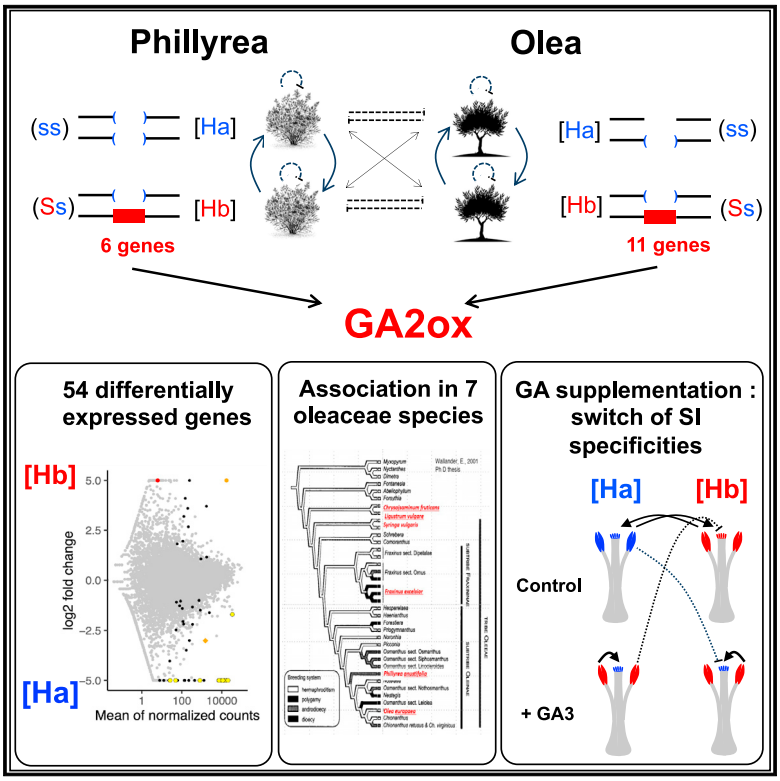
Article

Current Biology

The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene

Graphical abstract



Highlights

- The segregation of a genome fragment controls self-incompatibility (SI) in Oleaceae
- This is the first report of hemizygous control of a homomorphic SI system
- This fragment contains a single conserved gene, related to the gibberellin pathway
- Manipulation of gibberellin levels switches self-incompatibility specificities

Authors

Vincent Castric, Rita A. Batista, Amélie Carré, ..., Xavier Vekemans, Philippe Vernet, Pierre Saumitou-Laprade

Correspondence

pierre.saumitou-laprade@univ-lille.fr

In brief

The self-incompatibility system of the Oleaceae is unusual, with the long-term maintenance of only two groups of reproductive compatibility. Castric et al. identify a hemizygous chromosomal fragment unique to one specificity, containing a single conserved gene with predicted function in catabolism of gibberellin.

Castric et al., 2024, Current Biology *34*, 1–10
May 6, 2024
https://doi.org/10.1016/j.cub.2024.03.047

CellPress

Current Biology

CellPress

Article

The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene

Vincent Castric,[1] Rita A. Batista,[1,6] Amélie Carré,[1,5] Soraya Mousavi,[2] Clément Mazoyer,[1] Cécile Godé,[1] Sophie Gallina,[1] Chloé Ponitzki,[1] Anthony Theron,[3] Arnaud Bellec,[3] William Marande,[3] Sylvain Santoni,[4] Roberto Mariotti,[2] Andrea Rubini,[2] Sylvain Legrand,[1] Sylvain Billiard,[1] Xavier Vekemans,[1] Philippe Vernet,[1] and Pierre Saumitou-Laprade[1,7,*]

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

[2]CNR, Institute of Biosciences and Bioresources (IBBR), 06128 Perugia, Italy

[3]INRAE, CNRGV French Plant Genomic Resource Center, F-31326 Castanet Tolosan, France

[4]UMR DIAPC Diversité et adaptation des plantes cultivées, F-34398 Montpellier, France

[5]Present address: Algal Genetics Group, Integrative Biology of Marine Models Laboratory, CNRS, Sorbonne Université, Station Biologique de Roscoff, F-29680 Roscoff, France

[6]Present address: Department of Algal Development and Evolution, Max Planck Institute for Biology, Max-Planck-Ring 5 72076 Tübingen, Germany

[7]Lead contact

*Correspondence: pierre.saumitou-laprade@univ-lille.fr

https://doi.org/10.1016/j.cub.2024.03.047

SUMMARY

In flowering plants, outcrossing is commonly ensured by self-incompatibility (SI) systems. These can be homomorphic (typically with many different allelic specificities) or can accompany flower heteromorphism (mostly with just two specificities and corresponding floral types). The SI system of the Oleaceae family is unusual, with the long-term maintenance of only two specificities but often without flower morphology differences. To elucidate the genomic architecture and molecular basis of this SI system, we obtained chromosome-scale genome assemblies of *Phillyrea angustifolia* individuals and related them to a genetic map. The S-locus region proved to have a segregating 543-kb indel unique to one specificity, suggesting a hemizygous region, as observed in all distylous systems so far studied at the genomic level. Only one of the predicted genes in this indel region is found in the olive tree, *Olea europaea*, genome, also within a segregating indel. We describe complete association between the presence/absence of this gene and the SI types determined for individuals of seven distantly related Oleaceae species. This gene is predicted to be involved in catabolism of the gibberellic acid (GA) hormone, and experimental manipulation of GA levels in developing buds modified the male and female SI responses of the two specificities in different ways. Our results provide a unique example of a homomorphic SI system, where a single conserved gibberellin-related gene in a hemizygous indel underlies the long-term maintenance of two groups of reproductive compatibility.

INTRODUCTION

In hermaphroditic flowering plants, patterns of reproductive compatibility are commonly governed by genetic self-incompatibility (SI) systems (reviewed by Takayama and Isogai[1]). Such systems occur in about 40% of flowering plants (reviewed by Igic et al.[2]) and have evolved multiple times independently, generating diverse SI systems. Despite decades of effort, the genetic architecture and molecular mechanisms underlying SI have been discovered in only a few plant families. Two largely distinct types of genetic architectures and molecular mechanisms of SI have been described.[3] In species with homomorphic SI systems, individuals have morphologically indistinguishable flowers, but many different reproductive groups defined by distinct SI specificities are commonly found in natural populations (reviewed by Lawrence[4]). SI recognition specificities are controlled by large series of alleles maintained by strong negative-frequency-dependent selection.[5] In these systems, the pollen and pistil specificities are controlled by distinct but tightly linked genes, whose molecular functions differ between the SI systems so far studied (reviewed in Zhang et al.[6]). Pollen specificities may be determined by the pollen's own haploid genotype (gametophytic systems) or by the diploid genotype of the individual producing the pollen (sporophytic control). In contrast, species with heteromorphic SI typically show only two (in some taxa three) cross-compatible flower morphs. The morphological and incompatibility differences in distantly related taxa, including Primula,[7] Fagopyrum,[8] Turnera,[9] Linum,[10] and Gelsemium[11] are controlled by remarkable convergent genome regions (Data S1A) carrying 3–9 genes present only in the dominant morph, which, in all known cases, is the short-styled morph in which the genes are hemizygous. Control of pollen types is sporophytic. Because of the hemizygosity, the short-styled morph is dominant over the long-styled morph and cannot recombine in the heterozygotes.

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

Despite the recent advances just mentioned, the evolutionary relationships between homomorphic and heteromorphic SI systems are not fully understood. A first difficulty is that because the different genes are strictly co-segregating, teasing apart the genes determining the male SI reaction, the female SI reaction and, eventually, the various aspects of flower heteromorphy (pistil length, anther height and sometimes pollen size) has remained challenging in distylous species. A second challenge is to determine the order of events. Charlesworth and Charlesworth[12] modeled evolution from an ancestral state with a (homomorphic) biallelic sporophytic SI system, with the flower morphology differences evolving later, but Lloyd and Webb[13] proposed that flower heteromorphy came first, with SI evolving later to reinforce cross-fertilization.

Plants of the Oleaceae family have a highly unusual SI system, with only two incompatibility types ([Ha] and [Hb]), but often with no discernible morphological differences between them. This SI system was initially discovered in the anemophilous shrub *Phillyrea angustifolia*. In this androdioecious species, hermaphrodite individuals express either the [Ha] or the [Hb] incompatibility types, while the pollen of males can fertilize both types of hermaphrodites.[14] Inter-species pollinations showed that the same two SI specificities are shared between genera as distant as Olea (the olive tree[15]) and Fraxinus,[16] estimated to have been separated for at least 20 to 30 Ma—and up to 50 Ma—of divergence,[17,18] demonstrating the remarkable stability of this homomorphic SI system with only two specificities. The *P. angustifolia* system has sporophytic control of pollen incompatibility; the (ss) genotype has the [Ha] specificity and the [Hb] type has the heterozygous (Ss) genotype (respectively, S1S1 and S1S2 in Billiard et al.[19]). Genetic mapping in *P. angustifolia*[20] and *Olea europaea*[21] found a single mendelian S-locus in both species and showed that their genome positions are probably homologous.

In this study, we compared newly generated chromosome-scale genome assemblies from *P. angustifolia* individuals carrying the determinants of each SI group and found that, unlike any other homomorphic SI system so far studied, the SI specificities are controlled by the segregation of a hemizygous genome region. Close examination of the sequence and gene content of this segregating indel in *P. angustifolia* identified a single shared protein-coding gene with the homologous region in *O. europaea*. This gene is predicted to interfere with the gibberellin pathway, and its presence/absence appears to be perfectly associated with SI specificity phenotypes across six distantly related genera within the Oleeae tribe (a subgroup of the Oleaceae family). Experimental manipulation of gibberellic acid (GA) levels in developing buds modified the male SI response of one specificity and the female SI response of the other specificity. Together with the companion paper by Raimondeau et al. exploring the link with distyly in a broad phylogenetic context, our results identified several key features of this very unusual SI system, which is largely conserved across this economically important family of plants.

RESULTS

The S-locus in *P. angustifolia* corresponds to a 543-kb segregating indel

As no genome assembly was available for *P. angustifolia*, we first generated chromosome-scale genome assemblies for individuals carrying the determinants of each incompatibility type. In this androdioecious species, sex and SI segregate independently[19] and map to two different linkage groups in the genetic map of Carré et al.[20] Males show universal compatibility,[14] which makes it possible to breed homozygous (ss) and (SS) genotypes. To avoid sequencing problems due to potential rearrangements between haplotypes controlling the specificities, which are known in other homomorphic SI systems, such as Arabidopsis[22] and Petunia,[23] we selected individuals that were determined to be homozygous at the SI locus, based on the segregation of their SI specificities.[15]

Our first reference-level genome assembly was for a hermaphrodite individual derived from a population at locality Fabrègue, with [Ha] specificity (genotype ss) by combining HiFi PACBIO sequences (40X) and optical mapping (588X) datasets to obtain a hybrid scaffolding assembly, which we organized into 23 pseudo-chromosomes using the *P. angustifolia* genetic map of Carré et al.,[20] Figure S1A. The reference was highly contiguous (N50 = 33.5 Mb, Data S1B) and highly complete (99.7% of the genes in the benchmarking universal single-copy orthologs [BUSCO] database were complete in our assembly, and only 0.4% were missing). We obtained and assembled PACBIO HiFi reads from a second individual (a genotype SS male) and organized the resulting contigs using the reference. Given the high number of heterozygous sites (Figure S1B), the assembler was able to produce two alternative assemblies for each individual (referred to as "haplotype1" and "haplotype2").

