A hemizygous supergene controls homomorphic and heteromorphic self-incompatibility systems in Oleaceae

Pauline Raimondeau, Sayam Ksouda, William Marande, Anne-Laure Fuchs, Hervé Gryta, Anthony Theron, Aurore Puyoou, Julia Dupin, Pierre-Olivier Cheptou, Sonia Vautrin, et al.

A hemizygous supergene controls homomorphic and heteromorphic self-incompatibility systems in Oleaceae

Pauline Raimondeau[1,2], Sayam Ksouda[1], William Marande[3], Anne-Laure Fuchs[1], Hervé Gryta[1], Anthony Théron[3], Aurore Puyoou[1], Julia Dupin[1], Pierre-Olivier Cheptou[4], Sonia Vautrin[3], Sophie Valière[5], Sophie Manzi[1], Djamel Baali-Cherif[6], Jérôme Chave[1], Pascal-Antoine Christin[7], Guillaume Besnard[1*]

[1] CNRS, Université Paul Sabatier, IRD, UMR 5300, CRBE (Centre de Recherche sur la Biodiversité et l'Environnement), 118 Route de Narbonne, 31062 Toulouse, France; [2] Yale Institute of Biospheric Studies, New Haven, CT 06520, USA; [3] INRAE, Centre National de Ressources Génomiques Végétales, 31326 Castanet-Tolosan, France; [4] CEFE (Centre d'Ecologie Fonctionnelle et Evolutive), UMR 5175, CNRS, Université de Montpellier, Université Paul Valéry, EPHE, IRD, 34293 Montpellier, France; [5] INRAE, US 1426, GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France; [6] Laboratoire de Recherche sur les Zones Arides, USTHB/ENSA, 16000 Alger, Algeria; [7] Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield S10 2TN, UK

* Lead contact: guillaume.besnard@univ-tlse3.fr; Twitter handles: @PaulineRaim

Summary Self-incompatibility (SI) has evolved independently multiple times and prevents self-fertilization in hermaphrodite angiosperms. Several groups of Oleaceae such as jasmines exhibit distylous flowers, with two compatibility groups each associated to a specific floral morph[1]. Other Oleaceae species, in the olive tribe, have two compatibility groups without associated morphological variation[2–5]. The genetic basis of both homomorphic and dimorphic SI systems of Oleaceae are unknown. By comparing genomic sequences of three olive subspecies (*Olea europaea*) belonging to the two compatibility groups, we first locate the genetic determinants of self-incompatibility within a 700-kb hemizygous region present only in one compatibility group. We then demonstrate that the homologous hemizygous region also control distyly in jasmine. Phylogenetic analyses support a common origin of both systems following a segmental genomic duplication in a common ancestor. Examination of the gene content of the hemizygous region in different jasmine and olive species suggests that the mechanisms determining compatibility groups and floral phenotypes (whether homomorphic or dimorphic) in Oleaceae relies on the presence/absence of two genes involved in gibberellins and brassinosteroids regulation

Results and discussion

A hemizygous region determines self-incompatibility in olive

Within the Oleaceae family, several species in the olive tribe (Oleeae) have two compatibility groups without associated flower morphological variation[2–5]. We identified the locus responsible for these compatibility groups (*S*-locus) in olive based on the observation that self-incompatibility recognition specificities are shared among Oleeae lineages that have been diverging for at least 40 million years[5,6]. Self-incompatibility systems classically involve several genes controlling pollen and pistil determinants. If this is also the case for Oleaceae, conservation over a long evolutionary time of the SI system should imply a co-inheritance of these genes, since any recombination between them would break the self-incompatibility system. Tight-linkage of allelic combinations in a chromosomal region, referred to as a supergene, is a practical solution to achieve this. Following this hypothesis, we tracked supergene genomic footprints (Figure S1) through genome scans using RAD-seq data for olive trees phenotyped as belonging to the two compatibility groups [defined as the heterozygote *Ss* (G1) or the homozygote recessive *ss* (G2)][7,8] from three genetically distinct olive subspecies (*Olea europaea* subspp. *europaea*, *africana*, and *laperrinei*; Data S1A; Figure 1A). Coverage ratio analyses of the two compatibility groups revealed a single outlier region on chromosome 18, in which the *Ss* samples had higher coverage than the *ss* samples in all three subspecies (Wilcoxon-test p-value < 0.05; Figure 1B). Direct comparison of the chromosome 18 contigs in haplotype-resolved assemblies we generated for the Saharan olive (*Olea europaea* subsp.

laperrinei) confirmed the existence of a 700-kb region specific to *Ss* individuals (Figures 1C and S2; Data S1B and S1C). The *S*-specific sequence, always found at hemizygous state, is enriched in transposable elements compared to neighboring windows, suggesting that this region recombine rarely[9], which is guaranteed by hemizygosity (Wilcoxon test; p-value < 0.05; Figure 1C). The association between the hemizygous region and compatibility groups was supported by genome-wide association analyses (Figures 1D and 1E) as well as by the strict correlation between markers at the hemizygous region and the self-incompatibility phenotype in a larger sampling of phenotyped trees, from three diploid subspecies and hybrids (Data S1A and 1D). Our results therefore demonstrate that a hemizygous region on chromosome 18 determines self-incompatibility in olive.

The genetic controls of distyly and homomorphic self-incompatibility overlap

The 700-kb indel in olive contains three predicted genes - *Oe46719.t1, Oe46724.t1* and *Oe46727.t1* - with contrasted expression patterns (Figure 1C). Expression of *Oe46719.t1* (hereafter referred to as *BZR1-S*) was not detected in *O. e. europaea* inflorescences. In the closely related *Olea welwitschii,* we found only a 370-bp pseudogenic segment of its first exon, which contains several frame shifts. We found some transcripts for *Oe46724.t1* in *O. e. europaea*, but detected no orthologs of this gene in other Oleaceae*,* suggesting that it cannot be essential for this conserved function in distinct Oleaceae lineages. *Oe46727.t1* (referred to as *GA2ox-S*), however, is both expressed in olive inflorescences and detected within the homologous region in all self-incompatible Oleaceae investigated (Figure 2A). As *GA2ox-S* is the only gene systematically associated with self-incompatibility types, it appears to be sufficient for self-incompatibility responses in olive. The *GA2ox* gene family encodes oxidases that down-regulate gibberellins production[10]. While the precise self-incompatibility mechanism is unclear, we hypothesize that a mismatch between the *Ss* and *ss* reproductive tissue environments may be involved and that alleles of this single gene cause both male and female incompatibility reactions (see also the companion paper by Castric et al.[11] for an exploration of the role of gibberellins in self-incompatibility reactions in the olive tribe). Hormone regulation has frequently been found to be involved in the control of the floral dimorphism in distylous species (whose mechanism determining incompatibility reactions remains elusive) but, to our knowledge, this is the first time that gibberellins are suspected of being involved in self-incompatibility reactions[12-17] (Table S1).

In contrast to the olive tribe (Oleeae), in which self-incompatibility is not associated with morphological variation, other Oleaceae tribes, such as Jasmineae, exhibit distylous flowers, with two compatibility groups each associated with a different floral morph[1]. Distyly is controlled by a supergene in which several loci control different aspects of the flower dimorphism and incompatibility phenotype. In the distylous species so far studied, one of these genes controls the style length and its incompatibility, another controls the anther positioning, and another the pollen incompatibility (Table S1). The *S*-locus has invariably been found to be a hemizygous supergene specific to short-styled individuals[12-17] (Table S1). Genes controlling the short-styled morph are thus present in a haploid region, whereas the long-styled morph lacks them entirely, which guarantees the co-inheritance of the short-styled alleles at all the loci. The hemizygous architecture of the olive *S*-locus thus parallels that in distyly. To test whether homomorphic self-incompatible and distylous self-incompatible systems share the same genetic basis in Oleaceae, we compared gene expression between floral morphs of the distylous Mediterranean jasmine (*Chrysojasminum fruticans*; Jasmineae) using entire floral buds and identified 33 genes with morph-specific expression patterns (Figure 2B; Data S2). This included two *GA2ox* homologs. One of them, *Cf75129.t1*, was expressed only in long-styled individuals, while the other (*Cf41540.t1*) was exclusively expressed in short-styled individuals and corresponds to *GA2ox-S*, according to our phylogenetic reconstruction of the *GA2ox* gene family (Figure 3A). This gene, in our *C. fruticans* genome assembly, is physically close to three other differentially expressed genes (*Cf41478.t1*,

Cf41523.t1, and *Cf41549.t1*). In *C. fruticans*, these three genes were specifically expressed in short-styled buds, suggesting that they might all be part of a *S*-locus supergene. However, only *Cf41523.t1* and *Cf41540.t1* are hemizygous. PCR tests using 179 phenotyped jasmines indeed confirmed a strict association between the short-styled morph and presence of *Cf41523.t1* and *Cf41540.t1* (Data S1E). *Cf41523.t1* is an homolog of the non-expressed *BZR1-S* gene detected within the olive hemizygous *S*-locus. *BZR1-S* is also detected in short-styled individuals of other distylous Oleaceae (Figures 2A and 3A). It encodes a transcription factor (BRASSINAZOLE-RESISTANT 1) that regulates brassinosteroids expression[18], a function known to be involved in the control of female morphology in other distylous systems[13,14,16] (Table S1).

The present comparative approach therefore supports the importance of hormone regulation in the genetic control of self-incompatibility phenotypes in Oleaceae. The *GA2ox* family is particularly interesting as we identified two differentially expressed homologs in the distylous jasmine. The one that is located within the hemizygous region (*Cf41523.t1*) is orthologous to the gene we identified as controlling self-incompatibility response in olive (*GA2ox-S*). We postulate that it has the same function in jasmine as this is consistent with the absence of *GA2ox-S* in the self-compatible *Chrysojasminum bignoniaceum* (Jasmineae), and the absence of expression of the second *GA2ox* homolog (*Cf75129.t1*) in floral buds in this taxon. We speculate that the second homolog plays a complementary role in determining compatibility, for example, by being expressed in different floral parts of the two incompatility groups (e.g., stigmata and pollen).

