Supporting information

An ancestral hemizygous supergene controls homomorphic di-allelic self-incompatibility in olive (Oleaceae) - P. Raimondeau *et al.*

Supporting information includes the following items:

Supplementary Material S1. Test of the C6030 marker on a few Mediterranean and African brown olives of the CEFE collection

Supplementary Material S2. Development of simple PCR-based diagnoses of self-incompatibility groups in olive and the common yellow jasmine

Table S1 (separate xls file). Plant material:

(a) List of individuals of the CEFE collection used in this study, their SSR and DSI genotypes, and their cross-incompatibility group

(b) Determining cross-incompatibility group of CEFE individuals through paternity analyses

Table S2. RAD-sequencing and mapping statistics on the oleaster genome.

Table S3. List of windows harboring F_{ST} values greater than the ones obtained in any permutation in each olive assembly.

Table S4. List of Oleaceae genomic resources used for phylogenetics analyses.

Table S5. Synonymous divergence (dS) and sequencing depth ratio between *S*-locus genes and closest paralogs.

Table S6. Genomic location of the *S*-locus genes and their paralogs.

Figure S1. *k*-mer spectra of the two Saharan olive haplotype assemblies

Figure S2. Macrosynteny between four olive genome assemblies

Figure S3. PCA plot of sequenced olive individuals based on bi-allelic nuclear SNPs called on four different genome assemblies

Figure S4. Potential genomic footprints of the *S*-locus depending on its architecture.

Figure S5. Manhattan plots depicting the association between SNP genotype and compatibility group for olive chromosome 18.

Figure S6. Full representation of the maximum-likelihood phylogenetic tree of the G2BD gene family.

Supplementary Dataset 1. DNA segments of *S* and *s* alleles at the two insertion sites of the hemizygous DSI region

Supplementary Dataset 2 (separate pdf file). Variation in coverage ratio between incompatibility groups along each anchored scaffold in four olive genome assemblies

Supplementary Material S1. Test of the C6030 marker on a few Mediterranean and African brown olives of the Centre d'Ecologie Fonctionelle et Evolutive (CEFE) collection.

In a preliminary study, we first tested the DSI marker ('C6030') developed by Mariotti *et al.* (2020) that should allow distinguishing both self-incompatibility groups in Mediterranean olive. We analyzed a few trees of our collection for which self-incompatibility phenotype was determined based on realized matings through parentage analyses from seeds or seedlings (Besnard *et al.,* 2020).

Material and methods

Eight genotypes were first tested with the marker 'C6030'. We tested four genotypes of *Olea europaea* L. subsp. *europaea* (Mediterranean Olive; hereafter "*europaea*") and four genotypes of *O. e.* subsp. *africana* (Mill.) P.S. Green (African brown olive; hereafter "*africana*"). For *europaea*, three *Ss* and one *ss* were selected, while two *Ss* and two *ss* were analyzed on *africana* (Table SM1). In addition, a few selfings were available for the four *europaea* individuals (Besnard *et al.*, 2020) and were also characterized with the same protocol to check the inheritance of C6030 alleles; among the 34 self-progenies, 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 by Mariotti *et al.* (2020); PCR was not successful on some individuals (i.e. all tested accessions of *africana* and nine self-progenies of L4R14). The PCR protocol was slightly modified by decreasing the annealing temperature (T_a) to 55°C, which allowed us to get an amplification on all samples. PCR product sequencing (Sanger) was performed at GenoScreen (Lille).

Table SM1. List of accessions analyzed for the C6030 region, their incompatibility group (from Besnard *et al.*, 2020), their C6030 genotype, and for *europaea* individuals, those of their self-progenies.