The genetic map of Carré et al.[20] located the S-locus at 53.6 cM on linkage group 18, along with six fully linked genotyping by sequencing (GBS) markers. We used blast to position these GBS markers, as well as the flanking recombinant markers, in the *P. angustifolia* reference genome assembly. This delimited the S-locus to a 5.43-Mb genomic interval between 12,480,240–17,934,613 bp on chromosome 18 (Figures 1and S2A). This region has a low recombination rate (0.5 cM/Mb), low gene density and high transposable element (TE) content (Figure S2A), and probably corresponds to a pericentromeric region. The interval contains 65 predicted protein-coding genes. We first noted the presence of an inversion of about 5 Mb in one haplotype of the (ss) individual (Figure S2A), but the low divergence between the inverted and non-inverted haplotypes (Figure S2B) and inspection of the inversion breakpoint using long sequencing reads from a third *P. angustifolia* individual with the (Ss) genotype (Figure S2C) ruled out the possibility that the inversion itself is causally related to the SI specificities (supplemental information). To precisely delimit the region containing the SI determinants (Figure 1), we asked whether any of the 65 predicted protein-coding genes located in the previously identified genomic interval included single-nucleotide polymorphisms (SNPs) strictly associated with given SI phenotypes, using RNA sequencing (RNA-seq) data collected from entire flower buds and pistils from fourteen *P. angustifolia* individuals from the same local population (Fabrègues). Eight of these individuals were determined by controlled pollination assays to have incompatibility type [Ha] and six to be [Hb] (Data S1C). We found no fully associated SNP in the whole interval, which excludes these protein-coding genes as the SI determinants.

Sequence identity in this region between the two assemblies was high, but there was a 543-kb indel with sequences that were specific to the (SS) assembly (Figure 1), suggesting presence of a hemizygous region that might be the S-locus. This 543-kb indel contains six predicted protein-coding genes (Data

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

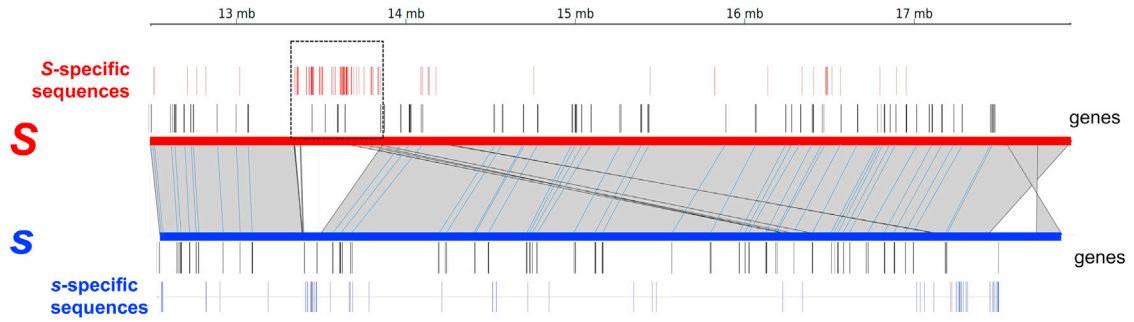


Figure 1. Chromosomes controlling the [Hb] specificity carry a 543-kb indel

The candidate 5-Mb interval contains 65 predicted protein-coding genes, and a 543-kb indel (box delimited by dotted line) spanning over six genes contains a large number of [Hb]-specific sequences. Haplotype 1 of the (SS) individual was aligned on haplotype 2 (non-inverted) of the (ss) individual. Gray areas delimited by thin black lines connect aligned portions of the two chromosomes. Predicted protein-coding genes are represented by black vertical lines along each of the two chromosomes. The s- and S-specific 300-bp sequences are represented by blue and red vertical lines, respectively.
See also Figures S1 and S2; Data S1A.

S1D). One is a probable sucrose-phosphate synthase, another encodes a replication protein A 70-kDa DNA-binding subunit D-like, and one has strong similarity to Gypsy retroelements. Two additional genes are predicted to encode uncharacterized proteins with no recognized Pfam domains. The sixth predicted gene (*PaGA2ox-S*) is a gibberellin-2-oxidase. None of the molecular functions predicted for these six genes have ever been associated with the control of SI in any species.

The S-locus hemizygous indel is shared with Olea and contains a single conserved gene

To determine the possible role of these genes in the control of SI, we reasoned that, because the male and female SI reactions are shared between distant genera in the Oleaceae, the genetic factors controlling them might also be shared. Therefore we first tested for a shared genetic architecture by comparing the region identified in *P. angustifolia* with the orthologous region in the recently published genome assembly of the olive tree, *O. europaea*, *var Arbequina*.[24] This cultivar is known to have the (Ss) genotype at the S-locus.[25] We found strongly conserved gene content and order on chromosome 18 in both genome assemblies, but poor conservation of intergenic regions (Figure 2), as expected after the estimated ca. 32 Ma of divergence between Olea and Phillyrea.[18] To determine whether a segregating indel was also present at the orthologous position in Olea, we mapped short reads from olive tree cultivars[26] with known SI genotypes onto the Arbequina reference assembly (four were ss and four Ss, see Mariotti et al.,[27] Data S1E). We found a 756-kb fragment where (Ss) cultivars consistently had mapping densities of about half the chromosomal median value, and no reads from (ss) cultivars mapped (Figure 2). Hence, it appears that, also in the olive, the SI groups are associated with an indel at an orthologous position to that in *P. angustifolia*, and one S-locus haplotype is again hemizygous. The indel sizes in Phillyrea and Olea differ somewhat (543 vs. 756 kb, respectively), similar to their overall relative genome size ratio (803 Mb vs. 1.3 Gb, or a ratio of 0.62[24]).

Comparison of the gene content of the indels of Phillyrea (see above) and Olea (with 13 annotated genes; Data S1F) showed that only one was common to both species, the candidate *GA2ox-S* gene already described above (Figures 2 and S3). Notably, the *BZR1-S* gene described in the companion paper by Raimondeau et al. in Oleae and several other Oleaceae is present in the Olea assembly (with two duplicated copies) but is absent from the Phillyrea indel (Figure S3;Data S1F). Hence, as the SI functional response is conserved in the two species, *GA2ox-S* remains a prime candidate for the control of SI.

The presence/absence polymorphism of the *GA2ox-S* gene is stably associated with SI groups across Oleaceae

To expand the phylogenetic scale of our analysis, we retrieved published genome assemblies from twenty Oleaceae species belonging to five genera and available from NCBI (Data S1G). We used BLAST to recover the sequence of the *PaGA2ox-S* orthologs in the other species. We found that *GA2ox-S* was present in ten of the twenty assemblies, Data S1G), consistent with the expected 1:1 frequencies of the [Ha] and [Hb] incompatibility types in natural populations of each species. When the *GA2ox-S* sequence was present in a species in any of the four genera of the Oleeae tribe included (Phillyrea, Olea, Fraxinus, and Syringa), their sequences cluster with those of other species in the same genus (Figure 3A).

To further establish the role of *GA2ox-S* in SI, we examined associations of its sequences in diverged species after many generations of recombination or mutation. We used the *GA2ox-S* sequences obtained above to design highly specific PCR primers and tracked the presence of *GA2ox-S* in individuals whose SI groups we had previously determined[14,25,28,29] or that we had newly phenotyped (in *Olea europaea*, *Syringa vulgaris*, and *Chrysojasminum fruticans*, Data S1H). The presence of *GA2ox-S* is completely associated with the dominant [Hb] specificity across all samples from the six tested species, including 47 *P. angustifolia* accessions from distant populations, 10 accessions of a different Phillyrea species, *P. latifolia*, 132 cultivated *O. europaea* varieties, 38 wild *O. europaea* var. *sylvestris*, as well as 26, 36, and 24 accessions of the even more distant *Fraxinus excelsior*, *Syringa vulgaris*, and *Ligustrum vulgare*, respectively (Figures 3B, 3C, 3D, 3E, 3F, and 3G; Data S1H). Notably, the association was also maintained in the distylous species *Chrysojasminum fruticans*, which is basal in the Oleaceae; in this species, *GA2ox-S* was present in all five short-styled individuals but absent from all seven long-styled individuals tested

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

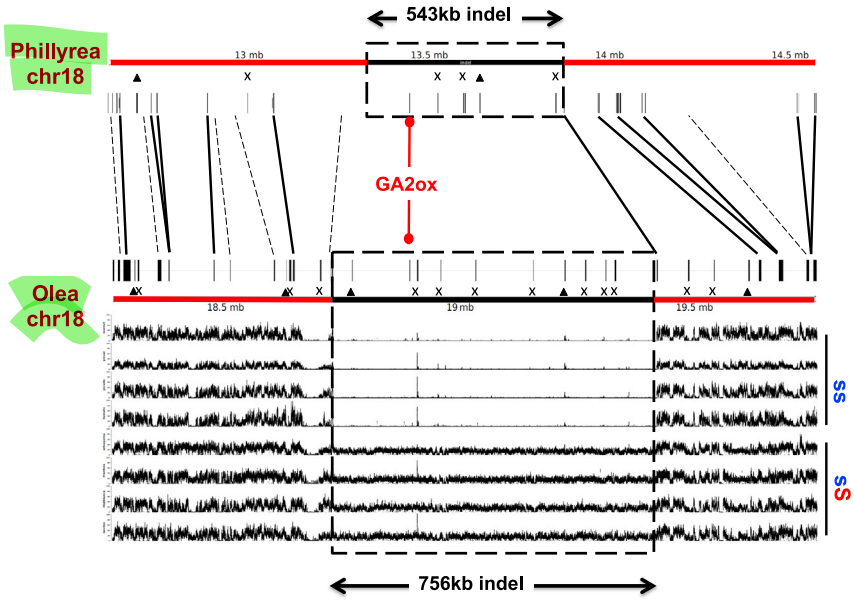


Figure 2. The indel in *P. angustifolia* contains six predicted protein-coding genes

Comparison of the indel sequence between *P. angustifolia* and *O. europaea* (var. Arbequina) reveals that *GA2ox-S* is the only conserved gene in the indel, with high divergence of intergenic sequences. Short reads mapping of *O. europaea* accessions (at the bottom of the figure, y axis indicates read mapping depth up to 100X) identifies a segregating 756-kb indel in the Arbequina chromosome 18 assembly. Triangles indicate annotated genes that putatively correspond to transposable elements. Crosses indicate genes with no sequence similarity (by blast) with anything in the chromosomal fragment of the other species. Solid lines (black and red) indicate orthologous genes. Interrupted lines indicate genes in one species with strong sequence similarity but no gene annotation in the chromosomal fragment of the other species.
See also Data S1D–S1F and Figures S3 and S4.

(Figure 3H). This suggests that the short-styled morph corresponds to the [Hb] specificity, whereas the long-styled morph corresponds to the [Ha] specificity, thus formally establishing a relationship between the determinants of the homomorphic SI in this tribe and those of distyly across the Oleaceae family (see the companion paper by Raimondeau et al. for a detailed phylogenetic analysis of the supergene).