The *BZR1-S* gene is either not expressed, pseudogenized or is even absent from the *S*-locus of non-distylous Oleeae lineages (including *Phillyrea*[11]). In contrast, our floral bud RNAseq data and genome assembly of *C. bignoniaceum*, a long-homostylous species (with styles of similar length to those of the long-styled individuals of the distylous *C. fruticans*, but with long anthers), show that *BZR1-S* is expressed, and that it is in a genomic region collinear with the *S*-locus supergene of *C. fruticans.* However, it is not hemizygous. The presence of a sequence homologous to the *S*-specific region implies that this long-homostylous state derives from a short-style individual (*Ss* genotype). Consequently, the phenotypic difference between short-styled *C. fruticans* and long-homostylous self-compatible *C. bignoniaceum* may be due to the absence of *GA2ox-S*. This single gene would thus control both female self-incompatibility reactions and style-length, as reported for other distylous systems where a single gene accomplishes both functions[14-17] (Table S1). *BZR1-S* would then determine anther length. Indeed, when it is expressed, in both *C. fruticans* and *C. bignoniaceum,* anthers are found in high positions (Figure 2A).

We cannot rule out the involvement of other genetic elements in the *S*-locus, or in other genomic regions, such as genes involved in downstream signaling cascades or modifier loci that have become fixed in the homostyled species. GA2ox-S and BZR1-S are both parts of complex transcriptional networks[11,18], and different hormone levels caused by the two *S*-locus genes may have downstream effects on metabolite levels or cell-wall properties that might affect pollen tube germination and growth. The specific expression of a *GA2ox* paralog in long-styled jasmine suggests that the *S*-locus copy may act as a dominant-negative mutation that interferes with the expression of the other copy[19]. Although further functional studies are required, the evidence presented here strongly supports the conclusion that both homomorphic and heteromorphic self-incompatibility systems are controlled by a hemizygous supergene.

Heteromorphic and homomorphic self-incompatibility systems have a common origin in Oleaceae

Within the Lamiales order, Oleaceae and its sister family Carlemanniaceae are sister to all other Lamiales (hereafter Core Lamiales)[20]. Phylogenetic analyses of the two gene families detected in the hemizygous *S*-locus of olives and jasmines (i.e., *GA2ox* and *BZR1*) support the orthology of their *S*-copies (Figure 3). For both genes, we detected a *S*-locus copy of one or both genes only in

short-styled individuals or in one of the homomorphic compatibility groups (Figure 3; Data S1F). For species for which we only had short-read data covering the *S*-locus, the sequencing depth was consistently lower than that of paralogs outside the *S*-locus (Figure 3), supporting the result that hemizygosity is shared by both systems.

Across the Oleaceae and other Lamiales, the most-closely related paralogs of both *S*-locus genes (*GA2ox-I-1* for *GA2ox-S*) are invariably located on a single chromosome (Figure 3). In addition, pairs of genes from the *S*-locus and their closest paralogs have similar synonymous divergence levels for both *GA2ox* and *BZR1* (Figure 3C). Furthermore, phylogeny of the *BZR1* gene family indicates that the duplication of the *S*-locus copy occurred prior to the divergence of Oleaceae from its sister family Carlemanniaceae, but after these families diverged from Core Lamiales (Figure 3B). This period coincides with an inferred whole-genome duplication event in their common ancestors (62 and 95 Mya)[20]. Whether or not there is a link with this polyploidization event, we conclude that the supergene architecture of self-incompatibility loci in Oleaceae originated from an ancestral duplication of the DNA segment encompassing these two genes.

Two models have been proposed for the evolution of distyly. In the first model, the pollen incompatibility evolved first, before morphological differentiation[21]. The second model postulates that floral dimorphism appeared first[22]. In either model, the distyly supergene evolved in several steps and dating the duplications of the different genes provide information about the sequence of events in the evolution of distyly[23,24]. In Oleaceae, the origin of the *S*-locus via a unique duplication event makes it even more complicated to determine whether floral dimorphism or pollen incompatibility evolved first. Distyly shows some lability in Jasmineae, where the co-occurrence of stigma-height dimorphism and distyly, has been interpreted as evidence for the evolution of distyly through the "style-length-variation-first" model[25]. Conservation of the recognition specificities has been studied only between homomorphic Oleaceae species, not between different distylous Oleaceae lineages (which are mostly distantly related), nor between any homomorphic and distylous lineages in the family. It is consequently possible that distyly evolved independently more than once in the Oleaceae, as it has been hypothesized in the case of Rubiaceae[26]. The ancestral duplication of *S*-locus genes may have provided genetic precursors facilitating multiple independent evolutions of self-incompatibility systems.

Alternatively, the homology of the hemizygous region in different distylous lineages would also be compatible with a single origin of distyly through the establishment of a hemizygous incompatibility locus in the common ancestor of Carlemanniaceae and Oleaceae. Carlemanniaceae includes two genera (five species), one of which (*Silvianthus*) is reported as distylous[27]. Instead of representing an intermediate stage on the way to distyly, non-distylous Jasmineae, such as the self-compatible species *C. bignoniaceum* (long homostylous) could have evolved by losses of distyly and self-incompatibility. This scenario would also imply that in non-distylous self-incompatible Oleaceae, such as the olive, *S*-loci represent ancient degenerated distyly supergenes. Independently of the exact scenario, our analyses demonstrate a common origin of homomorphic and heteromorphic incompatibility systems using a supergene generated by an ancestral segmental duplication.

While the detailed mechanisms controlling the observed phenotypes remain speculative, we identified two strong candidate genes involved in the regulation of brassinosteroids and gibberellins. These genes represent excellent candidates for further functional studies. In contrast to other investigated systems so far, molecular incompatibility and morphological determinants appear to be separable despite genetic overlap at the *S*-locus. Oleaceae self-incompatibility systems thus offer a compelling example of convergent evolution of gene functions and architecture, but also provides a unique opportunity to disentangle the contribution of different genes in this complex

phenotype.

Acknowledgements This work has received support from the grants GENRES (Occitanie-France Olive), REPOL (BIODIVOC, Occitanie), the side project JASMINE (EUR TULIP-GS), and the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie actions (FRUITFUL, H2020-MSCA-IF-2018-842234). P.R., J.C., H.G., and G.B. are members of the CRBE laboratory supported by the excellence projects LabEx CEBA (ANR-10-LABX-25-01) and LabEx TULIP (ANR-10-LABX-0041), managed by the French ANR. P-O.C., and G.B. are also supported by the European Union's Horizon 2020 project GEN4OLIVE (H2020-SFS-2020-1; G.A. No. 101000427). The olive collection is managed by the Platform "Terrains d'Expériences" of the LabEx CeMEB (ANR-10-LABX-04-01). We thank Thierry Mathieu, Pauline Durbin (CEFE), Hélène Lasserre (France Olive), and Edy Spagnol for their help in collection management, as well as Marie-Pierre Dubois and Lucie Moreau for their help in the lab. We are grateful to the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi:10.15454/1.5572369328961167E12) for providing computing and storage resources. We thank Laure Civeyrel for providing *C. bignoniaceum* and *C. odoratissimum* individuals. We thank Tanja Slotte, members of the Edwards lab as well as two anonymous reviewers for helpful discussion.

Author contributions

G.B., P-A.C., and P.R. designed the study with inputs from J.C. D.B-C. provided Laperrine's olive material. G.B. and P-O.C. managed olive collections. G.B. phenotyped plant material, established the sampling, and designed the markers. G.B., A.P., and H.G. contributed to the sampling of *Chrysojasminum*. G.B. and J.D. provided genome skims of Oleaceae. G.B. and S.K. assembled *BZR1* and *GA2ox* sequences for phylogenetic reconstructions. G.B., S.M., S.Val., and A-L.F. did the lab work. W.M., A.T., and S.Vau. managed the HMW DNA extraction, sequencing, and assembly of the olive and jasmine genomes. P.R. analyzed the genomic data and wrote the paper, with inputs from G.B. and P-A.C. All authors commented on and approved the final draft.

Declaration of interests

GB declares that a patent concerning the marker developed in this work to detect olive compatibility groups has been deposited under the application number EP22306777.