Subspecies	Individual	SI group	C6030 genotype	C6030 genotype of self-progenies
europaea	L4R19 ('Cailletier/Frantoio')	*Ss*	S-A/s-c	S-A/s-c (2/4) or S-A (2/4)
	L4R13 ('Arbequina')	*Ss*	S-A/s-b2	S-A/s-b2 (2/2)
	L4R14 ('Koroneiki')	*Ss*	s-b2	s-b2 (16/25) or Cm (9/25)
	L4R17 ('Verdale de Millas')	*ss*	s-b2	s-b2 (3/3)
africana	Campbelltown 28	*Ss*	C1	-
	Grahamstown 5	*Ss*	C1/C2	-
	Campbelltown 27	*ss*	C1	-
	Grahamstown 8	*ss*	C1/C2	-

Results

Among the *europaea* individuals, a novel haplotype (GenBank accession: ON667921) was detected in three individuals. This haplotye 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 according to Mariotti *et al.* (2020)]. It was detected in L4R13 (*Ss*), L4R14 (*Ss*) and L4R17 (*ss*; Table SM1). Unexpectedly, L4R14 (*Ss*) does not show a C6030 profile compatible with a *Ss* genotype as described in Mariotti *et al.* (2020). 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 Mariotti *et al.* (2020). 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. Supporting this hypothesis, the nine selfings for which Cm was observed were finally *SS* according to the DSI markers developed in the present study (see below; Material S2).

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 (Table SM1). 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 marker in our PCR conditions (T_a = 55°C).

Conclusion

Results of this study do not confirm the expectation that the C6030 profile should allow unambiguously determining the self-incompatibility group of each Mediterranean olive tree (e.g. L4R14; Table SM1). In addition, the presence of null alleles [already stated by Mariotti *et al.* (2020)] and/or paralogous sequences should limit the usefulness of this locus in olive, especially at a wide taxonomic scale (i.e. *O. europaea* complex). This calls for the development of a more robust test for determining cross compatibility of olive genotypes. Lastly, to facilitate its routine use, a test based on length variants allowing distinguishing *S* and *s* alleles (rather than Sanger sequences) on a large panel of *O. europaea* populations is required.

Supplementary Material S2. Development of simple PCR-based diagnoses of self-incompatibility groups in olive and the common yellow jasmine

Developing a robust PCR test of the DSI locus (i.e. useful on any olive or common yellow jasmine accession, and able to simultaneously amplify a segment specific to *s* and *S* alleles) requires comparing genetically divergent accessions in order to define primers in conserved sequences. For the olive tree, our test was thus developed and validated on individuals coming from Africa (*O. e.* subsp. *africana*), the Central Sahara [*O. e.* subsp. *laperrinei* (Batt. & Trab.) Cif.], the Mediterranean Basin (*O. e.* subsp. *europaea*) and on a few hybrids. We also characterized self-progenies of *Ss* cultivated olives to test the occurrence and viability of *SS* individuals. For the common yellow jasmine [*Chrysojasminum fruticans* (L.) Banfi], we validated our test on a set of populations recently sampled in southern France (Puyoou *et al.*, 2023).

Material and methods

Plant material

Olive belongs to a species complex (the so-called *O. europaea* complex) that diversified since the Late Miocene, about 6 to 8 Mya (Green, 2002; Besnard *et al.*, 2009). All wild olives are native to the Old World (from the Mediterranean Basin, the Canary and Madeira archipelagos, Tropical Africa and southern Asia; Green, 2002). In the collection at the common garden of the "Plateforme des Terrains d'Experience du LabExCeMEB," (CEFE, CNRS, Montpellier, France), we currently maintained more than 100 olive accessions belonging to subspecies *europaea* (Mediterranean olive), *laperrinei* (Saharan olive), *africana* (African brown olive), *maroccana* (Moroccan olive, hexaploid) and hybrids. Sixty-four nine trees were phenotyped for their self-incompatibility group (*Ss* = G1 or *ss* = G2; following the approach described in Besnard *et al.*, 2020) and retained for this study (listed in Table S1a; with paternity tests on their seeds/seedlings in Table S1b). They include 38 individuals of *laperrinei*, 16 *europaea* (i.e. 12 cultivars and four oleasters), 11 *africana*, and four hybrids (i.e. "*laperrinei* x *cuspidata*", "*cuspidata* x *europaea*", "*africana* x *europaea*", and "*laperrinei* x *africana*"). Thirty trees are *Ss*, while the 39 remaining are *ss*. In addition, the 34 self-progenies mentioned above (Material S1) were also characterized with our DSI markers, in order to test the presence of *SS* individuals in *Ss* progenies [issued from L4R19 ('Cailletier/Frantoio'), L4R13 ('Arbequina'), and L4R14 ('Koroneiki')].