The GA2ox-S gene encodes a class-I gibberellin oxidase expressed in floral buds

Sequence analysis of the S-locus *PaGA2ox-S* gene indicates that it is a class-I gibberellin 2 oxidase enzyme (Figure S4A). GA2ox proteins have specialized functions in the inactivation of GA precursor molecules at different stages of the GA biosynthesis pathway.[30] Its sequence similarity with class-I GA2 oxidases of other flowering plants predicts that the *PaGA2ox-S* protein functions to degrade "end" products of the GA biosynthesis pathway, including the bioactive forms GA1 and GA4. Two protein domains can be identified: the 2-oxoglutarate/Fe(II)-dependent oxygenase domain (2OG-FeII_Oxy), spanning exon 2 and 3, and a DIOX_N domain, specifically located in exon 1. The latter corresponds to the highly conserved N-terminal region of proteins with 2-oxoglutarate/Fe(II)-dependent oxygenase activity. Both these domains are consistently found in GA2 oxidases of other flowering plants and are a defining characteristic of this gene family (Figure S4B[31,32]). Quantitative real-time PCR (RT-qPCR) assays confirmed that *GA2ox-S* transcripts could be detected only in [Hb] individuals (Figure S4C), with robust expression in mature pistils. Inspection of the melting curve of the amplification product indicated that low but specific expression was also detectable in mature anthers of one of the three sampled biological replicates, but not in germinating pollen tubes. Due to the small size of immature anthers and pistils, we were not able to dissect these tissues at earlier developmental stages, where expression of SI determinants could potentially occur. Nevertheless, these results suggest that *PaGA2ox-S* has the potential to be expressed in both female and male reproductive tissues. Additionally, we could not detect expression in non-reproductive tissues (leaves) of any individual (Figure S4C).

Using RNA-seq data, we identified a total of 54 differentially expressed genes in whole buds when comparing *P. angustifolia* [Ha] and [Hb] individuals (Figure 4; Data S1I). Besides the [Hb]-specific *PaGA2ox-S* expression, 15 genes encoding proteins of the pollen Ole e 1 allergen and extensin family were much more highly expressed in [Ha] individuals. These genes are located on chromosomes 10, 12, and 17 and are not in the S-locus. They are homologs to either At1g29140, At5g45880, or At3g16670 in *A. thaliana*, where they are expressed broadly, including in mature and germinating pollen grains.[33,34] In Oleaceae, these proteins are produced in gametophytic and sporophytic tissues during pollen grain development and are thought to be important in recognition processes between pollen and stigma and between pollen tubes and style cells.[35,36] An additional RNA-seq experiment on pollinated pistils from the same individuals revealed 16 differentially expressed genes (Figure S5; Data S1J). *PaGA2ox-S* was identified in both RNA-seq experiments, along with two other non-S-locus genes. One is located on chromosome 23 and encodes an aspartic proteinase; this gene's expression is higher in plants with the [Hb] specificity. The other gene is on chromosome 10 and encodes a proteinase inhibitor PSI-1.2, and its expression is higher in the other specificity, [Ha]. There could be an antagonistic action between their proteinase and proteinase inhibitor actions. Antagonism occurs in the S-RNase SI system (a homomorphic system), where a female S-RNase protein is specifically detoxified by the male S-locus F-box (SLF) proteins.[37]

Treatment with gibberellin disrupts the pollen and stigma SI response in an S-allele-specific manner

Because *GA2ox-S* has the putative function of reducing the levels of active GA, we envision a simple model under which [Ha] pistils and pollen have basal levels of GA hormone, whose active form is lower in the presence of *GA2ox-S* in [Hb] individuals. To test this, we applied exogenous GA3 to immature *Ligustrum vulgare* floral buds and examined the resulting pollen and pistil SI specificities when the flowers opened. We chose *L. vulgare* because the structure, morphology, and size of its inflorescences allow easy

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

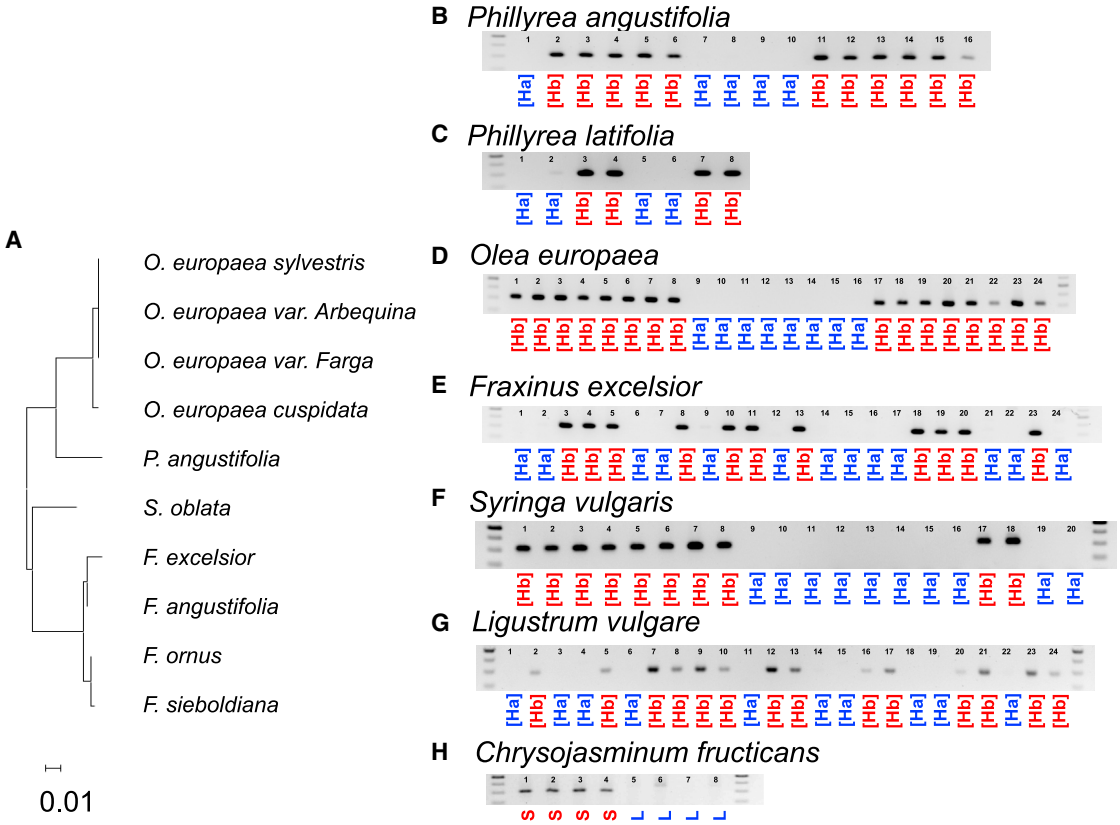


Figure 3. Presence of GA2ox-S is stably associated with SI phenotypes across distant Oleaceae species

(A) Maximum likelihood phylogeny of the first exon of GA2ox-S across ten Oleeae genomes.

(B–G) Presence/absence of PCR products fully correlates with SI groups: (B) in *P. angustifolia* from distant populations, (C) in a distinct *Phillyrea* species (*P. latifolia*), (D) in a diverse set of *O. europaea* accessions, (E) in *Fraxinus excelsior*, (F) in *Syringa vulgaris*, (G) in *Ligustrum vulgare*, and (H) in *Chrysojasminum fructicans* (S, short-styled; L, long-styled). The full list of samples tested is reported in Data S1H. Numbers above the sample lanes are provided for cross-referencing with Data S1H.

See also Data S1G and S1K.

treatment with GA3 by inflorescence dipping. GA treatment of [Ha] buds (the ss genotype) caused the specificity of pollen to switch to the opposite of the normal reaction: pollen of treated buds was rejected by untreated [Hb] pistils but germinated on untreated [Ha] pistils, while the pistil [Ha] specificity remained unchanged. Treatment of [Hb] buds (the Ss genotype, with the GA2ox-S gene present) did not change the pollen specificity, but the pistil specificity switched to that expected for a [Ha] plant (Figures 5A and 5B). Interestingly, the treated buds of both groups all became self-compatible because the pollen and pistil change induced by the treatment caused their pollen and pistil SI specificities to be different (and not because the rejection reaction of their pollen or pistils was abolished). This closely resembles the change to self-compatibility in homostyled individuals of distylous species (see, e.g., Haldane[38]).

DISCUSSION

Hemizygosity underlies the homomorphic SI system of Oleaceae

Our findings, together with the results of the companion paper by Raimondeau et al., are the first reports of control of a homomorphic SI system by presence/absence of a genome region, with one incompatibility type hemizygous for a sequence that is absent from the other type. Such a system has previously been found in distylous plants, i.e., in heteromorphic SI systems. In homomorphic SI systems, natural selection is expected to lead to the diversification of S-allele lineages, especially when the allele number is initially low,[39,40] and allelic diversity is indeed high in all species so far investigated.[4] The binary nature of the presence/absence system in the Oleaceae may constrain the number of SI specificities to two.

Interestingly, several basal Oleaceae species are distylous,[41–43] and the presence of *GA2ox-S* is specific to one of the two morphs of such species, a jasmine. The homomorphic SI system might therefore derive from the loss of flower heteromorphism in an ancestrally heteromorphic SI system.[13] However, the alternative—that SI evolved before flower heteromorphism[12]—cannot be definitively excluded. Testing this alternative will require clarifying the phylogeny of basal Oleaceae (including a broader range of homomorphic and heteromorphic species) and determining functionality and homology of their eventual SI systems. If the hemizygous region is flanked by genome regions that are likely to be homologous between homomorphic and heteromorphic species, based, e.g., on having the same gene content, this would

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

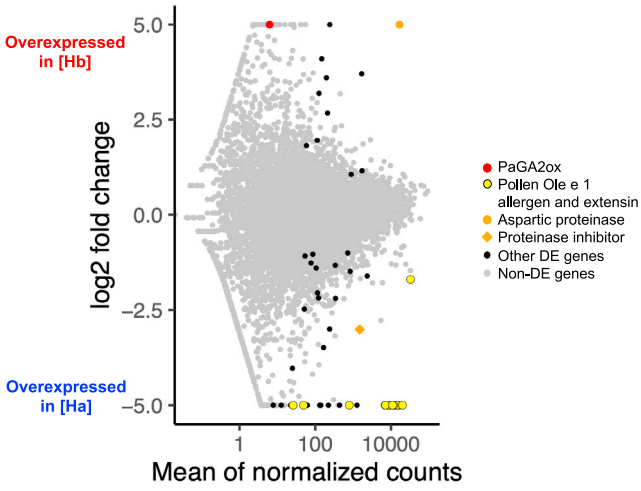


Figure 4. Differential expression of genes in buds from [Ha] vs. [Hb] individuals

The red dot corresponds to PaGA2ox-S transcripts. The orange dot corresponds to the aspartic proteinase and the orange diamond to the proteinase inhibitor. The yellow dots correspond to the pollen Ole e 1 allergen and extensin proteins. Black dots indicate the other differentially expressed genes (adjusted p value below 0.01 and log2 fold change above 1; full annotation list in Data S1I). Genes with no significant expression difference are represented by gray dots.

See also Data S1C and S1J and Figures S4 and S5.

support the conclusion that one system evolved from the other (though determining the direction of change will be more difficult). Our observation in Chrysojasminum and the phylogenetic analysis in the companion paper by Raimondeau et al. suggest that the heteromorphic vs. homomorphic SI systems are indeed related, such that decoupling between SI and floral heteromorphism in some Oleaceae will be a powerful way to tease apart the contributions of each S-locus gene to the determination of SI specificities and/or to the morphological differences between floral morphs. We note that functional SI is maintained without heteromorphy, not only in the anemophilous Phillyrea and Olea but also in the entomophilous Ligustrum and Syringa, where the selective pressure to specialize floral morphs would be much stronger. Gelsemium represents an interesting reciprocal situation, with flower dimorphism but no functional SI.[11] In other distylous species so far studied, the distinction is challenging to operate because the gene determining the female SI specificity also controls style length (TsBAHD in Turnera,[44] CYP734A50 in Primula,[45] and S-ELF3 in Fagopyrum[46]). Stylar incompatibility, controlled separately from style length, has been reported in Plumbaginaceae (e.g., in Armeria), where striking pollen and stigma polymorphisms exist without differences on their respective positions between the two mating types.[47] Identifying the molecular mechanism of the SI system in families with heterostylous species or with other types of floral heteromorphism will enable the systematic search for SI genes, not only in heterostylous species but also in non-heterostylous species within the same families. We anticipate that such an approach could reveal other cases of homomorphic SI systems with only two specificities and no floral heteromorphy. Early work by Baker[48] and Gibbs[49] mentioned the difficulty of establishing the number of SI specificities (which requires extensive crossing experiments), relative to the ease with which distyly can be spotted with the naked eye. Hence, it is possible that homomorphic SI systems with only two SI specificities may actually be more common across the flowering plants than is currently appreciated.