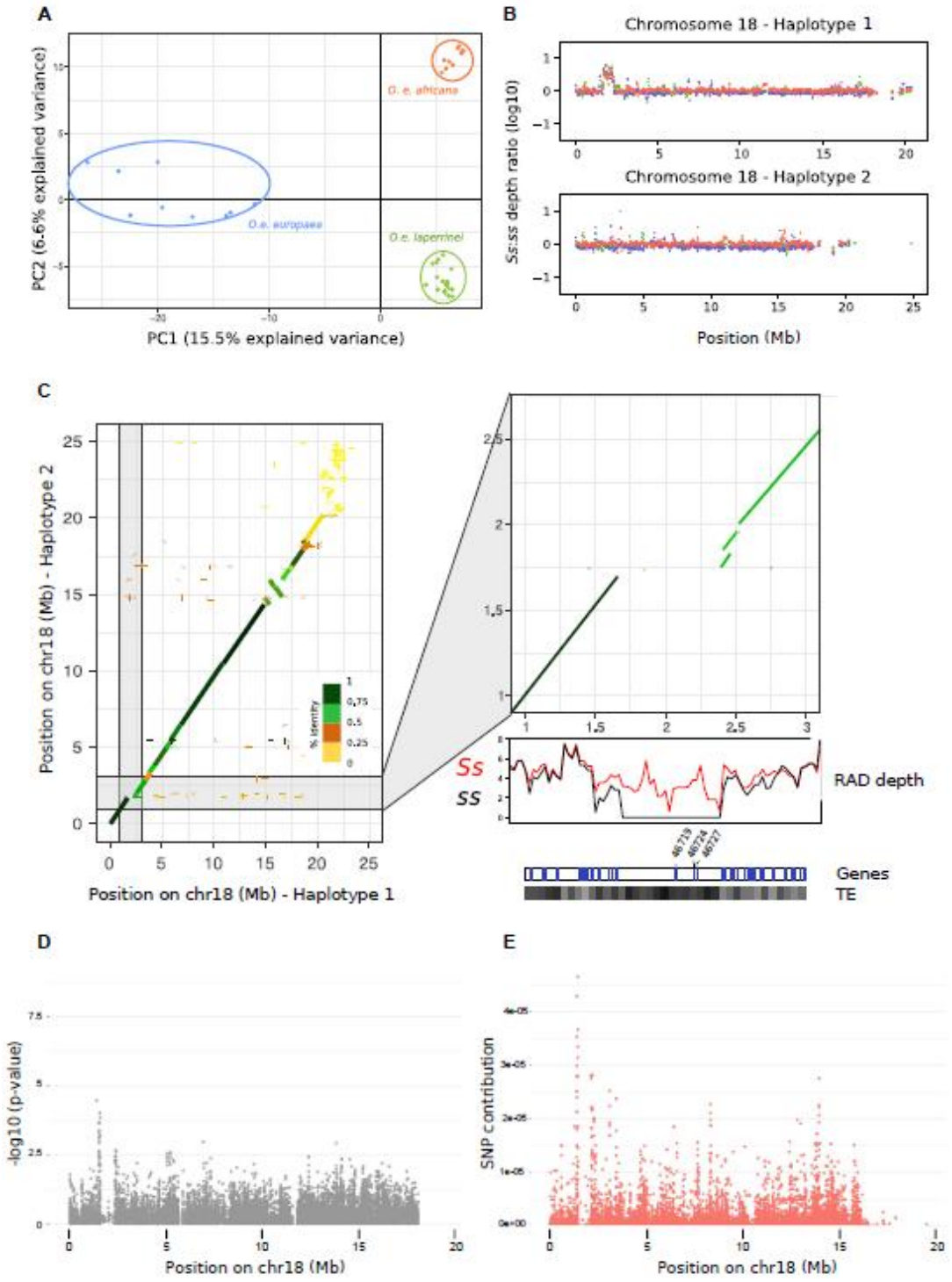

Figure 1. Identification of the *S*-locus in olive. A) Principal component analysis of RAD-sequenced olive individuals based on 43,525 bi-allelic nuclear SNPs. Full and open circle represents *ss* (G2) and *Ss* (G1) self-incompatibility phenotypes, respectively, as determined from parental analyses[8]. B) Ratio of depth of coverage between compatibility groups in three olive subspecies along haplotypes of chromosome 18 of the Saharan olive assembly. We plotted log10[(*Ss*+1)/(*ss*+1)] for sliding windows of 100 kb with 25-kb steps. Values in each of the three subspecies investigated are figured in the same color as in A. See also Figures S1, S2. C) Pairwise synteny between the two haploid sequences of chromosome 18 of the Saharan olive assembly. The left plot presents pairwise identity over the whole scaffold, and a zoom of the region identified with coverage ratio analysis is given on the right. The location of genes (blue segments; for which identifier numbers are given within the hemizygous region), the proportion of transposable elements (TEs) per 50-kb window (darker shades denote higher TE content), and median RAD depth of coverage per 100-kb sliding window in each compatibility group (*Ss* in red, *ss* in black) are plotted along 'haplotype 1' axis. D) Genome-wide association results showing peaks of phenotype-associated SNPs surrounding the hemizygous region (1.6-2.3 Mbp). E) Discriminant analysis of principal components results revealing SNPs strongly contributing to compatibility groups separation surrounding the hemizygous egion (1.6-2.3 Mbp).

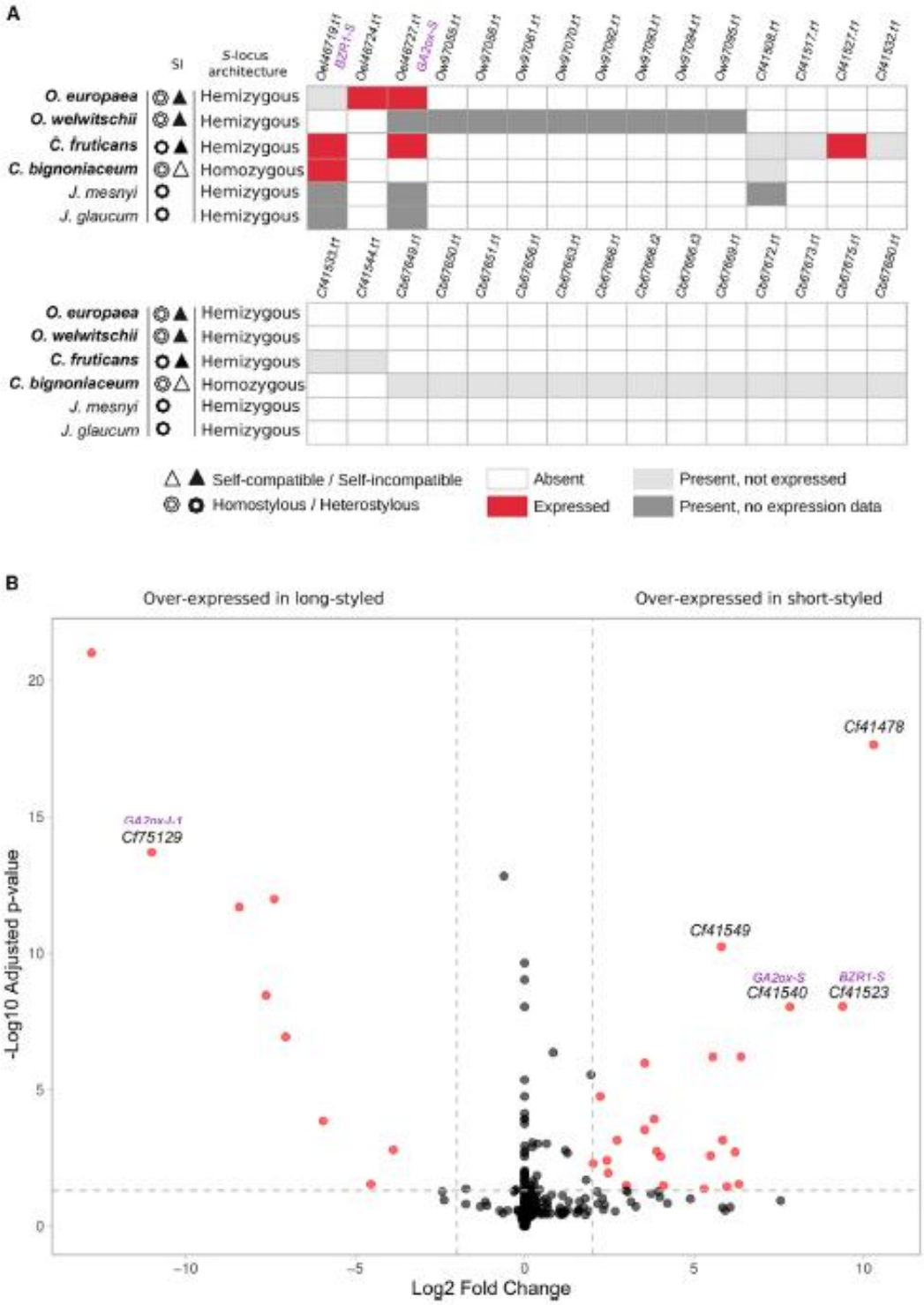

Figure 2. Genetic overlap between heteromorphic and homomorphic Oleaceae. A) *S*-locus gene content and associated expression profiles. The 30 distinct genes identified within *S*-locus regions in four Oleaceae genomes (in bold) are presented under their identifier in the first genome in which they were identified (i.e., *Olea europaea, Olea welwitschii, Chrysojasminum fruticans, Chrysojasminum bignoniaceum*). More information about these assemblies is presented in Data S1B. We also show if they were retrieved in short-read sequencing data of short-styled individuals from two *Jasminum* species (see Data S1F). Expression profiles are summarized to a binary state "expressed/not expressed". Details on expression profiles are available in Data S2. Investigations for the two other genes differentially expressed and neighboring the hemizygous region in *C. fruticans* (*Cf41478.t1* and *Cf41549.t1*) are described in STAR Methods where we explain why we do not consider them as part of the *S*-locus. B) Differential gene expression between long-styled and short-styled *C. fruticans* entire floral buds. Differentially expressed genes are highlighted in red, according to the log2 fold-change (>2) and adjusted *p*-value threshold (0.05), as delimited by the dashed lines. Annotations and expression levels for the differentially expressed genes are available in Data S2.