For the common yellow jasmine, we analyzed 179 individuals from 13 French populations for which the floral morph [L = Longistylous (90 individuals) or B = Brevistylous (89 individuals)] was noted during the flowering (Puyoou *et al.*, 2023).

Designing of primers to genotype the DSI locus in olive

After aligning the two 18-chromosome haplocontigs of the *Ss* Saharan olive (h2tg000033l and h1tg000037l), extremities of the hemizygous region were located (positions 1'658'780 and 2'393'593 in halpocontig h1tg000037l). On each extremity, a DNA segment (ca. 700-1200 bp) of both *S* and *s* alleles covering the insertion site of the hemizygous region were isolated and then used as references (full sequences of *europaea* and *laperrinei* are given in Data S1). HiSeq reads genome skimming data available for two olive *Ss* accessions of *europaea* and *africana* were mapped on these two alleles in order to compare these regions among distantly related olive accessions. Three primers were first defined in short conserved regions of *S*

and/or *s* alleles at the 5' extremity (DSI-A; Table SM2-1; Figure). Primer S/s_A_For (including a M13 tail in 5') was defined in a conserved region on both *S* and *s* alleles (less than 40 bp before the insertion of the hemizygous region). In contrast, one reverse primer was defined specifically for each *s* and *S* alleles - s_A(1)_Rev and S_A_Rev - the latter being located in the hemizygous region. Expected size of PCR products was 441-451 and 298-333 bp for the *s* and *S* alleles, respectively. Length variation was due to a 4-bp-microsatellite (TTAT motif) and a 10-bp indel in the *s* allele, while a 35-bp indel was observed in the *S* alleles between *europaea* and *africana/laperrinei*. In addition, we defined an alternative reverse primer to amplify a shorter segment of alleles *s*: s_A(2)_Rev. This primer can be used instead of s_A(1)_Rev to generate a *s* fragment of 186 bp in all studied olive subspecies. On the 3' extremity (DSI-B), we then used the same approach and designed two primers [S/s_B_For (including a M13 tail in 5') and S/s_B_Rev] to amplify a segment of 190-191 and 185 bp for the *s* and *S* alleles, respectively (Table SM2-1). Finally, we amplified specifically a short segment of the G2BD-S gene (Table SM2-1). Primers were designed in a region conserved between G2BD-S homologs of distantly related species [i.e. *O. europaea* (Oleeae) and *C. fruticans* (Jasmineae)], while as divergent as possible among their olive paralogs.

Designing of primers to genotype the DSI locus in the common yellow jasmine

For the jasmine, we aligned contigs xxx and xxx, to delimitate extremities of the hemizygous region (positions xxx in haplocontig xxx). Following a similar approach to olive, we designed primers to amplify at each extremity a region shared by both alleles s and S, and showing indels: Cf-DSI-A (125 bp for S, and 133 bp for s) and Cf-DSI-B (226 bp for S, 190 bp for s). We also developed a marker linked to the BZR1-S gene (Cf-BZR1-S) to amplify a S fragment of about 170 bp in brevistylous individuals (Table SM2-1).

Laboratory procedures

For olive DSI-A and DSI-B, *S* and *s* alleles of the DSI locus were amplified by PCR following the method described by Schuelke (2000). 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 ($^{5'}$TGTAAAACGACGGCCAGT$^{3'}$), 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 (Table SM2-1)], 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-labelled M13(21) primer, 0.1 µmol of the S/s_B_For primer, and 0.2 µmol of the S/s_B_Rev primer (Table SM2-1). 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 v.9 (Kearse *et al.*, 2012). For jasmine Cf-DSI-A, Cf-DSI-B, and Cf-BZR1-S, we used the same protocol as for olive DSI-B.

Furthermore, the G2BD-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 (Table SM2-1), 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.

Results and discussion

All targeted regions were all successfully amplified and generated *S* and *s* alleles at the expected size.

Different *s* size variants were revealed in olive on DSI-1(A) (4 alleles) and DSI-B (3 alleles; Table SM2-2). For the *S*-haplotype, DSI-A(1)_298 (*europaea*) or _332 (*laperrinei-africana*), DSI-B_185, and G2BD-S_168 are revealed only in *Ss* individuals (Table SM2-2). 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 (Table S1b). 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.