Finally, we note that, in heterostylous genera, the short-styled morph has been associated with the dominant S-locus haplotype in almost all cases investigated.[43] Just as in Primula, Fagopyrum, Linum, and Gelsemium, we found that the indel in the heterostylous Chrysojasminum is associated with the short-styled morph, and this has been confirmed in the companion paper by Raimondeau et al. Once we understand the developmental effects of the presence/absence polymorphism in these different systems better, it will be interesting to determine whether this molecular convergence is more than just a coincidence.

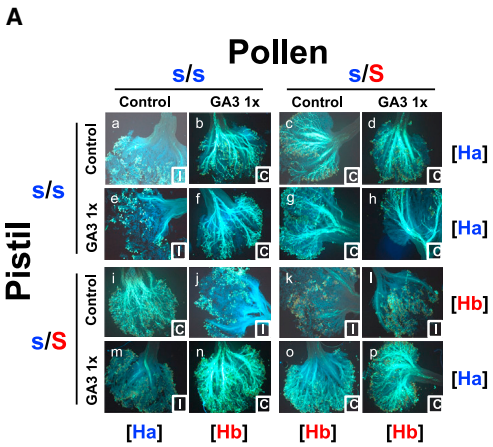
GA as a new pathway controlling SI

Our study establishes a role for GA in the control of the two SI specificities. The GA hormone represents a novel pathway for the control of SI, in particular in homomorphic SI, where receptor/ligand or toxin/antitoxin systems have been described in all species investigated so far. The involvement of a plant hormone, together with the occurrence of hemizygosity, is shared with heterostylous lineages (Data S1A). For instance, one of the three genes in the S-locus of Gelsemium is annotated as a GA3ox and is involved in style length.[11] In Primula, brassinosteroids are involved in the determination of style length female specificity,[45,50] while in the short-styled morph of Turnera, two S-genes are associated with an increased level of auxin in the anthers and a decreased level of brassinosteroids in pistils.[44] These differences in the biochemical environments of the anther and pistil are hypothesized to underlie SI.[9,51] Determining whether the control of SI specificities by GA in the Oleaceae system follows this "hormonal mismatch" mechanism will require biochemical methods to characterize how the relative amounts of the different forms of active/inactive GA differ between the two SI specificities in male and female reproductive organs. Although it is not yet clear how plant hormones might establish SI specificities, and the action of hormones commonly relies on negative feedback loops that can be complex, our observations show that a simple GA treatment can affect the specificities. This may be of use in olive production when the availability of compatible pollen is low, or to obtain selfed progenies.

Determination of the male and female SI specificities

A single gene, *GA2ox-S*, is shared between the S-locus in Phillyrea and Olea, even though pollen from one species can trigger the SI response on pistils of the other species and vice versa, indicating a shared SI mechanism. The conservation of a single gene is puzzling because, to our knowledge, all SI systems whose genetic basis has been understood so far, whether homomorphic or heteromorphic, involve at least two separate genes, one determining the male specificity and the other determining the female specificity.[52,53] The genes in different species function in very different ways,[54] and the male specificity can sometimes depend on multiple tandemly duplicated paralogs (as in Solanaceae[55]). It remains possible that an S-locus fragment might be missing from our genome assembly, but this is unlikely. First, all our assembly metrics indicate high completeness.

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047



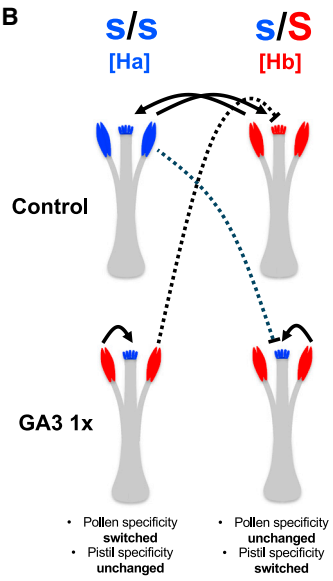


Figure 5. Treatment with gibberellin modifies the male and female SI responses of the two specificities in different manners

(A) Fluorescent microscopic images of controlled pollination assays in *Ligustrum vulgare*. Compatible and incompatible reactions (indicated in the bottom right corner of each image) are identified based on the presence of germinated pollen tubes reaching the style.

(B) Graphical summary of the effect of the GA supplementation assay, illustrating that pistils of the [Ha] specificity and pollen of the [Hb] specificity were GA insensitive, while, conversely, pollen of the [Ha] specificity and pistils of the [Hb] specificity switched specificity upon GA3 treatment.

See also Figure S4.

Furthermore, the two Phillyrea [Hb] specificity haplotypes we assembled both led to the conclusion that GA2ox-S is the single gene shared with the Olea S-locus. However, a genetic element other than a protein-coding gene, for example, a non-coding RNA, could act as a second factor. Small non-coding RNAs have been reported at other SI loci, including the Brassicaceae S-locus,[56,57] but they do not directly determine allelic specificities and rather control the allele's transcript levels.[58] The possibility that *GA2ox-S* could control both the male and female SI specificities is reinforced by our observation that GA3 treatment switched specificity of the male function of [Ha] individuals and of the female function of [Hb] individuals. We currently hypothesize that GA2ox-S acts by activating downstream non-S-locus genes, consistent with the finding that specific genes are expressed in individuals with the two different SI specificities. Formal proof of the possibility that *GA2ox-S* controls both the male and female specificities will require creation of knockin and knockout mutants by genetic transformation. Although the extended development time of Oleaceae species makes such experiments slow, our work provides the foundation required for their design.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Biological material and SI genotype determination
  - HMW DNA isolation
  - PACBIO sequencing and assembly
  - Optical map
  - Scaffolding contigs with the optical and genetic maps
  - HMW DNA isolation and Oxford Nanopore sequencing
  - Preparation of RNA samples
  - RNA-seq library construction and sequencing
  - A *de novo* transcriptome assembly for *P. angustifolia*
  - Annotation of protein-coding genes and transposable elements
  - Genome alignment and sequence comparison
  - Mapping short Illumina reads from *Olea europaea* accessions
  - Sequence comparison across distant Oleaceae species, specific primer design, PCR protocol and association study across Phillyrea and Olea accessions
  - Phylogeny of GA2ox-S proteins
  - RT-qPCR
  - GA2 supplementation experiment
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.03.047.

ACKNOWLEDGMENTS

Sequencing of the *P. angustifolia* genomes was funded by France_Olive as part of the study agreement entitled "Study of the self-incompatibility locus in olive and Phillyrea" to P.S.L. Resources for this work were also partly provided from the European Research Council (NOVEL project, grant #648321) to V.C. and ANR TE-MoMa (grant ANR-18-CE02-0020-01) to V.C. and S.L., as well as an EMBO fellowship (ALTF 657-2020) to R.A.B. We would like to thank Antònia Nino from the IRTA institute and Hélène Lasserre and Julien Balajas from France Olive for collecting inflorescences and providing access to the olive cultivars and oleaster collections. We would like to thank Arnaud Dowkiw and Lemonniers Nurseries for access to *Fraxinus excelsior* samples, Hanno Schaefer for collecting *Ligustrum vulgare* inflorescences in the Isar Valley (Germany), and Amelia Dumbravă for access to *Syringa vulgaris* populations in the Portes de Fer Natural Park (Romania). Finally, we would like to thank Michel and Christine Prat for their warm welcome and for collecting

Chrysojasminum fruticans. The authors thank Luciana Baldoni for her contribution to the study design at the early stages of the project, as well as Jacques Lepart†, Mathilde Dufay, Patrick Achard, Peter Hedden, Tanja Slotte, Michael Lenhard, Isabelle Fobis-Loisy, and Marie Monniaux for scientific discussions. We thank two anonymous reviewers for constructive comments. This work was performed using the infrastructure and technical support of the Plateforme Serre, cultures et terrains expérimentaux – Université de Lille for the greenhouse/field facilities. We thank David Degueldre and the staff of the Plateforme des Terrains d'Expérience du LabEx CeMEB (CEFE, CNRS) for plant cultivation.

AUTHOR CONTRIBUTIONS

Conceptualization, P.S.-L. and P.V.; methodology, S.L., R.A.B., and X.V.; investigation, P.S.-L., V.C., R.A.B., A.C., A.T., W.M., S.S., C.G., S.M., R.M., C.M., and S.L.; software, C.M., S.G., and A.R.; resources, A.B.; writing – original draft, V.C., P.S.-L., A.C., R.A.B., A.T., W.M., S.M., R.M., and A.R.; writing – review & editing, V.C., P.S.-L., S.B., X.V., P.V., and R.A.B.; funding acquisition, P.S.-L. and V.C.; supervision, P.S.-L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 19, 2023
Revised: February 29, 2024
Accepted: March 25, 2024
Published: April 15, 2024