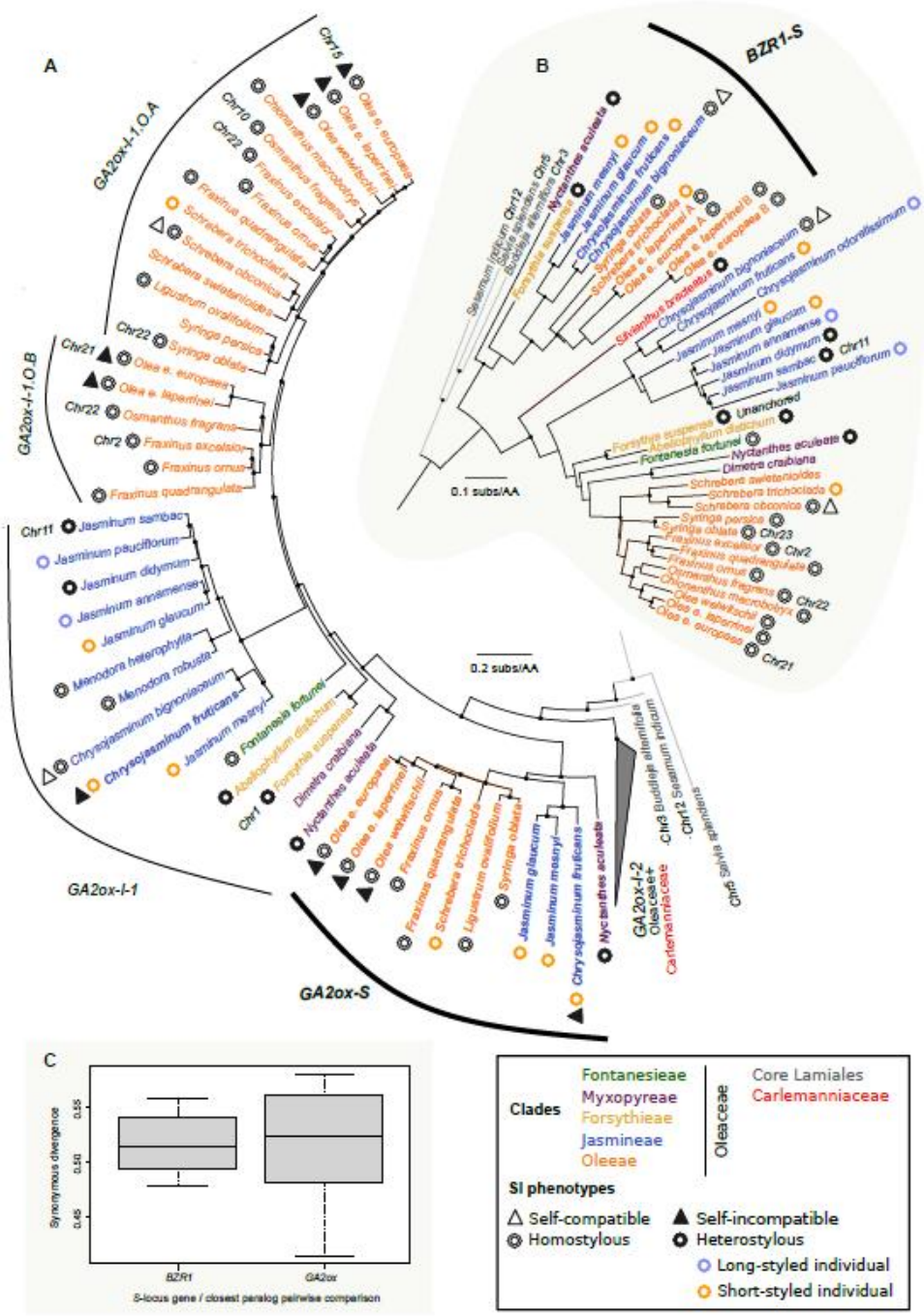

Figure 3. Ancestral origin of the *S*-locus. A) Reduced representation of the maximum likelihood phylogenetic tree of the *GA2ox* gene family in Oleaceae based on protein sequence alignment. Paralogs representing duplicates originating before Oleaceae divergence from Core Lamiales are not shown. An ancestral duplication common to Oleaceae and Carlemanniaceae and leading to *GA2ox-I-2* was collapsed to focus on the duplication that gave rise to the *GA2ox-S* copy (see STAR Methods). Bullets denote nodes with ultrafast bootstrap support greater than 90. Duplicates within the hemizygous supergene and their orthologs, as well as the *GA2ox* copy specifically expressed in long-styled *C. fruticans* (*Cf75129.t1* in Figure 2B), are highlighted in bold. B) Maximum likelihood phylogenetic tree of the *BZR1* gene family in Oleaceae based on protein sequence alignment. The tree was rooted on the split between Oleaceae+Carlemanniaceae and Core Lamiales. Bullets denote nodes with ultrafast bootstrap support greater than 90. Bold is used to highlight duplicates found within the hemizygous supergene and orthologs. *BZR1-S* A and B are tandem duplicates in olive. *BZR1-S* B is annotated in the two previously published olive genomes, but solely supported by computational prediction and no biological hints in our annotation. A-B) For both gene families, symbols denote self-incompatibility phenotypes of the sequenced individuals when known (details in Data S1F). Thus, even if *Fraxinus* and *Ligustrum* species reportedly share the same self-incompatibility systems as olive, they are not indicated as such. When determined, physical location of the *S*-locus paralogs is indicated at the tips. C) Pairwise synonymous divergence between *S*-genes (*GA2ox-S/BZR1-S*) and their closest paralogs in Oleaceae accessions where both gene duplicates were retrieved (i.e., *O. e. laperrinei*, *O. e. europaea, C. fruticans*, *J. mesnyii*, *J. glaucum*, *N. aculeata*). The closest paralogs for both genes were consistently found on the same chromosome across Oleaceae. The median divergence is not different between the two gene sets (Wilcoxon test; *p*-value = 0.84).

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Guillaume Besnard (guillaume.besnard@univ.tlse3.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All sequencing data generated in this study has been uploaded to NCBI. Accession numbers are listed in the key resources table.
- All original code is publicly available at Zenodo, with doi 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

Olive trees (*Olea europaea* and *O. welwitschii*), as well as *Chrysojasminum fruticans* individuals are maintained on the experimental field or in a frost-free greenhouse of CEFE in Montpellier, France. *Chrysojasminum bignoniaceum* and *C. odoratissimum* individuals are cultivated in a mixture of potting soil and sand (50/50) in a frost-free greenhouse at CRBE, Toulouse. Other samples were extracted from herbarium specimens. All voucher numbers are reported in the key resources tables and Data S1F.

METHOD DETAILS

Whole genome sequencing and assembly

We provide the first de novo genome assembly for the Saharan olive (*Olea europaea* subsp. *laperrinei*), found only in relict populations today[28]. High-molecular weight DNA from fresh leaves of the individual 'Adjelella 9_S4' (phenotyped as belonging to the *Ss* group)[8] was extracted using Qiagen Genomic-tips 500/G kit (Cat No./ID: 10262) for whole genome sequencing and an HIFI SMRTbell® library was constructed using the SMRTbell® Prep kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) according to PacBio recommendations (PN 102-166-600). Sequencing was performed at Gentyane platform (Clermont-Ferrand, France) using two PacBio SMRT cells. Raw reads were corrected using SMRTLink v10.1 and HiFi reads were assembled using HiFiasm v15.5[29], which produces partially phased assemblies, i.e., each phased chromosome copy is randomly assigned to one of the two haplotypes of a diploid genome. Assembly completeness was assessed with BUSCO against eudicots_odb10[30] and allele separation was checked via *k*-mer analysis with KAT[31]. All steps from extraction to assembly were performed at the CNRGV (Centre National de Ressources Génomiques Végétales, Toulouse, France). We used Minimap2[32] to map back long reads on the haplotype assembly and to perform whole-chromosome alignment. Synteny relationships were investigated using MCscan and the jcvi python library[33].

Hifiasm produced two highly-contiguous partially phased assemblies for the Saharan olive (Data S1B). The *k*-mer spectra confirm good allele separation between the two assemblies. The proportion of duplicated genes is congruent with expectation and observations in other Oleeae species due to the allopolyploid origin of the tribe[20,34]. Overall, 59,471 protein-coding genes were annotated with at least partial support from biological hints (OrthoDB or RNA-seq) in 'haplotype 1', of which 7,775 were excluded during our stringent filtering of TE-associated sequences. The final number of genes is close to those of the oleaster and 'Arbequina' olive genomes[34,35]. The

Saharan olive genome is composed of about 66% of repeats, DNA transposable elements cover 35% of the genome, the rest are retrotransposons. Mutator-like elements are the most abundant type of repeats, accounting for nearly 26% of the genome. Synteny analyses identified 24 contigs collinear with the 23 chromosomes of the previously published assembly in 'haplotype 1', as chromosome 18 is split into two contigs in our assembly (Data S1C). For 'haplotype 2', we retrieved 28 contigs that match the original 23 chromosomes. The first cultivated olive assembly (cultivar 'Farga'[36]) includes recently diverged copies of many genes, which has been interpreted as the remnants of a recent polyploidization in the cultivated olive followed by a rapid rediploidization[37]. However, these duplicated regions could also be an assembly artifact in which divergent alleles have not been merged due to the high heterozygosity of this olive cultivar[38]. The high synteny of our wild Saharan olive assembly with this of cultivar 'Arbequina'[35] supports that there has not been a WGD event specific to the cultivated olive.

Four other genome assemblies were generated following the same protocol for *O. welwitschii* ("Kakamega 1"), *C. bignoniaceum* (self-compatible and long homostylous; "CRBE"), and a brevistylous individual of *C. fruticans* ("CEFE S6") (Data S1B). Two SMRT cells per species were used except for *C. bignoniaceum* for which only one was used because the *k*-mer spectra reveal very little heterozygosity in this species. We also generated haplotype assemblies for *O. welwitschii* and *C. fruticans*. All assembly statistics are presented in Data S1B.

Genome annotation

Olive genome assembly was soft-masked using Red[39] and annotated using Braker2[40] with hints from OrthoDB[41] and RNA-seq data for *O. e.* subsp. *europaea* from leaves, flowers and fruits publicly available in the European Nucleotide Archive (SRR9203034, SRR9203035, SRR9203038, SRR9203039, SRR9203042, SRR9203067). Reads were cleaned with Fastp[42] and mapped using STAR2[43]. We only kept genes at least partially supported by external hints (proteins homology or expression data). Functional annotation was performed with InterProScan[44] with default parameters. We used EDTA[45] to identify transposable elements (TEs). We then filtered out genes associated with TEs from the annotation file. We first used BEDtools[46] to remove any exons that overlapped a TE, although genes containing both exons that overlapped TEs and exons that did not overlap TEs were retained. We also removed any gene functionally annotated with 'transpos*' (transposon, transposase, etc.), 'ribonuclease H', 'pol poly', or 'retrovirus'. We then used the non-overlapping EDTA output annotation ('split.gff3') to compute the proportion of TEs per 50-kb window with BEDtools. We performed a Wilcoxon test to compare the proportion of TEs in the hemizygous region to the one in the neighboring regions (50-kb windows, up and downstream). We followed the same pipeline to annotate the other genome assemblies we generated but without the use of RNA-seq data.

RAD-sequencing and reads processing

Olive belongs to a species complex that diversified since the Late Miocene, about 6 to 8 Mya[47,48]. We selected 37 accessions from three olive subspecies (nine *O. e.* subsp. *europaea*, 18 *O. e.* subsp. *laperrinei*, and 10 *O. e.* subsp. *africana*) for which the self-incompatibility phenotype has been previously determined using paternity tests on realized matings[8]. In total, 19 individuals are *Ss* and 18 are *ss* (Data S1A). DNAs were extracted with BioSprint (Qiagen), and 200 ng were then digested with *Pst*I. RAD-seq libraries were prepared at the GenoToul sequencing platform facility (Toulouse, France) following the protocol described by Etter et al.[49]. Samples were pooled into three distinct libraries and multiplexed. The three libraries were sequenced on one lane of Illumina NovaSeq run to produce 150-bp paired-end reads. Reads were demultiplexed and cleaned with the process_radtags module of Stacks v2.5[50], allowing the rescue of reads with two mismatches in barcodes (less than the distance between any two barcodes in the used set). Cleaned demultiplexed reads were then mapped using Bowtie2[51] with default parameters to the oleaster genome[35], as well

as the cultivar 'Arbequina' genome[34], and finally to the two haplotype assemblies of the Saharan olive generated in this study. A total of 618,291,848 paired end 150-bp reads of RAD markers were generated, 99% of which were retained after quality filtering and demultiplexing. The mapping rate over the oleaster genome assembly ranged from 74 to 85% per sample and was higher for the 'Arbequina' genome (78.5 to 90%) and for both haploid assemblies of the Saharan olive (84 to 93.5%).

