allele***	***0.08***	***4.55e-04***
		Number of segregated mutations by S-allele	***-0.26***	***<2e-16***
		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***	***2.3e-11***
		Number of segregated mutations by S-allele	***-0.34***	***<2e-16***
		Mean total number of mutation by S-allele (S_S)	*-0.01*	*0.33*
S_{NS}/S_S	*A. halleri*	*Ratio*	*-5.82e-03*	*0.81*
	A. lyrata	*Ratio*	*0.02*	*0.60*

Dominance at the S-locus was implemented as a fixed effect. Populations were implemented as random effects. The distribution of the genetic load 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.***

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

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

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

[a] *The individuals were mainly sequenced by capture approach. The genotypes of homozygotes were confirmed after a whole genome sequencing. The class of dominance is specified in parenthesis.*

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

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

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

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

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

Effect tested	*Trait*	*Mean with attack (sd)*	*Mean without attack (sd)*	*p value* [a]
	Time first leaf (day)	*15.37 (4.84)*	*15.38 (5.24)*	*0.49*
	Number of leaves	*13.03 (2.9)*	*13.46 (3.71)*	*0.23*
	Rosette area (cm²)	*31.93 (27.19)*	*37.08 (27.14)*	*0.13*
	Mean leave length (cm)	*1.78 (0.78)*	*1.91 (0.8)*	*0.16*
	St dev leave length (cm)	*0.29 (0.14)*	*0.28 (0.13)*	*0.38*
	Mean leave width (cm)	*1.42 (0.56)*	*1.51 (0.53)*	*0.17*
	St dev leave width (cm)	*0.25 (0.11)*	*0.27 (0.12)*	*0.23*
	Mean leave area (cm²)	*2.94 (2.33)*	*3.25 (2.15)*	*0.2*
	Time first flower (day)	***37.43 (4.56)***	***31.34 (5.55)***	***2e-03***
	Number of flowering stems	*12.05 (4.4)*	*11.84 (4.62)*	*0.05*
	Maximum flowering stem length (cm)	*56.78 (12.04)*	*52.55 (11.82)*	*0.1*
	Number of flowers by flowering stem	*56.56 (30.99)*	*56.22 (34.55)*	*0.11*
Phytopathogen	***Flowering duration (day)***	***41.44 (8.13)***	***39.51 (7.32)***	***9e-03***
	Time first leaf (day)	*14.48 (4.4)*	*15.56 (5.2)*	*0.16*
	Number of leaves	*12.59 (3.76)*	*13.44 (3.35)*	*0.13*
	Rosette area (cm²)	*32.15 (32.24)*	*35.77 (26.16)*	*0.27*
	Mean leave length (cm)	*1.68 (0.83)*	*1.9 (0.78)*	*0.1*
	St dev leave length (cm)	*0.29 (0.1)*	*0.28 (0.14)*	*0.36*
	Mean leave width (cm)	***1.29 (0.55)***	***1.51 (0.53)***	***0.03***
	St dev leave width (cm)	***0.22 (0.1)***	***0.27 (0.12)***	***0.03***
	Mean leave area (cm²)	*2.59 (2.3)*	*3.25 (2.19)*	*0.08*
	Time first flower (day)	***34.78 (6.93)***	***33.36 (4.92)***	***0.04***
Phytophagous	*Number of flowering stems*	*10.78 (4.76)*	*12.15 (4.55)*	*0.34*

	Maximum flowering stem length (cm)	***53.53 (9.81)***	***54.24 (12.2)***	***0.05***
	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*
	Number of leaves	***10.04 (3.88)***	***13.93 (2.96)***	***1e-^{04}***
	Rosette area (cm²)	***11.79 (16.54)***	***39.64 (26.56)***	***1e-^{04}***
	Mean leave length (cm)	***1.05 (0.49)***	***2.02 (0.74)***	***1e-^{04}***
	St dev leave length (cm)	*0.25 (0.1)*	*0.29 (0.14)*	*0.09*
	Mean leave width (cm)	***0.88 (0.31)***	***1.59 (0.5)***	***1e-^{04}***
	St dev leave width (cm)	***0.23 (0.12)***	***0.27 (0.12)***	***0.04***
	Mean leave area (cm²)	***1.05 (0.98)***	***3.54 (2.17)***	***1e-^{04}***
	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*