REFERENCES

1. Takayama, S., and Isogai, A. (2005). Self-incompatibility in plants. Annu. Rev. Plant Biol. *56*, 467–489. https://doi.org/10.1146/annurev.arplant.56.032604.144249.
2. Igic, B., Lande, R., and Kohn, J.R. (2008). Loss of self-incompatibility and its evolutionary consequences. Int. J. Plant Sci. *169*, 93–104. https://doi.org/10.1086/523362.
3. De Nettancourt, D. (2001). Incompatibility and incongruity in wild and cultivated plants (Springer-Verlag). https://doi.org/10.1007/978-3-662-04502-2.
4. Lawrence, M.J. (2000). Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. Ann. Bot. *85*, 221–226. https://doi.org/10.1006/anbo.1999.1044.
5. Wright, S. (1939). The distribution of self-sterility alleles in populations. Genetics *24*, 538–552. https://doi.org/10.1093/genetics/24.4.538.
6. Zhang, D., Li, Y.Y., Zhao, X., Zhang, C., Liu, D.K., Lan, S., Yin, W., and Liu, Z.J. (2024). Molecular insights into self-incompatibility systems: From evolution to breeding. Plant Commun. *5*, 100719. https://doi.org/10.1016/j.xplc.2023.100719.
7. Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C.V., and Gilmartin, P.M. (2016). Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. Nat. Plants *2*, 16188. https://doi.org/10.1038/nplants.2016.188.
8. Yasui, Y., Hirakawa, H., Ueno, M., Matsui, K., Katsube-Tanaka, T., Yang, S.J., Aii, J., Sato, S., and Mori, M. (2016). Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. DNA Res. *23*, 215–224. https://doi.org/10.1093/dnares/dsw012.
9. Shore, J.S., Hamam, H.J., Chafe, P.D.J., Labonne, J.D.J., Henning, P.M., and McCubbin, A.G. (2019). The long and short of the S-locus in Turnera (Passifloraceae). New Phytol. *224*, 1316–1329. https://doi.org/10.1111/nph.15970.
10. Gutiérrez-Valencia, J., Fracassetti, M., Berdan, E.L., Bunikis, I., Soler, L., Dainat, J., Kutschera, V.E., Losvik, A., Désamoré, A., Hughes, P.W., et al. (2022). Genomic analyses of the Linum distyly supergene reveal convergent evolution at the molecular level. Curr. Biol. *32*, 4360–4371.e6. https://doi.org/10.1016/j.cub.2022.08.042.
11. Zhao, Z., Zhang, Y., Shi, M., Liu, Z., Xu, Y., Luo, Z., Yuan, S., Tu, T., Sun, Z., Zhang, D., and Barrett, S.C.H. (2023). Genomic evidence supports the genetic convergence of a supergene controlling the distylous floral syndrome. New Phytol. *237*, 601–614. https://doi.org/10.1111/nph.18540.
12. Charlesworth, D., and Charlesworth, B. (1979). A model for the evolution of distyly. Am. Nat. *114*, 467–498. https://doi.org/10.1086/283496.
13. Lloyd, D.G., and Webb, C.J. (1992). The evolution of heterostyly. In Evolution and Function of Heterostyly, S.C.H. Barrett, ed. (Springer), pp. 151–178. https://doi.org/10.1007/978-3-642-86656-2_6.
14. Saumitou-Laprade, P., Vernet, P., Vassiliadis, C., Hoareau, Y., de Magny, G., Dommée, B., and Lepart, J. (2010). A self-incompatibility system explains high male frequencies in an androdioecious plant. Science *327*, 1648–1650. https://doi.org/10.1126/science.1186687.
15. Saumitou-Laprade, P., Vernet, P., Vekemans, X., Billiard, S., Gallina, S., Essalouh, L., Mhaïs, A., Moukhli, A., El Bakkali, A., Barcaccia, G., et al. (2017). Elucidation of the genetic architecture of self-incompatibility in olive: evolutionary consequences and perspectives for orchard management. Evol. Appl. *10*, 867–880. https://doi.org/10.1111/eva.12457.
16. Vernet, P., Lepercq, P., Billiard, S., Bourceaux, A., Lepart, J., Dommée, B., and Saumitou-Laprade, P. (2016). Evidence for the long-term maintenance of a rare self-incompatibility system in Oleaceae. New Phytol. *210*, 1408–1417. https://doi.org/10.1111/nph.13872.
17. Unver, T., Wu, Z., Sterck, L., Turktas, M., Lohaus, R., Li, Z., Yang, M., He, L., Deng, T., Escalante, F.J., et al. (2017). Genome of wild olive and the evolution of oil biosynthesis. Proc. Natl. Acad. Sci. USA *114*, E9413–E9422. https://doi.org/10.1073/pnas.1708621114.
18. Olofsson, J.K., Cantera, I., Van de Paer, C., Hong-Wa, C., Zedane, L., Dunning, L.T., Alberti, A., Christin, P.A., and Besnard, G. (2019). Phylogenomics using low-depth whole genome sequencing: a case study with the olive tribe. Mol. Ecol. Resour. *19*, 877–892. https://doi.org/10.1111/1755-0998.13016.
19. Billiard, S., Husse, L., Lepercq, P., Godé, C., Bourceaux, A., Lepart, J., Vernet, P., and Saumitou-Laprade, P. (2015). Selfish male-determining element favors the transition from hermaphroditism to androdioecy. Evolution *69*, 683–693. https://doi.org/10.1111/evo.12613.
20. Carré, A., Gallina, S., Santoni, S., Vernet, P., Godé, C., Castric, V., and Saumitou-Laprade, P. (2021). Genetic mapping of sex and self-incompatibility determinants in the androdioecious plant *Phillyrea angustifolia*. Peer Community J. *1*, e15. https://doi.org/10.24072/pcjournal.23.
21. Mariotti, R., Fornasiero, A., Mousavi, S., Cultrera, N.G.M., Brizioli, F., Pandolfi, S., Passeri, V., Rossi, M., Magris, G., Scalabrin, S., et al. (2020). Genetic mapping of the incompatibility locus in Olive and development of a linked sequence-tagged site marker. Front. Plant Sci. *10*, 1760. https://doi.org/10.3389/fpls.2019.01760.
22. Goubet, P.M., Bergès, H., Bellec, A., Prat, E., Helmstetter, N., Mangenot, S., Gallina, S., Holl, A.C., Fobis-Loisy, I., Vekemans, X., and Castric, V. (2012). Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in Arabidopsis. PloS Genet. *8*, e1002495. https://doi.org/10.1371/journal.pgen.1002495.
23. Wu, L., Williams, J.S., Sun, L., and Kao, T.H. (2020). Sequence analysis of the *Petunia inflata* S-locus region containing 17 S-Locus F-Box genes and the S-Rnase gene involved in self-incompatibility. Plant J. *104*, 1348–1368. https://doi.org/10.1111/tpj.15005.
24. Rao, G., Zhang, J., Liu, X., Lin, C., Xin, H., Xue, L., and Wang, C. (2021). *De novo* assembly of a new Olea europaea genome accession using nanopore sequencing. Hortic. Res. *8*, 64. https://doi.org/10.1038/s41438-021-00498-y.
25. Saumitou-Laprade, P., Vernet, P., Vekemans, X., Castric, V., Barcaccia, G., Khadari, B., and Baldoni, L. (2017). Controlling for genetic identity of varieties, pollen contamination and stigma receptivity is essential to characterize the self-incompatibility system of *Olea europaea* L. Evol. Appl. *10*, 860–866. https://doi.org/10.1111/eva.12498.
26. Jiménez-Ruiz, J., Ramírez-Tejero, J.A., Fernández-Pozo, N., Leyva-Pérez, M.O., Yan, H., Rosa, R., Belaj, A., Montes, E., Rodríguez-Ariza, M.O., Navarro, F., et al. (2020). Transposon activation is a major driver in the

genome evolution of cultivated olive trees (*Olea europaea* L.). Plant Genome *13*, e20010. https://doi.org/10.1002/tpg2.20010.

27. Mariotti, R., Pandolfi, S., De Cauwer, I., Saumitou-Laprade, P., Vernet, P., Rossi, M., Baglivo, F., Baldoni, L., and Mousavi, S. (2021). Diallelic self-incompatibility is the main determinant of fertilization patterns in olive orchards. Evol. Appl. *14*, 983–995. https://doi.org/10.1111/eva.13175.
28. Saumitou-Laprade, P., Vernet, P., Dowkiw, A., Bertrand, S., Billiard, S., Albert, B., Gouyon, P.H., and Dufaÿ, M. (2018). Polygamy or subdioecy? The impact of diallelic self-incompatibility on the sexual system in *Fraxinus excelsior* (Oleaceae). Proc. Biol. Sci. *285*, 20180004. https://doi.org/10.1098/rspb.2018.0004.
29. De Cauwer, I., Vernet, P., Billiard, S., Godé, C., Bourceaux, A., Ponitzki, C., and Saumitou-Laprade, P. (2021). Widespread coexistence of self-compatible and self-incompatible phenotypes in a diallelic self-incompatibility system in *Ligustrum vulgare* (Oleaceae). Heredity *127*, 384–392. https://doi.org/10.1038/s41437-021-00463-4.
30. Ouellette, L., Anh Tuan, P.A., Toora, P.K., Yamaguchi, S., and Ayele, B.T. (2023). Heterologous functional analysis and expression patterns of gibberellin 2-oxidase genes of barley (*Hordeum vulgare* L.). Gene *861*, 147255. https://doi.org/10.1016/j.gene.2023.147255.
31. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.
32. Cheng, J., Ma, J., Zheng, X., Lv, H., Zhang, M., Tan, B., Ye, X., Wang, W., Zhang, L., Li, Z., et al. (2021). Functional analysis of the gibberellin 2-oxidase gene family in peach. Front. Plant Sci. *12*, 619158. https://doi.org/10.3389/fpls.2021.619158.
33. Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z., and Wu, W.H. (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. Plant Physiol. *148*, 1201–1211. https://doi.org/10.1104/pp.108.126375.
34. Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nat. Genet. *37*, 501–506. https://doi.org/10.1038/ng1543.
35. de Dios Alché, J., M'rani-Alaoui, M., Castro, A.J., and Rodríguez-García, M.I. (2004). Ole e 1, the major allergen from olive (*Olea europaea* L.) pollen, increases its expression and is released to the culture medium during in vitro germination. Plant Cell Physiol. *45*, 1149–1157. https://doi.org/10.1093/pcp/pch127.
36. Rodríguez-Rajo, F.J., Vega-Maray, A., Asturias, J.A., Jato, V., Seoane-Camba, J.A., and Suárez-Cervera, M. (2010). The relationship between tapetum cells and microspores based on protein localization in *Fraxinus angustifolia* (Oleaceae) pollen grains. Int. J. Plant Sci. *171*, 34–52. https://doi.org/10.1086/647922.
37. Kubo, K., Entani, T., Takara, A., Wang, N., Fields, A.M., Hua, Z., Toyoda, M., Kawashima, S., Ando, T., Isogai, A., et al. (2010). Collaborative non-self recognition system in S-RNase-based self-incompatibility. Science *330*, 796–799. https://doi.org/10.1126/science.1195243.
38. Haldane, J.B.S. (1933). Two new allelomorphs for heterostylism in Primula. Amer. Nat. *67*, 559–560. https://doi.org/10.1086/280515.
39. Gervais, C.E., Castric, V., Ressayre, A., and Billiard, S. (2011). Origin and diversification dynamics of self-incompatibility haplotypes. Genetics *188*, 625–636. https://doi.org/10.1534/genetics.111.127399.
40. Harkness, A., Goldberg, E.E., and Brandvain, Y. (2021). Diversification or collapse of self-incompatibility haplotypes as a rescue process. Am. Nat. *197*, E89–E109. https://doi.org/10.1086/712424.
41. Taylor, H. (1945). Cyto-taxonomy and phylogeny of the oleaceae. Brittonia *5*, 337–367. https://doi.org/10.2307/2804889.
42. Wallander, E., and Albert, V.A. (2000). Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data. Am. J. Bot. *87*, 1827–1841. https://doi.org/10.2307/2656836.
43. Ganders, F.R. (1979). The biology of heterostyly. N. Z. J. Bot. *17*, 607–635. https://doi.org/10.1080/0028825X.1979.10432574.
44. Matzke, C.M., Hamam, H.J., Henning, P.M., Dougherty, K., Shore, J.S., Neff, M.M., and McCubbin, A.G. (2021). Pistil mating type and morphology are mediated by the Brassinosteroid inactivating activity of the S-locus gene BAHD in heterostylous Turnera species. Int. J. Mol. Sci. *22*, 10603. https://doi.org/10.3390/ijms221910603.
45. Huu, C.N., Plaschil, S., Himmelbach, A., Kappel, C., and Lenhard, M. (2022). Female self-incompatibility type in heterostylous Primula is determined by the brassinosteroid-inactivating cytochrome P450 CYP734A50. Curr. Biol. *32*, 671–676.e5. https://doi.org/10.1016/j.cub.2021.11.046.
46. Fawcett, J.A., Takeshima, R., Kikuchi, S., Yazaki, E., Katsube-Tanaka, T., Dong, Y., Li, M., Hunt, H.V., Jones, M.K., Lister, D.L., et al. (2023). Genome sequencing reveals the genetic architecture of heterostyly and domestication history of common buckwheat. Nat. Plants *9*, 1236–1251. https://doi.org/10.1038/s41477-023-01474-1.
47. Costa, J., Torices, R., and Barrett, S.C.H. (2019). Evolutionary history of the buildup and breakdown of the heterostylous syndrome in Plumbaginaceae. New Phytol. *224*, 1278–1289. https://doi.org/10.1111/nph.15768.
48. Baker, H.G. (1966). The evolution, functioning and breakdown of heteromorphic incompatibility systems. I. The Plumbaginaceae. Evolution *20*, 349–368. https://doi.org/10.1111/j.1558-5646.1966.tb03371.x.
49. Gibbs, P.E. (1986). Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism? Plant Syst. Evol. *154*, 285–323. https://doi.org/10.1007/BF00990130.
50. Huu, C.N., Kappel, C., Keller, B., Sicard, A., Takebayashi, Y., Breuninger, H., Nowak, M.D., Bäurle, I., Himmelbach, A., Burkart, M., et al. (2016). Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. eLife *5*, e17956. https://doi.org/10.7554/eLife.17956.
51. Henning, P.M., Shore, J.S., and McCubbin, A.G. (2020). Transcriptome and network analyses of heterostyly in *Turnera subulata* provide mechanistic insights: are S-loci a red-light for pistil elongation? Plants (Basel) *9*, 713. https://doi.org/10.3390/plants9060713.
52. Fujii, S., Kubo, K.I., and Takayama, S. (2016). Non-self- and self-recognition models in plant self-incompatibility. Nat. Plants *2*, 16130. https://doi.org/10.1038/nplants.2016.130.
53. Rohner, M., Manzanares, C., Yates, S., Thorogood, D., Copetti, D., Lübberstedt, T., Asp, T., and Studer, B. (2023). Fine-mapping and comparative genomic analysis reveal the gene composition at the S and Z self-incompatibility loci in Grasses. Mol. Biol. Evol. *40*, msac259. https://doi.org/10.1093/molbev/msac259.
54. Broz, A.K., and Bedinger, P.A. (2021). Pollen-pistil interactions as reproductive barriers. Annu. Rev. Plant Biol. *72*, 615–639. https://doi.org/10.1146/annurev-arplant-080620-102159.
55. Kubo, K.I., Paape, T., Hatakeyama, M., Entani, T., Takara, A., Kajihara, K., Tsukahara, M., Shimizu-Inatsugi, R., Shimizu, K.K., and Takayama, S. (2015). Gene duplication and genetic exchange drive the evolution of S-Rnase-based self-incompatibility in Petunia. Nat. Plants *1*, 14005. https://doi.org/10.1038/nplants.2014.5.
56. Tarutani, Y., Shiba, H., Iwano, M., Kakizaki, T., Suzuki, G., Watanabe, M., Isogai, A., and Takayama, S. (2010). Trans-acting small RNA determines dominance relationships in Brassica self-incompatibility. Nature *466*, 983–986. https://doi.org/10.1038/nature09308.
57. 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 a complex small RNA regulatory network. Science *346*, 1200–1205. https://doi.org/10.1126/science.1259442.
58. Burghgraeve, N., Simon, S., Barral, S., Fobis-Loisy, I., Holl, A.C., Ponitzki, C., Schmitt, E., Vekemans, X., and Castric, V. (2020). Base-pairing requirements for small RNA-mediated gene silencing of recessive self-Incompatibility alleles in *Arabidopsis halleri*. Genetics *215*, 653–664. https://doi.org/10.1534/genetics.120.303351.
59. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and

annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

60. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.
61. Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., and Clavijo, B.J. (2017). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics *33*, 574–576. https://doi.org/10.1093/bioinformatics/btw663.
62. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.
63. Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and Bioconductor. Methods Mol. Biol. *1418*, 335–351. https://doi.org/10.1007/978-1-4939-3578-9_16.
64. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
65. Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864. https://doi.org/10.1093/bioinformatics/btr026.
66. Evangelistella, C., Valentini, A., Ludovisi, R., Firrincieli, A., Fabbrini, F., Scalabrin, S., Cattonaro, F., Morgante, M., Mugnozza, G.S., Keurentjes, J.J.B., et al. (2017). De novo assembly, functional annotation, and analysis of the giant reed (*Arundo donax* L.) leaf transcriptome provide tools for the development of a biofuel feedstock. Biotechnol. Biofuels *10*, 138. https://doi.org/10.1186/s13068-017-0828-7.
67. Wang, S., and Gribskov, M. (2017). Comprehensive evaluation of *de novo* transcriptome assembly programs and their effects on differential gene expression analysis. Bioinformatics *33*, 327–333. https://doi.org/10.1093/bioinformatics/btw625.
68. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
69. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M., and Kelly, S. (2016). TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. Genome Res. *26*, 1134–1144. https://doi.org/10.1101/gr.196469.115.
70. Gilbert, D.G. (2019). Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes?. Preprint at BiorXiv. https://doi.org/10.1101/829184.
71. Armero, A., Baudouin, L., Bocs, S., and This, D. (2017). Improving transcriptome *de novo* assembly by using a reference genome of a related species: Translational genomics from oil palm to coconut. PloS one *12*, e0173300. https://doi.org/10.1371/journal.pone.0173300.
72. Bao, E., Jiang, T., and Girke, T. (2013). BRANCH: boosting RNA-Seq assemblies with partial or related genomic sequences. Bioinformatics *29*, 1250–1259. https://doi.org/10.1093/bioinformatics/btt127.
73. Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. Genome Res. *12*, 656–664. https://doi.org/10.1101/gr.229202.
74. Gouzy, J., Carrere, S., and Schiex, T. (2009). FrameDP: sensitive peptide detection on noisy matured sequences. Bioinformatics *25*, 670–671. https://doi.org/10.1093/bioinformatics/btp024.
75. Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinformatics *48*, 4.11.1–4.11.39. https://doi.org/10.1002/0471250953.bi0411s48.
76. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics *7*, 62. https://doi.org/10.1186/1471-2105-7-62.
77. Cruz, F., Julca, I., Gómez-Garrido, J., Loska, D., Marcet-Houben, M., Cano, E., Galán, B., Frias, L., Ribeca, P., Derdak, S., et al. (2016). Genome sequence of the olive tree, *Olea europaea*. GigaScience *5*, 29. https://doi.org/10.1186/s13742-016-0134-5.
78. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.
79. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.
80. Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics *37*, 4572–4574. https://doi.org/10.1093/bioinformatics/btab705.
81. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
82. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis, Version 11. Mol. Biol. Evol. *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120.
83. Han, F., and Zhu, B. (2011). Evolutionary analysis of three gibberellin oxidase genes in rice, Arabidopsis, and soybean. Gene *473*, 23–35. https://doi.org/10.1016/j.gene.2010.10.010.
84. Giacomelli, L., Rota-Stabelli, O., Masuero, D., Acheampong, A.K., Moretto, M., Caputi, L., Vrhovsek, U., and Moser, C. (2013). Gibberellin metabolism in *Vitis vinifera* L. during bloom and fruit-set: functional characterization and evolution of grapevine gibberellin oxidases. J. Exp. Bot. *64*, 4403–4419. https://doi.org/10.1093/jxb/ert251.
85. Chen, S., Wang, X., Zhang, L., Lin, S., Liu, D., Wang, Q., Cai, S., El-Tanbouly, R., Gan, L., Wu, H., and Li, Y. (2016). Identification and characterization of tomato gibberellin 2-oxidases (GA2oxs) and effects of fruit-specific SlGA2ox-S1 overexpression on fruit and seed growth and development. Hortic. Res. *3*, 16059. https://doi.org/10.1038/hortres.2016.59.
86. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. Nucleic Acids Res. *51*, D418–D427. https://doi.org/10.1093/nar/gkac993.
87. Brewbaker, J.L., and Kwack, B.H. (1963). The essential role of calcium ion in pollen germination and pollen tube growth. Am. J. Bot. *50*, 859–865. https://doi.org/10.1002/j.1537-2197.1963.tb06564.x.
88. Ray, D.L., and Johnson, J.C. (2014). Validation of reference genes for gene expression analysis in olive (*Olea europaea*) mesocarp tissue by quantitative real-time RT-PCR. BMC Res. Notes *7*, 304. https://doi.org/10.1186/1756-0500-7-304.
89. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. *29*, e45. https://doi.org/10.1093/nar/29.9.e45.

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Gibberellic acid 3 (GA3)	Duchefa biochemie	G0907.0005CAS N°77-06-5
Critical Commercial Assays		
Kit for high molecular weight DNA isolation	QIAGEN (Qiagen, MD, USA)	Genomic-tips 500/G kit
High Fidelity (HiFi) libraries	Pacific Biosciences, Menlo Park, CA, USA	SMRTbell® Template Prep kit 2.0; PN: 100-938-900)
Genome assembler	HiFiasm assembler	v0.15.556; https://github.com/chhylp123/hifiasm
Optical map	Bionano Genomics	Bionano Prep Plant Tissue DNA Isolation Base Protocol (30068)
Hybrid scaffolding	hybridScaffold pipeline	https://bionano.com/wp-content/uploads/2023/01/30073-Bionano-Solve-Theory-of-Operation-Hybrid-Scaffold.pdf; solve version 3.6
Nanopore sequencing libraries	Oxford Nanopore Technologies	kit SQK-LSK110
Oxford Nanopore sequencing	Oxford Nanopore Technologies	MinION (MIN-101B) device, flowcells FLO-MIN106 Rev D, MinKNOW software version 4.3.12 (https://community.nanoporetech.com/downloads)
RNA isolation for RNA-seq	Sigma, Inc., USA	Spectrum Plant Total RNA kit
RNA-seq libraries	Illumina Inc., USA	TruSeq RNA sample Preparation v2 kit
RNA extraction for qRT-PCR	Macherey-Nagel	NucleoSpin RNA Plus kit
Reverse transcriptase	Thermo Scientific	RevertAid First Strand cDNA Synthesis Kit
Deposited Data		
Genome assemblies and annotations, corrected HiFi PACBIO reads from the S1S1 and S2S2 Phillyrea angustifolia individuals, RNA sequencing reads, the *de novo* transcriptome assembly, nucleotide sequence of the GBS markers	https://www.ncbi.nlm.nih.gov/bioproject/	PRJNA993866; Reviewer access : https://dataview.ncbi.nlm.nih.gov/object/PRJNA993866?reviewer=ca576ikhj9e5p4o136as1u1s87
Raw nanopore sequencing reads for the S1S2 Phillyrea angustifolia individual	https://www.ncbi.nlm.nih.gov/bioproject/	bioproject # PRJNA984176 biosample # SAMN35742618
Olea europeae var. Arbequina genome	https://doi.org/10.1038/s41438-021-00498-y	https://bigd.big.ac.cn/gwh/Assembly/10300/show
Oligonucleotides		
PCR and RT-qPCR primers	See Table S1 for a list of oligonucleotide primers for PCR and RT-qPCR	N/A
Software and Algorithms		
Demultiplexing RNA-seq reads	Illumina	CASAVA 1.8.1
Reads cleaning and filtering	https://doi.org/10.14806/ej.17.1.200	CutAdapt
RNA-seq reads assembly	https://doi.org/10.1093/bioinformatics/btw625	Trinity v 2.5.1; Trans-Abyss v.1.5.5
Genome annotation	https://doi.org/10.1002/0471250953.bi0411s48	MAKER
library of Olea transposable elements	https://doi.org/10.1002/tpg2.20010	http://olivegenome.org/genome_datasets/Olea_europaea.denovo.library.fa.zip
RNA-seq reads aligners	https://doi.org/10.1038/nmeth.3317 https://doi.org/10.1038/nbt.3122	HiSat, StringTie

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
genome alignments	https://doi.org/10.1093/bioinformatics/btab705	Minimap2
Short read alignment onto reference Olea genome	https://doi.org/10.1038/nmeth.1923	Bowtie2

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Pierre Saumitou-Laprade (pierre.saumitou-laprade@univ-lille.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Genome assemblies and annotations, corrected PACBIO reads, RNA sequencing reads, the *de novo P. angustifolia* transcriptome assembly, nucleotide sequence of the GBS markers and raw nanopore reads of the S1S2 individual (13A06) have been deposited on the NCBI database. Accession numbers are listed in the key resources table.
- All original code has been deposited at https://github.com/vincentcastric/oleaceae and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plants were grown under field conditions either at the Plateforme des Terrains d'Expérience du LabEx CeMEB (CEFE, CNRS Montpellier, France) or at the Plateforme Serre, cultures et terrains expérimentaux – Université de Lille, France.