Variant calling and population structure

From read alignments on each reference genome, we called bi-allelic SNPs with bcftools[52], keeping only reads with a mapping quality above 20 (excluding multi-mappers), a base-calling quality of at least 20, and a mean read depth over all individuals between 5 and 60 (twice the mean depth to exclude errors due to mapping in paralogous/repetitive regions). We further filtered sites missing in more than 90% of individuals and those with a minor allele frequency inferior to 0.05 using vcftools[52]. For an individual genotype to be called, we also required a minimum coverage of 5. We only kept one site every 1000 bp to perform a principal component analysis (PCA) of the 37 samples using vcfR v1.14 and adegenet v2.1.10 packages in R[53,54]. There was a greater diversity in the Mediterranean olive sampling set than in *laperrinei* and *africana* subspecies.

Genome scans

Given that both compatibility groups are functionally conserved in divergent Oleeae lineages, one could expect that the two non-recombining haplotypes of the *S*-locus (*s* and *S*) are highly divergent[55] (Figure S1A). Pairwise F_{ST} were thus calculated in 50-kb sliding windows along the different reference genomes with vcftools using the complete SNP set generated in the variant calling section. To test whether the differentiation between compatibility groups deviates from null expectations, we performed 1,000 permutations by randomly shuffling individuals among the two groups before re-estimating F_{ST} for each window. We then calculated *p*-values as the proportion of permuted F_{ST} values that were equal to or larger than the observed one, applying a Benjamini-Hochberg correction for multiple testing with the p.adjust function in R. While acknowledging the limits of our dataset (small sample size and large windows due to the fragmented nature of RAD-seq coverage) that limit power to correctly reject the null hypothesis (a single permutation over 1000 with a F_{ST} greater than the observed one is sufficient to obtain a *p*-value close to 1 after correction for multiple testing), we looked at the windows in which the observed F_{ST} was greater than that observed in every 1000 permutations (*p*-values < 0.001) in the Saharan olive individuals. Regardless of the reference genome used, four to 11 windows F_{ST} exceeded this threshold. We identified F_{ST} peaks at different coordinates on the sequences corresponding to chromosome 18 (none of them coinciding with the region previously linked to self-incompatibility[56]) in each assembly: one in the oleaster genome (of a total of 11 windows with high F_{ST}), five windows out of nine in the 'Arbequina' genome, and three windows out of 11 in 'haplotype 1' of the Saharan olive. None of the four windows identified in 'haplotype 2' were on chromosome 18. The coordinates of these windows are available on Zenodo.

Haplotypes *s* and *S* may have diverged enough to prevent reads from one haplotype to properly map onto the other, similarly to what is observed for sex chromosomes[57,58]. Alternatively, hemizygosity of the *S*-locus, a common feature of distylous systems[12-17] (Table S1), could explain the absence of recombination that underlies the conservation of the Oleaceae self-incompatibility system over a long evolutionary period. Both configurations would result in sequencing-depth variation between groups (Figures S1B and S1C). We used SeqKit[59] to predict *Pst*I restriction sites in the three reference genomes to which reads were mapped. We then estimated the depth of coverage at each predicted *Pst*I locus with samtools[52]. For each sample, we normalized the coverage by dividing it by the millions of mapped reads. We filtered out sites, removing those with a median normalized coverage smaller than 0.5 read per million of mapped reads across both groups for each

olive subspecies. The ratio between the median depth in *Ss* individuals and *ss* individuals along the genome was calculated as log10[(*Ss*+1)/(*ss*+1)], where *Ss* and *ss* stand for the normalized number of reads in each group. It was plotted by 100-kb sliding windows with a step size of 25 kb, using the python library matplotlib[60]. The compatibility phenotype of the individual sequenced for the oleaster reference genome is unknown. Both cultivar 'Arbequina' and the Saharan olive we used for whole-genome sequencing are *Ss*[8] and the expected profile will depend on the assembled haplotype (Figures S1B and S1C). If the assembled haplotype is *s*, this log-transformed *Ss*/*ss* ratio would be lower or equal to 0, depending on the degree of divergence between the two haplotypes. Indeed, in the eventuality of a hemizygous *S*-locus with the *s* allele being a deletion compared to the *S* allele (Figure S1C), we would not observe a drastic change in the coverage ratio (in reality, it will be a bit reduced in the flanking region but this would not be visible in our windowed RAD-seq data). If the *s* allele is not a deletion but strongly divergent, the log-transformed ratio would be negative, as *ss* individuals will have twice the number of reads mapping compared to the heterozygous individuals (Figure S1B). Conversely, if the assembled haplotype is *S*, the log-transformed *Ss*/*ss* ratio would be greater than 0, as long as the two alleles are divergent enough to impact mapping success. We also examined the occurrence of a significant depth difference between *Ss* and *ss* individuals in 50-kb windows along each reference genome (including non-anchored smaller contigs) using two-sided Wilcoxon tests. To increase our power and as we are looking for a genetic region conserved across species, we pooled individuals from each compatibility group together, regardless of the olive subspecies to perform these tests. Similar results were obtained regardless of the reference genome used for depth ratio scan. We here provide the detailed information of the significant regions in each genome. The regions underlying the large peaks in coverage difference in each olive genome assembly are syntenic (Data S1C). In the oleaster reference genome, a modest increase (to 0.2 for a unique window) in *Ss*:*ss* coverage ratio was observed in the region previously associated with the self-incompatibility phenotype (8.5-9.1 Mb on chromosome 18[56]), though only using Saharan olive individuals. On the same chromosome, a greater peak (up to 0.8), common to the three subspecies was detected between 16.7 and 16.9 Mb. Our Wilcoxon tests for difference in depth between groups (individuals pooled by compatibility groups regardless of the subspecies) identified three windows at these coordinates as significantly more covered in *Ss* individuals (FDR < 0.05). Three more windows with this profile were detected on two unanchored scaffolds: NW_019268110.1 (from 0.5 to 1 kb over a total size of 262 kb) and NW_019238463.1 (0.5 to 1 kb for a total size of 145 kb). NW_019238463.1 overlaps with one of the markers that showed partial association with SI in *Phillyrea*[61]. We did not detect any windows shared across the three subspecies with the opposite profile (higher coverage in *ss* than in *Ss*). Using the genome of olive cultivar 'Arbequina' as reference, a unique and large increase in coverage ratio (to a value of 0.8) was observed on chromosome 18 in the three subspecies between 18.7 and 19.5 Mb. This was supported by Wilcoxon tests, as we only detected nine 50-kb windows with significant differences in depth genome-wide, all between 18.85 and 19.35 Mb. In the Saharan olive haplotype assembly, we detected a large peak in 'haplotype 1': on the scaffold corresponding to chromosome 18 (h1tg0000037l; Data S1C) between 1.6 and 2.3 Mb (Figure 1B). All 50-kb windows in this interval (15 windows) had significantly greater coverage in *Ss* individuals (see Figure 1C for normalized coverage values in each group). No significant windows were identified outside of this region. As a consequence, we suggest the assembled haplotype is the dominant *S* haplotype in both reference genomes as well as in our 'haplotype 1' assembly. In contrast, we did not detect any peaks (Figure 1B) or significant windows (FDR < 0.05) indicating a differential coverage between groups in 'haplotype 2', that we infer to be the recessive *s* haplotype. It means that individuals from both compatibility types were sequenced and mapped with similar rates on this haplotype assembly. This last observation is consistent with a moderate divergence between the two haplotypes at the *S*-locus (Figure S1A) or with the absence of this *S*-locus (Figure S1C). As the former is not congruent with the results of our F_{ST} scans and given the peaks identified in other genome assemblies, we postulate the dominant *S* haplotype is hemizygous.

Comparison of the contigs corresponding to chromosome 18 in the two Saharan olive haplotype assemblies revealed that a large indel of over 700-kb explains the coverage difference between the two groups (Figure 1C). This region in 'haplotype 1' has no homolog in 'haplotype 2', yet the flanking regions have homologs in 'haplotype 2', where they are adjacent. These patterns indicate that a 0.7 Mb-stretch of DNA is present in 'haplotype 1', yet missing from 'haplotype 2'. We confirmed this insertion by mapping back HiFi reads on the two haplotype-resolved assemblies. In the oleaster assembly, the sequences underlying the peak detected on chromosome 18 (at 16.8 Mb) as well as on the two unanchored scaffolds are also syntenic fragments of the 700-kb *S*-specific region. The full *S*-locus is thus also present, but scattered across this genome assembly. The fragment on chromosome 18 corresponds to the 5' end of the indel sequence, NW_019238463.1 to the middle, and NW_019268110.1 to the 3' end, with some downstream sequence. The newly identified region is, in both 'haplotype 1' and 'Arbequina' assemblies, contiguous to the candidate region defined by Mariotti et al.[56], but 8 Mb apart in the oleaster assembly.

We performed a genome-wide association study to detect loci associated with compatibility groups using the SNP set generated for the 37 individuals mapped on the Saharan olive genome. Although RAD-seq data is not the most adapted to genome-wide association analysis, we capitalized on the binary nature of the considered phenotype to provide a strong signal. We used GEMMA[62] to fit a linear mixed model incorporating population structure as a relatedness matrix. Input files for GEMMA were prepared with PLINK[63]. Genetic associations were identified using the *p*-value of a Wald test. In parallel, we also performed on the same set of SNPs a discriminant analysis of principal components, as implemented in adegenet, to identify sites that most strongly contribute to the separation of compatibility groups.