[a] *The P value represents the proportions of the distribution with the expected difference of mean between individuals with and without attack after 10,000 random resamples equal to or less than the value observed. Significant values are represented in bold.*

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

Allele	*Family (Male/Female)*	*Trait (unit)*	*Mean (sd)*	*p value*[a]
Ah01	*1 (d29.1/d24.1)*	***Time first leaf (day)***	***20.8 (5.43)***	***$2e^{-03}$***
		Number of leaves	*11.9 (3.75)*	*0.09*
		Rosette area (cm²)	*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 (cm²)	*2.47 (1.93)*	*0.17*
		Time first flower (day)	*28.8 (5.29)*	*0.44*
		Number of flowering stems	***8 (4.07)***	***0.02***
		Maximum flowering stem length (cm)	***40.65 (6.76)***	***0.02***
		Number of flowers by flowering stem	***35.85 (10.72)***	***0.04***
		Flowering duration (day)	***40.35 (12.45)***	***$8e^{-03}$***
Ah01	*2 (d12.1/d29.1)*	***Time first leaf (day)***	***18.76 (5.57)***	***$1e^{-04}$***
		Number of leafs	*13.45 (3.22)*	*0.42*
		Rosette area (cm²)	*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 (cm²)	*2.84 (1.89)*	*0.21*
		Time first flower (day)	***29.86 (5.43)***	***0.04***
		Number of flowering stems	*9.59 (3.54)*	*0.32*
		Maximum flowering stem length (cm)	***34.85 (9.35)***	***$1e^{-04}$***
		Number of flowers by flowering stem	***30.27 (25.45)***	***$1e^{-04}$***

		Flowering duration (day)	*31.62 (7.17)*	*0.06*
Ah03	*3 (d208.1/d17.1)*	*Time first leaf (day)*	*13.33 (5.27)*	*0.22*
		Number of leaves	*12.05 (4.25)*	*0.17*
		Rosette area (cm²)	*31.23 (25.7)*	*0.35*
		Mean leaf length (cm)	*1.67 (0.78)*	*0.32*
		St dev leaf length (cm)	*0.28 (0.15)*	*0.42*
		Mean leaf width (cm)	*1.39 (0.62)*	*0.5*
		St dev leaf width (cm)	*0.24 (0.13)*	*0.42*
		Mean leaf area (cm²)	*2.87 (2.24)*	*0.45*
		Time first flower (day)	***34.81 (7.87)***	***0.02***
		Number of flowering stems	***7.67 (3)***	***$3e^{-04}$***
		Maximum flowering stem length (cm)	***44.4 (13.87)***	***0.03***
		Number of flowers by flowering stem	*61.55 (46.06)*	*0.06*
		Flowering duration (day)	***39 (6.6)***	***$5e^{-03}$***
Ah03	*4 (d10.1/d17.1)*	Time first leaf (day)	18 (6.2)	0.12
		Number of leaves	12.33 (4.46)	0.22
		Rosette area (cm²)	**17.71 (11.23)**	**0.04**
		Mean leaf length (cm)	1.52 (0.53)	0.14
		St dev leaf length (cm)	0.23 (0.07)	0.14
		Mean leaf width (cm)	1.23 (0.35)	0.13
		St dev leaf width (cm)	0.23 (0.08)	0.23
		Mean leaf area (cm²)	2.01 (1.34)	0.1
		Time first flower (day)	33.33 (4.23)	0.09
		Number of flowering stems	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
Ah04	*5 (d205.1/d191.1)*	***Time first leaf (day)***	***13.42 (3.28)***	***$1e^{-04}$***
		Number of leaves	*13.45 (2.97)*	*0.32*
		Rosette area (cm²)	*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 (cm²)	*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*
		Maximum flowering stem length (cm)	***62.96 (9.48)***	***$1e^{-04}$***
		Number of flowers by flowering stem	***70.13 (32.5)***	***$1e^{-04}$***
		Flowering duration (day)	*42.44 (7.05)*	*0.3*
Ah04	*6 (d205.1/d50.2)*	*Time first leaf (day)*	*14.31 (4.36)*	*0.30*
		Number of leaves	*13.5 (3.9)*	*0.13*
		Rosette area (cm²)	***44.1 (33.48)***	***0.02***
		Mean leaf length (cm)	***2.04 (0.89)***	***0.04***
		St dev leaf length (cm)	*0.29 (0.12)*	*0.21*
		Mean leaf width (cm)	*1.38 (0.51)*	*0.36*
		St dev leaf width (cm)	***0.21 (0.09)***	***0.03***
		Mean leaf area (cm²)	*3.33 (2.27)*	*0.21*
		Time first flower (day)	***31.92 (3.66)***	***$2e^{-03}$***
		Number of flowering stems	***15.77 (5.25)***	***$1e^{-04}$***
		Maximum flowering stem length (cm)	*60.01 (9.34)*	*0.07*