We examined plants or sequence data from the following species: *Phillyrea angustifolia*, *P latifolia*, *Olea europeae*, *Fraxinus excelsior*, *Syringa vulgaris*, *Ligustrum vulgare*, *Chyrsojasminum fruticans*, *Osmanthus fragrans*, *Jasminum sambac*, *Fraxinus floribunda*, *F. paxiana*, *F. bungeana*, *F. angustifolia angustifolia*, *F. angustifolia syriaca*, *F. chinensis*, *F. pennsylvanica*, *F. ornus F. sieboldana*, *F. excelsior*, *F. angustifolia oxycarpa*, *Syringa oblata*.

METHOD DETAILS

Biological material and SI genotype determination

We generated high-quality chromosome-scale genome assemblies of two *P. angustifolia* individuals. The first one (01N25 in Billiard et al.[19]) was a hermaphrodite plant belonging to SI group Ha,[19] with S-locus genotype (ss), as determined by the segregation of SI groups in its progeny. The other one (040A11 in Billiard et al.[19]) was a male individual with genotype (SS), also as determined by the segregation of SI groups in its progeny. A third individual (13A06 in Carré et al.[20]) was sequenced using the Oxford Nanopore Technology. This plant is a male, and was used as pollen donor to generate the controlled progeny to produce the GBS map[20]. According to the segregation of SI groups in its progeny, this plant has a heterozygous genotype (Ss) at the S-locus. The ancestors of these individuals (parents or great parents) originate from a local population in Fabrègues (southern France).

HMW DNA isolation

High molecular weight (HMW) DNA was extracted from frozen leaves using QIAGEN Genomic-tips 500/G kit (Qiagen, MD, USA), following the tissue protocol extraction. Briefly, 2g of young leaves material were grounded in liquid nitrogen with mortar and pestle. After 3h of lysis and one centrifugation step, the DNA was immobilized on a column. After several washing steps, DNA was eluted from the column, then desalted and concentrated by isopropyl alcohol precipitation. A final wash in 70% ethanol was performed and the DNA was resuspended in EB buffer. DNA quantity and quality were assessed using NanoDrop and Qubit (Thermo Fisher Scientific, MA, USA). DNA integrity was also assessed using the Agilent FP-1002 Genomic DNA 165 kb on the Femto Pulse system (Agilent, CA, USA).

PACBIO sequencing and assembly

High Fidelity (HiFi) libraries were constructed using the SMRTbell® Template Prep kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) according to PacBio recommendations (SMRTbell® express template prep kit 2.0 - PN: 100-938-900). HMW DNA samples were first

purified with 1× Agencourt AMPure XP beads (Beckman Coulter, Inc, CA USA), and sheared with Megaruptor 3 (Diagenode, Liège, BELGIUM) at an average size of 20 kb. After end repair, A-tailing and ligation of SMRTbell adapter, the library was size-selected on the BluePippin System (Sage Science, MA,USA) at range sizes of 10-50kb. The size and concentration of libraries were assessed using the Agilent FP-1002 Genomic DNA 165 kb on the Femto Pulse system and the Qubit dsDNA HS reagents Assay kit.

Sequencing primer v5 and Sequel® II DNA Polymerase 2.2 were annealed and bound, respectively, to the SMRTbell libraries. Each library was loaded on two SMRTcell 8M at an on-plate concentration of 90pM. Sequencing was performed on the Sequel® II system at the Gentyane Genomic Platform (INRAE Clermont-Ferrand, France) with Sequel® II Sequencing kit 3.0, a run movie time of 30 hours with an Adaptive Loading target (P1 + P2) at 0.75. HiFi reads were produced with the PacBio Sequel II system on four SMRTCells and were assembled using the HiFiasm assembler (v0.15.5[31]; https://github.com/chhylp123/hifiasm).

To assess the completeness and quality of the assemblies, we used the Benchmarking Universal Single-Copy Orthologs (BUSCO) pipeline with the viridiplantae database.[59] We obtained a 99.3% complete BUSCO score on the primary assembly. In addition, we performed k-mer analysis to quality control the dataset using Jellyfish tool[60] and the assemblies using module "comp" of the k-mer Analysis Toolkit.[61] All the metrics are reported in Data S1B.

Optical map

To achieve a reference-level genome assembly for the first individual (ss), we combined the 40X HiFi PACBIO sequences obtained above with 588X optical mapping datasets. Briefly, ultra HMW DNA (uHWM DNA) was purified from 1g of fresh dark treated very young leaves according to the Bionano Prep Plant Tissue DNA Isolation Base Protocol (30068 - Bionano Genomics) with the following specifications and modifications. Briefly, the leaves were fixed in a buffer containing formaldehyde. After three washes, leaves were cut in 2 mm pieces and disrupted with a rotor stator in homogenization buffer containing spermine, spermidine and beta-mercaptoethanol. Nuclei were washed, purified using a density gradient and then embedded in agarose plugs. After overnight proteinase K digestion (Qiagen) in the presence of lysis buffer and a one hour treatment with RNAse A (Qiagen), plugs were washed and solubilized with 2 μL of 0.5 U/μL AGARase enzyme (ThermoFisher Scientific). A dialysis step was performed in TE Buffer (ThermoFisher Scientific) to purify DNA from remaining residues. The DNA samples were quantified by using the Qubit dsDNA BR Assay (Invitrogen). The presence of megabase-sized DNA molecules was visualized by pulsed field gel electrophoresis (PFGE). Labelling and staining of the uHMW DNA were performed according to the Direct Label and Stain (DLS) protocol (30206 - Bionano Genomics). Briefly, labelling was performed by incubating 750 ng genomic DNA with 1× DLE-1 Enzyme for 2 hours in the presence of 1× DL-Green and 1× DLE-1 Buffer. Following proteinase K digestion and DL-Green clean-up, the DNA backbone was stained by mixing the labelled DNA with DNA Stain solution in the presence of 1× Flow Buffer and 1× DTT, and incubating overnight at room temperature. The DLS DNA concentration was measured with the Qubit dsDNA HS Assay (Invitrogen, Carlsbad, CA, USA). Labelled and stained DNA was loaded on 1 Saphyr chip and was run on the BNG Saphyr System according to the Saphyr System User Guide. Digitalized labelled DNA molecules were assembled to optical maps using the BNG Access software (solve version 3.5). The molecule N50 was 250kb.

Scaffolding contigs with the optical and genetic maps

A hybrid scaffolding was then performed between the sequence assembly and the optical genome map with the hybridScaffold pipeline (https://bionano.com/wp-content/uploads/2023/01/30073-Bionano-Solve-Theory-of-Operation-Hybrid-Scaffold.pdf; solve version 3.6).

Finally, the genetic map of Carré et al.[20] was used to finalize scaffolding. This map consists of an overall total of 15,814 SNPs contained in 10,388 GBS fragments (some GBS fragments contained more than one SNP) genotyped in 196 offspring. We used BLAST[62] to align the sequence of these GBS fragments onto the contigs obtained from the assemblies. These alignments were used to organize contigs containing fragments belonging to the same linkage group and achieve chromosome-scale assemblies. The position of these alignments along the pseudo-chromosomes were then displayed using GViz.[63] The SI phenotype ([Ha] *vs.* [Hb]) was mapped at position 53.619 cM on linkage group 18 by Carré et al.,[20] and we delimited the chromosomal interval containing the S-locus based on the immediately flanking upstream and downstream markers mapped at position 53.126cM (one GBS sequence, one SNP) and 53.864cM (two GBS sequences, three SNPs), respectively.

HMW DNA isolation and Oxford Nanopore sequencing

HMW DNA was extracted from leaves of a third *P. angustifolia* individual (13A06), whose S-locus genotype was (Ss). HMW DNA was extracted using the Carlson lysis buffer followed by purification using the QIAGEN Genomic-tip 500/G (Qiagen, MD, USA) with slight modification. 1g of fresh leaves was grinded in a mortar to a fine powder in presence of liquid nitrogen. The powdered material was dispensed into two 50 mL centrifuge tubes containing 20 mL of pre-warmed (65°C) lysis buffer. After one hour of lysis, 20 mL of chloroform has been added to each tube, followed by vortexing and centrifugation. The supernatant was collected, mixed with 0.7x volumes of isopropanol and pelleted by centrifugation. The pellets were resuspended with 19 mL of G2 buffer, from the QIAGEN Blood and Cell Culture DNA Maxi Kit t (Qiagen, MD, USA). Purification was then performed using QIAGEN Genomic-tips 500/G based on the indications of the protocol. Then DNA was precipitated by isopropanol, washed in 70% ethanol and re-suspended in TB buffer. To enhance recovery of long DNA fragments, 9 μg of DNA were processed using the Short Read Eliminator Kit XL (Circulomics, Baltimore, MD, Cat #SS-100-101-01) according to the supplier's instructions. DNA quantity and quality were assessed using NanoDrop and Qubit (Thermo Fisher Scientific, MA, USA) before and after size selection.

Please cite this article in press as: Castric et al., The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.03.047

DNA libraries were prepared using the Oxford Nanopore Technologies kit SQK-LSK110 with the following modifications. DNA repair and end-prep (New England BioLabs, Ipswich, MA, Cat #E7546 and Cat #M6630) were performed with 3 μg DNA, in two separate tubes (1.5 μg in each tube) in a total reaction volume of 60 μl each, incubated at 20°C for 5 minutes, and 60°C for 5 minutes. The two DNA repair and end-prep reactions were combined and cleaned with 120 μl of Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, Cat #A63880) with an incubation time ranging from ten to 20 minutes and an elution time of five to ten minutes. Ligation was performed at room temperature for 10 minutes. The ligation reaction was cleaned using Agencourt AMPure XP beads with an incubation time of ten to 20 minutes and an elution time of 15 to 25 minutes at room temperature or 37°C.

Sequencing was performed with the Oxford Nanopore Technologies (Oxford, UK) MinION (MIN-101B) device with a total of 13 FLO-MIN106 Rev D flow cells. The MinKNOW software version 4.3.12 (https://community.nanoporetech.com/downloads) was used to collect data. The running parameters were set to default and the fast basecalling model was used to generate real-time run statistics. After each run, a new basecalling was performed by using Guppy v5.0.13 with the super accurate configuration model on an Intel core I9 workstation equipped with an Nvidia RTX 2080ti GPU. Adapters were trimmed out with Porechop software (https://github.com/rrwick/Porechop).

Preparation of RNA samples

We performed two replicate RNA-seq experiments on 14 genotypes (8 (ss) and 6 (Ss)) from the Fabrègues population. For each of these individuals, the first experiment was based on five mixed-stages whole flower buds, and the second was based on ten dissected pistils that had been pollinated *in vitro* using pollen from a [Hb] individual. Samples were ground in liquid nitrogen and total cellular RNA was extracted using a Spectrum Plant Total RNA kit (Sigma, Inc., USA) with a DNAse treatment. RNA concentration was first measured using a NanoDrop ND-1000 Spectrophotometer then with the Quant-iT™ RiboGreen® (Invitrogen, USA) protocol on a Tecan Genius spectrofluorimeter. RNA quality was assessed by running 1 μL of each RNA sample on RNA 6000 Pico chip on a Bioanalyzer 2100 (Agilent Technologies, Inc., USA). Samples with an RNA Integrity Number (RIN) value greater than eight were deemed acceptable.