Phylogenetic analyses

To gain insight on the origin of the *S*-locus, we then investigated the gene family of the three genes present in the olive hemizygous region. We first performed blast search to extract paralogs from Oleaceae genomes and other Lamiales as outgroups (Data S1F). To get a sample of taxa representative of all Oleaceae tribes and subtribes, we also used shotgun data of species with a minimum genome sequencing depth of 6.5×. On these taxa, a reference-guided approach was used to assemble targeted genomic regions (in Geneious v9[64]) as described in Bianconi et al.[65]. Briefly, exons of genes from *O. europaea* were used as seeds to map homologous reads and assemble them into contigs, which were then elongated by recursively incorporating reads identical to at least 30 bp at the ends of the contig. Reads were then mapped onto the assembled contigs (to estimate their sequencing depth), manually checked to identify potential heterozygous sites and majority-rule consensus sequences were extracted. For each gene family, complete coding sequences were translated into proteins and aligned using MAFFT[66]. Maximum-likelihood phylogenetic trees were inferred using IQ-tree v2.1.3[67]. The best evolution model was selected by ModelFinder[68] and we assessed branch support with 1000 ultrafast bootstrap (UFB) replicates[69]. Trees in newick format were deposited on Zenodo. We retrieved homologs to build phylogenies for each of the genes predicted within the olive hemizygous region:

- For *GA2ox-S*, excluding duplicates originating before Oleaceae and Carlemanniaceae divergence from Lamiales, *S*-locus orthologs were retrieved in Myxopyreae, Jasmineae, and Oleeae (Figure 3A; Data S2). Their co-ortholog '*GA2ox-I-1*' experienced a second duplication event in the common ancestor of all Oleeae leading to '*GA2ox-I-1.O.A*' and '*GA2ox-I-1.O.B*' in-paralogs.
- *BZR1-S* orthologs were retrieved in four Oleaceae tribes (Myxopyreae, Forsythieae, Jasmineae, and Oleeae). In addition, co-orthologs of *BZR1-S* with respect to the divergence with core Lamiales were also retrieved in most Oleaceae and Carlemanniaceae (Figure 3B). We did not detect any in-paralogs in respect to the whole genome duplication event in Oleaceae.

- For *Oe47624.t1*, we did not retrieve orthologs for this gene in other Oleaceae, even in other closely related Oleeae (such as *Olea welwitschii*). This fact itself weakens the probability of its key implication in self-incompatibility. Moreover, two duplicates exist in olive, but are highly divergent from any other paralogous sequences identified in other Oleaceae. We consequently suggest this gene is a DNA fragment that was incidentally copied within the *S*-locus.

Homologs of *BZR1-S* and *GA2ox-S* were thus detected in 13 Oleaceae individuals (out of 30 examined), but not systematically together (Figures 2 and S4). At the family scale, the presence of the two genes is associated in distantly related lineages, namely Myxopyreae (*Nyctanthes*), Jasmineae (*Jasminum*, *Chrysojasminum*), and Oleeae (*Schrebera*, *Syringa*, and *Olea*). In contrast, while *BZR1-S* was found in the autogamous and long-homostylous individual of *C. bignoniaceum* and in *Forsythia suspensa* (unknown phenotype), the *GA2ox-S* copy was missing. Conversely, in four individuals of non-distylous Oleeae where *GA2ox-S* is present, *BZR1-S* was not found (in *Fraxinus ornus* and *Fr. quadrangulata*), or was degenerated/truncated (in *Olea welwitschii* and *Ligustrum ovalifolium*).

In olive, the two *S*-locus genes (*BZR1-S* and *GA2ox-S*) are separated by 230 kb on chromosome 18. Their closest paralogs are both found on chromosome 21, 969 kb apart. In every species where they are both present, *BZR1-S* and *GA2ox-S* are found within a distance of 143 to 404 kb on the same chromosome. Interestingly, their closest paralogs were 0.75 to 2.5 Mb apart from each other in Oleaceae, a distance that appears to scale with the extent of the hemizygous *S*-locus in their respective genomes. In Oleeae, where *GA2ox* underwent an extra duplication, the *GA2ox-I-1-O.A* copy was systematically on a distinct chromosome, while the *GA2ox-I-1-O.B* copy was located on the same chromosome as the *BZR1* paralog, indicating that the *GA2ox-I-1-O.B* copies are likely the co-orthologs of *GA2ox-S*. *GA2ox-I-1-O.B* was not detected in *Sy. oblata* and *Sc. trichoclada*.

Comparative genomics of the candidate S-*locus across Oleaceae*

Using synteny analyses, we identified the regions homologous to the olive *S*-locus in four other species for which we produced genome assemblies. We were thus able to compare the gene content of the region in each genome. We further delimited the extent of hemizygosity in *C. fruticans* and *O. welwitschii* by comparing contigs homologs to the olive *S*-locus. We checked that hemizygosity was supported by mapping Hifi reads on the assembled contigs and check we had reads spanning both types of junctions (with or without the indel). In *O. welwitschii*, nine genes were annotated within the hemizygous region*, GA2ox-S* being the only one shared with another species. In *C. fruticans*, the hemizygous region on h2tg00002751 contains eight genes, including homologs for *BZR1-S* and *GA2ox-S*. Only one of the extra genes has a predicted function (*Cf41508.t1*). It is annotated as GLUTAREDOXIN-C9, a gene that seems involved in oxidative stress response. The homologous region in *C. bignoniaceum* contains 15 genes. Only two of them had homologs in *C. fruticans*: *BZR1-S* and the glutaredoxin gene (see Data S2 for gene names correspondence between assemblies). Few of these species-specific genes have predicted functions and even fewer are expressed (Figure 2A). We consequently suggest these extra genes are spurious annotations of gene fragments that were incidentally copied within the *S*-locus, a process facilitated by the high concentration of TEs. For instance, we did not find any orthologs for the third *S*-locus gene in olive (*Oel46724.t1*) and it is highly divergent from any other paralogous sequences identified in other Oleaceae. Whatever these represent real functional genes or not, the fact that they are not widely distributed in self-incompatible Oleaceae weakens the probability of their implication in self-incompatibility.

Finally, in the set of differentially expressed genes in *C. fruticans* (see below), we identified two extra genes (*Cf41549.t1* and *Cf41478.t1*) physically-linked with the hemizygous region. *Cf41549.t1* is a hemizygous duplicate of *Cf41567.t1* (200-kb downstream on the same

chromosome). Their divergence level is really low (only two non-synonymous mutations), but the last exon of *Cf41549.t1* is missing. A single homolog is found in the *C. bignoniaceum* genome, on a different chromosome than the rest of the *S*-locus homologs in this species. While we had some hits for *Cf41549.t1* with RNA-seq from *C. bignoniaceum* and *C. odoratissimum* (Data S2), we thus deemed their assignation to the hemizygous copy as unreliable. Moreover, inspection of the mapped reads revealed most of them harbored one or two of the substitutions specific to *Cf41567.t1*. For *Cf41478.t1*, we did not detect expression of its only homolog in olive (located on chromosome 12), but it is also expressed in *C. bignoniaceum* and *C. odoratissimum* (long-styled). If it was linked to self-incompatibility/distyly, we would expect its expression to follow the same pattern as for *C. fruticans* (expressed in short-styled only). The expressed homolog in *C. bignoniaceum* is also not on the same chromosome as the other *S*-locus homologs. We consequently consider their association with self-incompatibility/distyly questionable and did not include them in the *C. fruticans* supergene.

We computed synonymous divergence between *S*-linked genes and their closest paralogs, when both present, as a proxy for the time since their duplication using the yn00 program from PAML v4[70]. We also investigated the genomic position of paralogs of *S*-linked genes in available genome assemblies. If the *S*-locus formed via a single segmental duplication, their closest paralogs of *S*-linked genes would stem from the same genomic region. Conversely, if the supergene formed by stepwise addition of genes, in-paralogs would be scattered across the genome.

We finally looked at the sequencing depth for five species (*Nyctanthes aculeata*, *Jasminum mesnyi*, *J. glaucum*, *Schrebera trichoclada*, and *Ligustrum ovalifolium*) analyzed with short-read shotgun data in which *S*-genes were detected. We compared the mean sequencing depth on *S*-genes compared to their paralogs.

Transcriptomics of self-incompatibility

We generated floral bud RNA-seq data for both morphs in *C. fruticans* (six short-styled, four long-styled individuals at both 0.5- and 1-cm length stages). Plants were collected (on April 29, 2021) at the common garden of CEFE in Montpellier, France. We also sampled buds for two congeneric jasmines, a long-styled individual of *C. odoratissimum*, and a long-homostylous and self-compatible individual of *C. bignoniaceum*. Samples were placed in liquid nitrogen on the field right after their collection, transported on dry ice and kept at -80°C. RNA extractions were performed with the Qiagen RNeasy Plant Mini Kit following manufacturer's instructions. Extracts were then purified with Promega RQ1 RNAse Free DNAse to limit DNA contamination. Quality controls and library preparation were performed at the GenoToul sequencing platform. One SP lane of Illumina NovaSeq was used to generate 150-bp paired-end reads. Raw reads were cleaned and trimmed using Fastp with default parameters to remove any low-quality bases or adapters. We used STAR2 to map reads to our annotated *C. fruticans* genome and quantify read expression (--quantMode GeneCounts). We then used the matrix of raw counts to perform differential gene expression with DESeq2[71]. We included morph, stage and their interaction in the design and used DESeq2's metric 'median of ratios' to normalize counts. For data representation, we finally used log2FoldChange shrinkage with the method apeglm[72].

For olive, we used publicly available expression data for inflorescences. We retrieved 15 runs for *Ss* cultivars (buds from six different branches of one tree at flower induction and bud differentiation stages[73], plus three distinct trees from PRJNA525000[74]). We also checked we did not retrieve any expression in leaf tissue (SRP101652), in whole inflorescences from five trees from PRJNA525000. We also checked their absence in *ss* pollen and pistil using transcriptomes from ReprOlive[75]. We used STAR2 as before for mapping and quantification. Finally, we also used RT-PCR on eight olive inflorescence RNA samples (4 *Ss*, 4 *ss*), to check *BZR1-S* was not missed

because of insufficient sequencing effort in these transcriptome studies.