For the jasmine, alleles Cf-DSI-A_125, Cf-DSI-B_226, Cf-BZR1-S_171(or 173) and G2BD-S_168 co-segregate and are revealed only in brevistylous individuals (Table SM2-3) as expected for the *S*-haplotype. Allelic variants (4 alleles, and possibly a null allele on one individual) were also observed on locus Cf-DSI-B for the *s* haplotype.

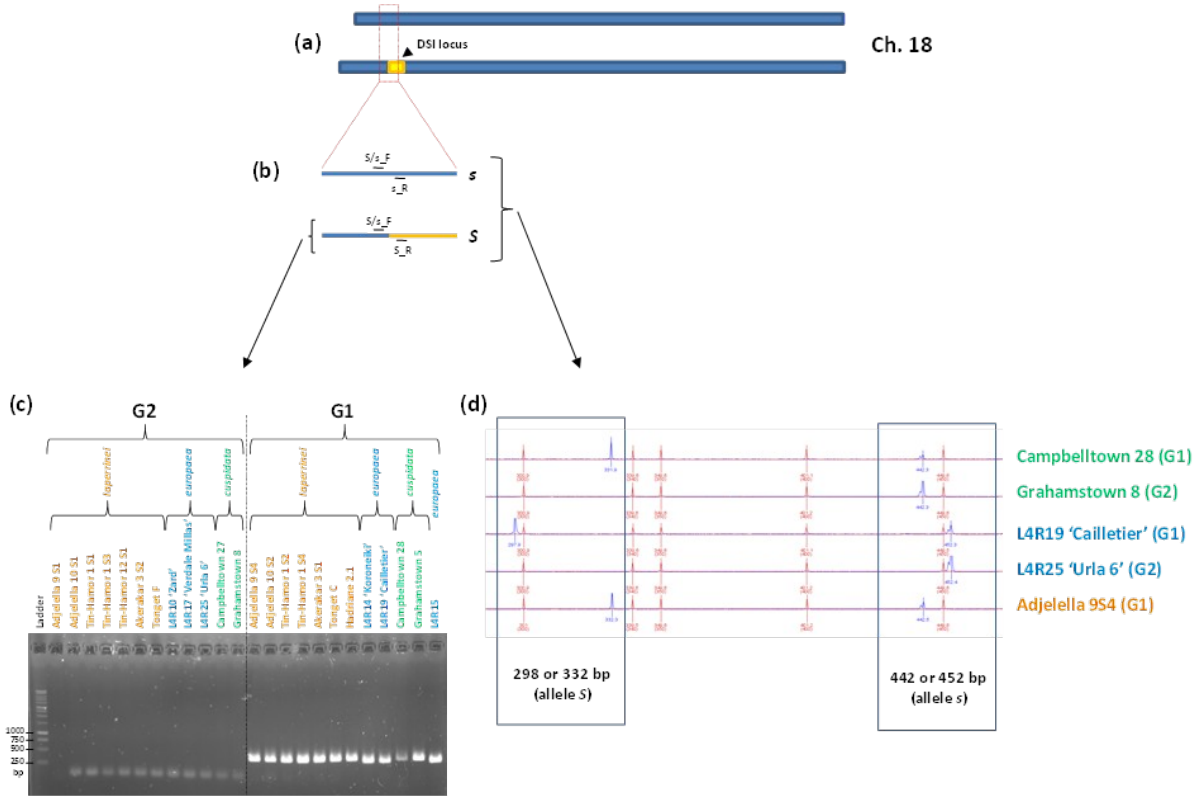


Figure. General approach to develop a PCR test aiming at determining the self-incompatibility group in olive (here for marker DSI-A). **(a)** Schematic representation of the pair of chromosomes 18 in a *Ss* individual (expected to harbor the *Ss* genotype at the DSI locus; Saumitou-Laprade *et al.*, 2010). The DSI region (in yellow) is hemizygous on ca. 730 kb and is present only in *Ss* individuals. It thus corresponds to allele *S*, while the absence of this region corresponds to allele *s*. **(b)** On the 5' extremity covering the hemizygous region insertion site, three primers were defined in short conserved regions