RNA-seq library construction and sequencing

The TruSeq RNA sample Preparation v2 kit (Illumina Inc., USA) was used according to the manufacturer's protocol with the following modifications. In brief, poly-A containing mRNA molecules were purified from 1 ug total RNA using poly-T oligo attached magnetic beads. The purified mRNA was fragmented by addition of the fragmentation buffer and was heated at 94°C in a thermocycler for four minutes to yield library fragments of 250-500 bp. First-strand cDNA was synthesized using random primers to eliminate the general bias towards 3'end of the transcripts. Second strand cDNA synthesis, end repair, A-tailing, and adapter ligation was done in accordance with the manufacturer's protocols. Purified cDNA templates were enriched by 15 cycles of PCR for 10s at 98°C, 30s at 65°C and 30s at 72°C using PE1.0 and PE2.0 primers and with Phusion DNA polymerase (NEB, USA). Each indexed cDNA library was verified and quantified using a DNA 100 Chip on a Bioanalyzer 2100, then pooled in equimolar amounts by sets of ten samples. The final library was quantified by real-time PCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) adjusted to 10 nM in water and provided to the Get-PlaGe core facility (GenoToul platform, INRA, Toulouse, France http://www.genotoul.fr) for sequencing.

Final pooled cDNA libraries were sequenced using the Illumina mRNA-Seq, paired-end protocol on a HiSeq2000 sequencer, for 2 x 100 cycles. Libraries were diluted to 2 nM with NaOH and 2.5 μL transferred into 497.5 μL HT1 to give a final concentration of 10 pM. 120 μL were then transferred into a 200 μL strip tube and placed on ice before loading onto the cBot. Mixed libraries from ten individual indexed libraries were run on a single lane. The flow cell was clustered using TruSeq PE Cluster Kit v3, following the Illumina PE_Amp_Lin_Block_V8.0 protocol. Following the clustering procedure, the flow cell was loaded onto the Illumina HiSeq 2000 instrument. The sequencing chemistry used was v3 (FC-401-3001, TruSeq SBS Kit) with 2x100 cycles, paired-end, indexed protocol. Image analyses and basecalling were performed using the HiSeq Control Software (HCS 1.5.15) and Real-Time Analysis component (RTA 1.13.48).

A *de novo* transcriptome assembly for *P. angustifolia*

Demultiplexing was performed using CASAVA 1.8.1 (Illumina) to produce paired sequence files containing reads for each sample in Illumina FASTQ format. Reads were cleaned and filtered with CutAdapt[64] (option: *–overlap=30*) and Prinseq[65] (option: *-min_len 80 -trim_tail_left 5 -trim_tail_right 5 -lc_method entropy -lc_threshold 70*). The *de novo* assembly protocol was based on Evangistella et al.[66] Following Wang and Gribskov,[67] we first pre-assembled the reads with Trinity v 2.5.1 and Trans-Abyss v.1.5.5 using the default *K-mer*, *i.e. K-mer = 25* for Trinity (min length=200bp) and *K-mer = 32* for Trans-Abyss, (min length=100bp). Contigs of the pre-assemblies with nucleotide sequence identity above 0.98 were merged using the CD-HIT-EST tool[68] and verified using Transrate v1.0.[69] These individual *de novo* assemblies were then merged again with Trans-Abyss and we used the EvidentialGene tr2aacds pipeline[70] to reduce this assembly into a first set of non-redundant unitigs. Following Armero et al.,[71] we used BRANCH[72] to improve unitig sequences by aligning the RNA sequencing reads onto the unitigs with a modified version of BLAT.[73] This latter step identifies novel unitigs, extends incomplete unitigs and joins fragmented ones.[72] Finally, we sequentially used FrameDP v1.2.2[74] and the scripts tr2aacds.pl of the EvidentialGene pipeline[70] and main.pl[71] to remove redundant and/or chimeric unitigs based on their translated polypeptide sequences.

Annotation of protein-coding genes and transposable elements

We used MAKER[75] to predict protein-coding genes on the two primary assemblies. Briefly, MAKER starts from *ab initio* gene prediction by Augustus v. 3.3.3.[76] trained on the Arabidopsis genome and then searches for a series of additional evidences using the set of predicted proteins and CDS from the olive tree genome[77] as well as unitig sequences from the *de novo P. angustifolia* transcriptome described above. We then aligned raw RNA-seq reads from bud and pistil tissues obtained from fourteen *P. angustifolia* individuals with known SI phenotypes (eight Ha individuals and six Hb individuals, assumed to carry (ss) and (Ss) genotypes, respectively, Data S1C) on the genome using the splice-aware aligner HiSat2,[78] and used Stringtie[79] to refine the prediction of transcripts using information from the RNA-seq reads split across intron-exon boundaries. We retained Augustus gene models for which at least one additional evidence was present. We used RepeatMasker to identify TEs based on the olive tree genome[26]; http://olivegenome.org/genome_datasets/Olea_europaea.denovo.library.fa.zip), and we eliminated gene predictions overlapping with TE annotations.

Genome alignment and sequence comparison

To identify major chromosomal rearrangements, we used minimap2[80] to align the alternative assembly of each individual to their respective primary assembly (using the -asm5 option), and to align them to one another. The resulting paiwise alignments were displayed using the *pafr* library in R (https://github.com/dwinter/pafr). We retrieved the sequence interval between the GBS markers aligned on the 040A11 hap1 assembly and used minimap2 with default parameters to align it to the 01N25 hap2 assembly, and displayed the alignment using *pafr*.

To identify nucleotide sequences specific to the s or S chromosomes, we split the chromosomal intervals between the non-recombining GBS markers linked to the S-locus into consecutive stretches of 300bp and blasted them onto the rest of the genome. We retained only those with no hit above 80% identity and concatenated all overlapping fragments.

We noted the presence of an inversion of about 5Mb in one of the haplotypes of the primary assembly of the (ss) individual, approximately corresponding to the S-locus region (Figure S2A). To evaluate the possibility that the inversion itself could be related to the SI determinant, we computed the rate of synonymous divergence (K_S) between the protein-coding genes in the s and S haplotypes along chromosome 18. We could not detect any elevation of K_S for genes inside the inversion as compared to the rest of the chromosome (Figure S2B), suggesting that the inversion is indeed very recent and probably unrelated to SI determination given the ancestrality of the two allelic specificities.[16] In fact, comparing the two s haplotypes of the (ss) assembly revealed that only one of these (hap1) contained the reverse orientation (Figure S2A). The second haplotype (hap2) had the same orientation as that in the two S haplotypes of the (SS) assembly, where this ‘‘standard’’ orientation was homozygous. Re-mapping of raw HiFi PACBIO and BioNano signals onto the assembly of the (ss) individual, and inspection of the inversion breakpoints confirmed that the inversion was truly heterozygous in the sequenced individual, and did not correspond to an assembly error (Figure S2C). Finally, to further investigate the presence of this inversion, we produced Nanopore reads from high molecular weight (HMW) genomic DNA of a third *P. angustifolia* individual from the same population (Fabrègues) with a (Ss) genotype. By mapping the raw unassembled reads on the breakpoints of the inversion, we confirmed that the inversion was absent from the (Ss) individual, and is therefore specific to only one of the two chromosomes of the sequenced (ss) individual (Figure S2B). Hence, this inversion appears to be very recent, since it is not fixed among *s* chromosomes, and thus should be independent from the SI determinants that presumably have a single origin, and have remained remarkably stable over extended evolutionary times.[16]

Mapping short Illumina reads from *Olea europaea* accessions

To determine whether the indel we identified in P. angustifolia was also segregating in *O. europaea*, we then used bowtie2[81] to map publically available short Illumina reads from eight *O. europaea* accessions whose SI phenotype had been determined previously (Data S1E) on the complete *O. europaea var*. Arbequina genome[24] (https://bigd.big.ac.cn/gwh/Assembly/10300/show). We used samtools to vizualise and quantify variation of the depth of aligned sequences along chromosome 18 with a quality threshold of Q30.

Sequence comparison across distant Oleaceae species, specific primer design, PCR protocol and association study across Phillyrea and Olea accessions

We retrieved assembled genomes from 20 Oleacea species available from the literature (Data S1G). We used blast with default parameters to search for *PaGA2ox-S* orthologs. Based on the aligned sequences of the first exon of the GA2ox-S orthologs, we designed PCR primers and optimized amplification conditions to track the presence of *GA2ox-S* in a series of samples with known SI phenotype (Data S1H).

Phylogeny of GA2ox-S proteins

To place PaGA2ox-S in the phylogenetic tree of GA2 oxidase enzymes, we collected previously published protein sequences of GA2ox-S enzymes of different flowering plant species for which the enzyme class was described. The phylogenetic tree was inferred by the Neighbor-Joining method, using MEGA11.[82] The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. This analysis involved a total of 41 amino acid sequences from *Arabidopsis thaliana*,[32] Rice (*Oryza sativa*[83]), Grapevine (*Vitis vinifera*[84]), Tomato (*Solanum lycopersicum*[85]), Peach (*Prunus persica*[32]) and Barley (*Hordeum vulgare*[30]). Domains within the PaGA2ox-S protein were identified by InterPro.[86]

RT-qPCR

To study the expression of GA2ox-S in different *P. angustifolia* tissues we manually dissected anthers and stigmas from closed buds one day before anthesis. We also isolated non-dissected immature buds two weeks before anthesis, as well as leaves. We collected mature pollen on dehiscent anthers and stored it at -80°C. For germination, we incubated 5 mg of pollen in 2 ml of Brewbacker and Kwack[87] solution containing 11% sucrose solution during 3 hours at 28°C. Proper germination was verified by microscopic observation. We then briefly centrifuged to gently pellet the pollen grains and remove the supernatant. All samples were flash-frozen in liquid nitrogen immediately upon collection and RNA was extracted using the NucleoSpin RNA Plus kit (Macherey-Nagel) following the supplier's instructions. The RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) was used to synthesize cDNA and qPCR was performed using the iTaq Universal SYBR Green Supermix (BioRad) on a Lightcycler 480 instrument (Roche). Primer sequences and amplification conditions are detailed in Data S1K. We selected *PaPP2A* (Protein Phosphatase 2A) to be used as a reference gene, based on its previous validation for use in expression analysis in *Olea europaea*.[88] Relative *PaGA2ox-S* transcript abundance was estimated using the Pfaffl method.[89] Specificity of the qPCR amplifications was confirmed by examination of the obtained melting curves.

GA2 supplementation experiment

In order to test the effect of GA3 on SI specificities, we took advantage of the relatively large and flexible inflorescences of *Ligustrum vulgare*. We selected three [Ha] and two [Hb] individuals whose SI phenotype had been characterized[21] and cultivated them in an insect-proof greenhouse to avoid pollen contamination of the opening buds. On each individual plant, we selected and labeled four inflorescences - one for each of the treatments described below. Each treatment consisted of immersing the complete inflorescence for a few seconds in a Falcon tube containing 50 mL of solution. We applied four treatments (1) a control with no GA3, (2) a 20μM GA3 solution (0.1×), (3) a 200μM GA3 solution (1×) and (4) a 2mM GA3 solution (10×). Each treatment was applied twice a week from the end of April until the stage "white bud" at the end of May. When an inflorescence contained at least three open flowers, it was immediately collected and transferred to the laboratory for the stigma test. Open flowers were eliminated, and the inflorescence containing only closed buds was kept overnight at 20°C under protection against pollen contamination. After 16 hours the newly opened flowers were collected, emasculated by removing the corollae and planted in agar medium. The corollae containing the two non-dehisced anthers were placed under dry air conditions (for two to three hours) until dehiscence. Then, receptive emasculated flowers in agar and the dehisced anthers were used to perform the different crossing schemes presented in Figure 4. Each cross was performed with five technical replicates.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses are described in the STAR Methods, methods details section, in the main text and Figure/Table legends. Scripts for all analyses are available as detailed in the data and code availability statement