Markers design and amplification procedures

Eight genotypes were first tested with the marker 'C6030' previously proposed as a diagnostic marker for self-incompatibility type[56] in the Mediterranean cultivated olive (*O. europaea* subsp. *europaea*). We tested four Mediterranean cultivars and four genotypes of *O. e.* subsp. *africana* (African Brown Olive) that were previously phenotyped for compatibility groups (Data S1A). In addition, we also checked the inheritance of C6030 alleles among 34 self-progenies of Mediterranean cultivars [25 were from L4R14 ('Koroneiki'), four from L4R19 ('Cailletier/Frantoio'), two from L4R13 ('Arbequina'), and three from L4R17 ('Verdale de Millas')]. The C6030 region was amplified following the protocol described in the paper[56] but PCR was not successful on some individuals (i.e., all tested accessions of subsp. *africana* and nine self-progenies of L4R14). Decreasing the annealing temperature (T_a) to 55°C ultimately allowed us to get an amplification on all samples. Sanger sequencing of the PCR products was performed at GenoScreen (Lille, France). Among the tested *O. e. europaea* individuals, a novel C6030 haplotype (GenBank accession: ON667921) was detected in three individuals. This haplotype was named s-b2 (as it was distinguished from s-b by only one single nucleotide polymorphism in the 3' part of the sequence; position 426[56]). It was detected in L4R13 (*Ss*), L4R14 (*Ss*), and L4R17 (*ss*; Data S1A). Unexpectedly, L4R14 (*Ss*) does not show a C6030 profile compatible with a *Ss* genotype as previously described[56]. Indeed, the apparent homozygous genotype for sb should be associated with *ss* (as observed here on L4R17). The detection of only one haplotype could, however, be due to the presence of null allele as mentioned by the authors[56]. The analysis of 25 selfings obtained from L4R14 shows that the s-b2 haplotype is not present in nine progenies supporting that a C6030 null allele is indeed present in L4R14. But, another divergent sequence (named Cm; GenBank accession: ON667922) was finally amplified (at T_a = 55°C) in these nine progenies. We hypothesize that it may correspond to a paralogous locus that can only be amplified in presence of a homozygous genotype for the null allele at locus C6030. Very divergent sequences were also generated on *africana* individuals and two haplotypes were distinguished (GenBank accessions: ON667923 and ON667924, for C1 and C2, respectively). The combination of these haplotypes does not allow recognizing the self-incompatibility group in the four studied *africana* accessions (Data S1A). In addition, the Cm sequence (detected in nine L4R14 progenies) is similar to *africana* sequences suggesting the amplification of a paralogous region, and a lack of specificity of the C6030 marker in our PCR conditions (T_a = 55°C).

As the C6030 marker did not show a systematic association with compatibility groups in our sampling, we developed simple PCR tests able to simultaneously amplify a segment specific to *s* and *S* alleles. These markers were used to validate the genetic association between the hemizygous region and compatibility groups on a larger sampling. We aligned sequences from genetically divergent *O. europaea* accessions (subspp. *africana*, *laperrinei*, and *europaea*) in order to define primers in conserved regions on both extremities of the hemizygous region and on *GA2ox-S*. We thus defined three PCR markers and tested them on 69 phenotyped trees (Data S1A). For DSI-A and DSI-B, *S* and *s* alleles of the *S*-locus were amplified by PCR with the following protocol. Each DSI-A PCR reaction (25 µL) contained 10 ng DNA template, 1× reaction buffer, 3.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 µmol of 6-FAM-labeled M13(21) primer (5'TGTAAAACGACGGCCAGT3'), 0.1 µmol of the S/s_A_For primer, 0.1 µmol of each reverse primer [preferentially s_A(1)_Rev and S_A_Rev to reveal different s alleles, or alternatively s_A(2)_Rev and S_A_Rev that generate a unique shorter s allele (Data S1D)], and 0.5 U of *Taq* DNA polymerase (Promega). For DSI-B, we used a similar PCR protocol except for primers: 0.1 µmol of HEX-labeled M13(21) primer, 0.1 µmol of the S/s_B_For primer, and 0.2 µmol of the S/s_B_Rev primer (Data S1D). We conducted PCR of both loci in a Mastercycler pro PCR System (Eppendorf) for 2 min at 94°C, followed by 25 cycles of 30 s at 94°C, 45 s at 56°C, and 1 min at 72°C, and then by 10 cycles of 30 s at 94°C, 45 s

at 51.5°C, and 45 s at 72°C. The last cycle was followed by a 20-min extension at 72°C. PCR products were finally multiplexed with GenScan-500ROX (Applied Biosystems) and separated on an ABI Prism 3730DNA Analyzer (Applied Biosystems). Chromatograms were read with Geneious v9[64]. The *GA2ox-S* locus was amplified following the next procedure: Each PCR reaction (25 µL) contained 10 ng DNA template, 1× reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µmol of each primer (Data S1D), and 0.5 U of *Taq* DNA polymerase (Promega). We conducted PCR in a Mastercycler pro PCR System (Eppendorf) for 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 45 s at 56°C, and 1 min at 72°C. The last cycle was followed by a 20-min extension at 72°C. PCR products were then migrated on a 2%-agarose gel. For this marker, presence or absence of amplification was scored. Compatibility groups based on parentage analyses of seedlings agreed perfectly with those deduced from our markers in all 64 individuals across the three tested olive subspecies. Different *s* size variants were revealed in olive on DSI-1(A) (4 alleles) and DSI-B (3 alleles; Data S1A and S1D). For the *S*-haplotype, DSI-A(1)_298 (*europaea*) or _332 (*laperrinei-africana*), DSI-B_185, and GA2ox-S_168 are revealed only in *Ss* individuals (Data S1A and S1D). As being co-dominant and easy to amplify and score, we recommend the use of DSI-B for a fast, reliable characterization of self-incompatibility groups in olive germplasm collections. Finally, we also obtained nine progenies of 'Koroneiki' (L4R14) showing only the *S* allele at locus DSI-A and DSI-B supporting the presence of *SS* genotypes (Data S1A). After four years of growth in a greenhouse, eight of them are still maintained in our collection indicating these genotypes are viable in such conditions.

We tested the primers of the *GA2ox-S* locus in *C. fruticans* and developed a marker linked to the *BZR1-S* gene (Cf-*BZR1-S*; Data S1D) to check a potential association between the presence of these genes with one floral morph in this distylous species. Following a similar approach to olive, we also designed primers to amplify at each extremity a region shared by both alleles *s* and *S*, and showing linked indels (Data S1D): Cf-DSI-A (118 bp for *S*, and 124 bp for *s*) and Cf-DSI-B (226 bp for *S*, 190 or 199 bp for *s*). For the amplifications of Cf-DSI-A, Cf-DSI-B, and Cf-*BZR1-S*, we used the same protocol as for olive DSI-B. We genotyped with this set of markers 179 common yellow jasmine individuals from 13 French populations for which the floral morph [long-styled (90 individuals) or short-styled (89 individuals)] was noted during the flowering period in 2022 and 2023[76]. Alleles Cf-DSI-A_118, Cf-DSI-B_226, Cf-*BZR1-S*_171/173 and GA2ox-S_168 co-segregate and are detected only in brevistylous individuals (Data S1D and S1E) as expected for the *S*-haplotype. Allelic variants (4 alleles) were also observed on locus Cf-DSI-B for the *s*-haplotype. Among 89 brevistylous individuals, one mature *SS* genotype was observed in the Bages population (Bages 10). The reduced heterozygosity of this genotype compared to other individuals analyzed from this population (SSR dataset provided in Puyoou et al.[76]) indicates 'Bages 10' was very likely issued from selfing.

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.

Data S1. Sampling details. Related to STAR Methods. A) List of olive individuals of the CEFE collection used in this study, their cross-incompatibility group based on paternity analyses[8], their DSI genotypes. For eight individuals, the C6030 genotype was determined, as well as those of self-progenies for four *europaea* cultivars. B) Assembly statistics of the four Oleaceae genomes generated in this study. C) Macrosynteny between four olive genome assemblies. Only the largest anchored scaffolds are shown for each assembly. D) Primers used to amplify extremities or genes of the *S*-locus. E) Genotypic profiles at four loci of the DSI region observed in 179 individuals of *Chrysojasminum fruticans* and associated floral morph. Bold underlined alleles associated to

brevistyly. F) Oleaceae genomic resources used for phylogenetic analyses. SC = Self-compatible, S = Short-styled (Ss genotype), L = Long-styled (ss genotype), ? = unknown.

Data S2. Presence/absence and expression patterns of *S*-locus genes and jasmine DEG. Related to Figure 2. A) Normalized reads count and annotation of genes differentially expressed between *C. fruticans* morphs. We also include raw counts for *C. bignoniaceum* (long-homostylous) and *C. odoratissimum* (long-styled). B) Presence/absence and expression patterns of *S*-locus genes in different Oleaceae species.

References

1. Green, P.S. (2004). Oleaceae. In Flowering Plants · Dicotyledons: Lamiales (except Acanthaceae including Avicenniaceae), The Families and Genera of Vascular Plants, J.W. Kadereit, ed. (Springer), pp. 296–306. 10.1007/978-3-642-18617-2_16.
2. 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. 10.1038/s41437-021-00463-4.
3. 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. 10.1126/science.1186687.
4. 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. Applic. *10*, 860–866. 10.1111/eva.12498.
5. 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. 10.1111/nph.13872.
6. 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. Resourc. *19*, 877–892. 10.1111/1755-0998.13016.
7. 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. 10.1111/evo.12613.
8. Besnard, G., Cheptou, P.-O., Debbaoui, M., Lafont, P., Hugueny, B., Dupin, J., and Baali-Cherif, D. (2020). Paternity tests support a diallelic self-incompatibility system in a wild olive (*Olea europaea* subsp. *laperrinei*, Oleaceae). Ecol. Evol. *10*, 1876–1888. 10.1002/ece3.5993.
9. Kent, T.V., Uzunović, J., and Wright, S.I. (2017). Coevolution between transposable elements and recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160458. 10.1098/rstb.2016.0458.
10. Ouellette, L., Anh Tuan, P., 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. 10.1016/j.gene.2023.147255.
11. Castric, V., Batista, R.A., Carré, A., Mousavi, S., Mazoyer, C., Godé C., Gallina, S., Ponitzki, C., Theron, A., Bellec, A. et al. (2024). The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene. Curr. Biol. *Accepted pending minor revisions*.
12. Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., and Ota, T. (2012). *S-LOCUS EARLY FLOWERING 3* is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. PLoS ONE *7*, e31264. 10.1371/journal.pone.0031264.
13. Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., van Oosterhout, C., and Gilmartin, P.M. (2016). Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*. Nat. Plants *2*, 16188. 10.1038/nplants.2016.188.
14. 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. 10.1111/nph.15970.
15. 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. 10.1016/j.cub.2022.08.042.
16. Zhao, Z., Zhang, Y., Shi, M., Liu, Z., Xu, Y., Luo, Z., Yuan, S., Tu, T., Sun, Z., Zhang, D., et al. (2023). Genomic evidence supports the genetic convergence of a supergene controlling the distylous floral syndrome. New Phytol. *237*, 601–614. 10.1111/nph.18540.
17. Yang, J., Xue, H., Li, Z., Zhang, Y., Shi, T., He, X., Barrett, S.C.H., Wang, Q., and Chen, J. (2023). Haplotype-resolved genome assembly provides insights into the evolution of *S*-locus supergene in distylous *Nymphoides indica*. New Phytol. *240*, 2058–2071. 10.1111/nph.19264.
18. Yu, X., Li, L., Zola, J., Aluru, M., Ye, H., Foudree, A., Guo, H., Anderson, S., Aluru, S., Liu, P., et al. (2011). A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in *Arabidopsis thaliana*. Plant J. *65*, 634–646. 10.1111/j.1365-313X.2010.04449.x.
19. Veitia, R.A. (2007). Exploring the molecular etiology of dominant-negative mutations. Plant Cell *19*, 3843–3851. 10.1105/tpc.107.055053.
20. Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., Rees, M., Frohlich, M.W., Qi, J., Weigend, M., et al.