Number of flowers by flowering stem	***43.51 (15.45)***	***$5e^{-04}$***
Flowering duration (day)	***39.96 (7.38)***	***0.04***

[a] *The P value represents the proportions of the distribution with the expected difference of mean between each family and the other individuals after 10,000 random resamples equal to or less than the value observed. Significant values are represented in bold.*

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., Hoeppner, 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-
967 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.

12. Uyenoyama, M.K. (2003). Genealogy-dependent variation in viability among self-incompatibility genotypes. Theoretical Population Biology 63, 281–293. https://doi.org/10.1016/S0040-5809(03)00020-0.

13. Llaurens, V., Gonthier, L., and Billiard, S. (2009). The sheltered genetic load linked to the S locus in plants: new insights from theoretical and empirical approaches in sporophytic self-incompatibility. Genetics 183, 1105–1118. https://doi.org/10.1534/genetics.109.102707.

14. Castric, V., and Vekemans, X. (2004). Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. Molecular Ecology 13, 2873–2889.https://doi.org/10.1111/j.1365-294X.2004.02267.x.

15. Llaurens, V., Billiard, S., Leducq, J.-B., Castric, V., Klein, E.K., and Vekemans, X. (2008). Does frequency-dependent selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in *Arabidopsis halleri*? Evolution 62, 2545–2557. https://doi.org/10.1111/j.1558-5646.2008.00469.x.

16. Durand, E., Méheust, R., Soucaze, M., Goubet, P.M., Gallina, S., Poux, C., Fobis-Loisy, I., Guillon, E., Gaude, T., Sarazin, A., et al. (2014). Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. Science 346, 1200–1205. https://doi.org/10.1111/eva.12933.

17. Billiard, S., Castric, V., and Vekemans, X. (2006). A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. Genetics 175, 1351–1369. https://doi.org/10.1016/j.tig.2011.06.005.

18. Goubet, P.M., Bergès, H., Bellec, A., Prat, E., Helmstetter, N., Mangenot, S., Gallina, S., Holl, A.-C., Fobis-Loisy, I., Vekemans, X., et al. (2012). Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in Arabidopsis. PLoS Genetics 8, e1002495. https://doi.org/10.1371/journal.pgen.1002495.

19. Stift, M., Hunter, B.D., Shaw, B., Adam, A., Hoebe, P.N., and Mable, B.K. (2013). Inbreeding depression in self-incompatible North-American *Arabidopsis lyrata*: disentangling genomic and S-locus-specific genetic load. Heredity 110, 19–28. https://doi.org/10.1038/hdy.2012.49.

20. Le Veve, A., Burghgraeve, N., Genete, M., et al. (2023) Long-term balancing selection and the genetic load linked to the self-incompatibility locus in *Arabidopsis halleri* and *A. lyrata*. Molecular Biology and Evolution msad120. https://doi.org/10.1093/molbev/msad120.

21. Hu, T.T., Pattyn, P., Bakker, E.G.,, Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43, 476–481. https://doi.org/10.1038/ng.807.

22. Foxe, J.P., Stift, M., Tedder, A., Haudry, A., Wright, S.I., and Mable, B.K. (2010). Reconstructing origins of loss of self-incompatibility and selfing in North american *Arabidopsis lyrata*: a population 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. Clauss, M.J., and Mitchell-Olds, T. (2006). Population genetic structure of *Arabidopsis lyrata* in
1038 Europe. Molecular Ecology 15, 2753–2766. Clauss, 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 metallicolous 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-
1047 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.