of *S* and/or *s* alleles. S/s_A_F was defined in a conserved region on both *S* and *s* alleles (less than 40 bp before the insertion of the hemizygous region), while one reverse primer was defined specifically for each *S* and *s* alleles. **(c)** Electrophoresis (on 2% agarose) of the PCR product obtained with S/s_A_For and S_A_Rev on 12 *ss* and 12 *Ss* trees belonging to subspecies *laperrinei*, *europaea* and *africana*. A fragment of 298 or 332 bp is revealed only in all *Ss* individuals (for *europaea* or *laperrinei-africana*, respectively), while no amplification is observed in *ss* individuals. **(d)** Chromatograms of the PCR product of five individuals after the simultaneous amplification of both *S* and *s* alleles in the same reaction with S/s_A_For, s_A(1)_Rev, and S_A_Rev. One *S* allele is detected only in *Ss* individuals at an expected size of 298 or 332 bp, whereas *s* alleles are detected in all individuals at 442 or 452 bp.

Table SM2-1. Primers used to amplify extremities of the DSI locus (DSI_A and DSI_B), and a segment of the G2BD-S gene or BZR1-S genes in olives and/or *Chrysojasminum fruticans*. PCR product size is given in bp for *S* and *s* haplotypes of the *S* locus (see Tables SM2 and SM3 for genotypic profiles in olive and common yellow jasmine, respectively).

Locus	**For (5' → 3')**	**Rev (5' → 3')**	**Allele size (bp)**
DSI-A(1)	S/s_A: *GAACAAACAATGAGCTTGTATTTC	s_A(1): CCCAACAATTCCACGTTGCT S_A: TGGTAAAAGGATATCATTGTTACC	*S* = 298, 332; *s* = 438, 442, 446, 452
DSI-A(2)	S/s_A: *GAACAAACAATGAGCTTGTATTTC	s_A(2): TTACCCTAGTACATTAACTTTCCA S_A: TGGTAAAAGGATATCATTGTTACC	*S* = 332; *s* = 186
DSI-B	S/s_B: *STATTTTAGGAGACAAACATCCA	S/s_B: ATCGAAATCAACATATGTGCATCAA	*S* = 185; *s* = 190, 191
G2BD-S	TGCAAGCCCATTTCAASCAYYTTC	CTTAWTGGCTTCACATTCCAATCTA	*S* = 168; *s* = null
Cf-BZR1-S	ACTACGTTTATCCTCTCGTACT	*GGAAACATTGTGAAAAGTTGTAAC	*S* = 171, 173; *s* = null
Cf-DSI-A	S/s_A: *TGCTGGTCACACTGGATTC	S/s_A: ATAGAACAAGCCCAAGTGTTCA	*S* = 125; *s* = 133

*A M13 tail (tgtaaaacgacggccagt) was added in 5'

Table SM2-2. Genotypic profiles at three loci of the DSI region observed in three olive subspecies and hybrids (i.e. 69 trees from the CEFE collection), and their associated incompatibility group. As expected for the *S*-haplotype, DSI-A(1)_298/332, DSI-B_185, and G2BD-S_168 (all in bold and underlined), co-segregate and are revealed only in *Ss* individuals. N = Number of trees.

Taxon	SI group	DSI genotypes			N
		DSI-A(1)	G2BD-S	DSI-B	
europaea (16)	*Ss*	**298**:452	**168**	**185**:190	4
	ss	452:452	0	190:190	8
		446:446	0	191:191	2
		446:452	0	190:191	1
		438:452	0	190:191	1
laperrinei (38)	*Ss*	**332**:442	**168**	**185**:191	19
	ss	442:442	0	191:191	19
africana (11)	*Ss*	**332**:442	**168**	**185**:190	5
	ss	442:442	0	190:190	6
Hybrids (4)	*Ss*	**298**:452	**168**	**185**:190	1
		332:452	**168**	**185**:190	1
	ss	442:442	0	191:191	1
		442:452	0	190:191	1

Table SM2-3. Genotypic profiles at four loci of the DSI region observed in 179 individuals of *Chrysojasminum fruticans*, and their associated floral morph. Alleles of the *S*-haplotype, Cf-DSI-A_125, Cf-DSI-B_226, G2BD-S_168 and Cf-BZR1-171/173 (all in bold and underlined), co-segregate and are revealed only in *Ss* individuals. N = Number of genotypes.