(2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. Mol. Biol. Evol. *37*, 3188–3210. 10.1093/molbev/msaa160.
21. Charlesworth, D., and Charlesworth, B. (1979). A model for the evolution of distyly. Am. Nat. *114*, 467–498. 10.1086/283496.
22. Lloyd, D.G., and Webb, C.J. (1992). The evolution of heterostyly. In Evolution and Function of Heterostyly Monographs on Theoretical and Applied Genetics, S.C.H. Barrett, ed. (Springer), pp. 151–178. 10.1007/978-3-642-86656-2_6.
23. 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. 10.7554/eLife.17956.
24. Potente, G., Léveillé-Bourret, É., Yousefi, N., Choudhury, R.R., Keller, B., Diop, S.I., Duijsings, D., Pirovano, W., Lenhard, M., Szövényi, P., et al. (2022). Comparative genomics elucidates the origin of a supergene controlling floral heteromorphism. Mol. Biol. Evol. *39*, msac035. 10.1093/molbev/msac035.
25. Ganguly, S., and Barua, D. (2020). High herkogamy but low reciprocity characterizes isoplethic populations of *Jasminum malabaricum*, a species with stigma-height dimorphism. Plant. Biol. *22*, 899–909. 10.1111/plb.13127.
26. Ferrero, V., Rojas, D., Vale, A., and Navarro, L. (2012). Delving into the loss of heterostyly in Rubiaceae: Is there a similar trend in tropical and non-tropical climate zones? Perspect. Plant Ecol. Evol. Syst. *14*, 161–167. 10.1016/j.ppees.2011.11.005.
27. Tange, C. (1998). *Silvianthus* (Carlemanniaceae) a genus and family new to Thailand. Thai Forest Bull. *26*, 59–65. https://li01.tci-thaijo.org/index.php/ThaiForestBulletin/article/view /25001.
28. Besnard, G., Gorrilliot, O., Raimondeau, P., Génot, B., El Bakkali, A., Anthelme, F., and Baali-Cherif, D. (2021). Contrasting genetic footprints among Saharan olive populations: potential causes and conservation implications. Plants *10*, 1207. 10.3390/plants10061207.
29. 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. 10.1038/s41592-020-01056-5.
30. 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. 10.1093/bioinformatics/btv351.
31. 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. 10.1093/bioinformatics/btw663.
32. Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics *37*, 4572–4574. 10.1093/bioinformatics/btab705.
33. Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and collinearity in plant genomes. Science *320*, 486–488. 10.1126/science.1153917.
34. 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. 10.1073/pnas.1708621114.
35. 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. 10.1038/s41438-021-00498-y.
36. 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. 10.1186/s13742-016-0134-5.
37. Julca, I., Marcet-Houben, M., Vargas, P., and Gabaldón, T. (2018). Phylogenomics of the olive tree (*Olea europaea*) reveals the relative contribution of ancient allo- and autopolyploidization events. BMC Biol. *16*, 15. 10.1186/s12915-018-0482-y.
38. Besnard, G., El Bakkali, A., Haouane, H., Baali-Cherif, D., Moukhli, A., and Khadari, B. (2013). Population genetics of Mediterranean and Saharan olives: geographic patterns of differentiation and evidence for early generations of admixture. Ann. Bot. *112*, 1293–1302. 10.1093/aob/mct196.
39. Girgis, H.Z. (2015). Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinform. *16*, 227. 10.1186/s12859-015-0654-5.
40. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. *3*, lqaa108. 10.1093/nargab/lqaa108.
41. Zdobnov, E.M., Kuznetsov, D., Tegenfeldt, F., Manni, M., Berkeley, M., and Kriventseva, E.V. (2021). OrthoDB in 2020: evolutionary and functional annotations of orthologs. Nucleic Acids Res. *49*, D389–D393. 10.1093/nar/gkaa1009.
42. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. 10.1093/bioinformatics/bty560.
43. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. 10.1093/bioinformatics/bts635.
44. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. 10.1093/bioinformatics/btu031.
45. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. *20*, 275. 10.1186/s13059-019-1905-y.
46. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. 10.1093/bioinformatics/btq033.
47. Besnard, G., Rubio de Casas, R., Christin, P.-A., and Vargas, P. (2009). Phylogenetics of *Olea* (Oleaceae) based on plastid and nuclear ribosomal DNA sequences: Tertiary climatic shifts and lineage differentiation times. Ann. Bot. *104*, 143–160. 10.1093/aob/mcp105.
48. Green, P.S. (2002). A revision of *Olea* L. (Oleaceae). Kew Bull. *57*, 91–140. 10.2307/4110824.
49. Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., and Cresko, W.A. (2012). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In Molecular Methods for Evolutionary Genetics, Methods in Molecular Biology, Vol. 772, V. Orgogozo and M. Rockman, eds (Humana Press), pp. 157–178. 10.1007/978-1-61779-228-1_9.

50. Rochette, N.C., Rivera-Colón, A.G., and Catchen, J.M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. *28*, 4737–4754. 10.1111/mec.15253.
51. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. 10.1038/nmeth.1923.
52. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. 10.1093/gigascience/giab008.
53. Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics *27*, 3070–3071. 10.1093/bioinformatics/btr521.
54. Knaus, B.J., and Grünwald, N.J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resourc. *17*, 44–53. 10.1111/1755-0998.12549.
55. Kamau, E., and Charlesworth, D. (2005). Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. Curr. Biol. *15*, 1773–1778. 10.1016/j.cub.2005.08.062.
56. 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. 10.3389/fpls.2019.01760.
57. Palmer, D.H., Rogers, T.F., Dean, R., and Wright, A.E. (2019). How to identify sex chromosomes and their turnover. Mol. Ecol. *28*, 4709–4724. 10.1111/mec.15245.
58. Qiu, S., Bergero, R., Guirao-Rico, S., Campos, J.L., Cezard, T., Gharbi, K., and Charlesworth, D. (2016). RAD mapping reveals an evolving, polymorphic and fuzzy boundary of a plant pseudoautosomal region. Mol. Ecol. *25*, 414–430. 10.1111/mec.13297.
59. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE *11*, e0163962. 10.1371/journal.pone.0163962.
60. Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. *9*, 90–95. 10.1109/MCSE.2007.55.
61. 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 Journal *1*, e15. 10.24072/pcjournal.23.
62. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824. 10.1038/ng.2310.
63. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. 10.1086/519795.
64. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649. 10.1093/bioinformatics/bts199.
65. Bianconi, M.E., Hackel, J., Vorontsova, M.S., Alberti, A., Arthan, W., Burke, S.V., Duvall, M.R., Kellogg, E.A., Lavergne, S., McKain, M., et al. (2020). Continued adaptation of C_4 photosynthesis after an initial burst of changes in the Andropogoneae grasses. Syst. Biol. *69*, 445–461. 10.1093/sysbio/syz066.
66. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. 10.1093/molbev/mst010.
67. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. 10.1093/molbev/msaa015.
68. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. 10.1038/nmeth.4285.
69. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522. 10.1093/molbev/msx281.
70. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. 10.1093/molbev/msm088.
71. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1186/s13059-014-0550-8.
72. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics *35*, 2084–2092. 10.1093/bioinformatics/bty895.
73. Salimonti, A., Forgione, I., Sirangelo, T.M., Puccio, G., Mauceri, A., Mercati, F., Sunseri, F., and Carbone, F. (2021). A complex gene network mediated by ethylene signal transduction TFs defines the flower induction and differentiation in *Olea europaea* L. Genes *12*, 545. 10.3390/genes12040545.
74. Gros-Balthazard, M., Besnard, G., Sarah, G., Holtz, Y., Leclercq, J., Santoni, S., Wegmann, D., Glémin, S., and Khadari, B. (2019). Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. Plant J. *100*, 143–157. 10.1111/tpj.14435.
75. Carmona, R., Zafra, A., Seoane, P., Castro, A.J., Guerrero-Fernández, D., Castillo-Castillo, T., Medina-García, A., Cánovas, F.M., Aldana-Montes, J.F., Navas-Delgado, I., et al. (2015). ReprOlive: a database with linked data for the olive tree (*Olea europaea* L.) reproductive transcriptome. Front. Plant Sci. *6*, 625. 10.3389/fpls.2015.00625.
76. Puyoou, A., Gryta, H., Fuchs, A.-L., Blanchard, P., Cheptou, P.-O., Civeyrel, L., Dufay, M., Dupin, J., Jargeat, P., Lecompte, E., et al. (2024). Limitation of gene flow by distance in the common yellow jasmine (*Chrysojasminum fruticans*, Oleaceae): implications for the study of its mating strategies. Bot. J. Linn. Soc. *204*, 23–34. 10.1093/botlinnean/boad035.