Floral morph	DSI genotypes				N
	Cf-DSI-A	G2BD-S	Cf-BZR1-S	Cf-DSI-B	
Brevistylous	**125**:133	**168**	**171**	190:**226**	28
(*Ss*)	**125**:133	**168**	**173**	190:**226**	1
	125:133	**168**	**171**	199:**226**	59
	125:133	**168**	**171**	202:**226**	1
	125:133	**168**	**171**	0:**226**	1
Longistylous	133:133	0	0	190:190	25

(*ss*)	133:133	0	0	190:199	18
	133:133	0	0	199:199	44
	133:133	0	0	199:202	1
	133:133	0	0	199:214	1

Table S2. RAD-sequencing and mapping statistics on the oleaster genome.

Sample	Raw reads	Clean reads	Mapping rate (%)	Sample	Raw reads	Clean reads	Mapping rate (%)
Adjelella 10_S10	10048146	10011485	82.0	Kirst_1	21128268	21058689	82.2
Adjelella 10_S11	16924894	16840808	80.0	Kirst_2	18835768	18732614	80.7
Adjelella 10_S12	11666302	11583285	80.1	Kirst_4	21805072	21708485	82.0
Adjelella 10_S9	14910068	14866348	80.4	Koroneiki (L4R14)	9141566	9092291	80.5
Adjelella 9_S1	25275302	25107646	80.7	Manzanilla (L4R11)	24268672	24114103	82.1
Adjelella 9_S4	27142444	27027369	82.1	Cailletier /Frantoio (L4R19)	20343488	20287882	80.2
Akerekar 3_S1	11921620	11885430	82.1	Tin-Hamor1_S1	18233404	18162892	80.2
Al_Ascharinah_9 (L4R25)	17485822	17415930	83.0	Tin-Hamor1_S10	18628334	18553433	81.5
Arbequina (L4R14)	13088002	13038573	80.1	Tin-Hamor1_S11	15629066	15572008	80.0
Verdale de Millas (L4R17)	18660040	18573078	83.1	Tin-Hamor1_S14	12274968	12228457	80.4
Ca 27	14619204	14555327	79.9	Tin-Hamor1_S4	16914898	16848388	80.3
Ca 28	12719166	12624495	79.1	Tin-Hamor1_S6	15823176	15751436	78.4
Ca 29	14819554	14688804	76.7	Tin-Hamor1_S7	20462340	20304506	73.9
Picholine (L4R15)	16395184	16354789	82.9	Tonget A	12489254	12410717	77.9
Gr 3	20845246	20740413	79.8	Tonget C	183532	18303637	78.0

					44		
Gr 5	15587608	15520209	79.2	Tonget D	18371078	18306346	79.7
Gr 7	19708102	19625567	81.2	Tonget F	14995700	14944628	79.0
Gr 8	14484088	14397980	80.3	Urla 6 (L4R24)	14229824	14174075	81.7
				Zard (L4R10)	12930964	12879725	84.4

Table S3. List of windows harboring F_{ST} values greater than the ones obtained in any permutation in each olive assembly.

Assembly	Windows coordinates	Assembly	Windows coordinates
	NC_036242.1:15450000-15500000		h1tg000001l:17000000-17050000
	NC_036248.1:25850000-25900000		h1tg000002l:22200000-22250000
	NC_036249.1:8650000-8700000		h1tg000005l:16800000-16850000
	NC_036254.1:16450000-16500000		h1tg000010l:38350000-38400000
	NC_036256.1:1050000-1100000		h1tg000011l:50750000-50800000
Oleaster	NW_019229365.1:0-50000	Haplotype 1	h1tg000013l:9950000-10000000
	NW_019235924.1:0-50000		h1tg000025l:28200000-28250000
	NW_019242006.1:0-50000		h1tg000027l:13350000-13400000
	NW_019249636.1:400000-450000		h1tg000037l:1600000-1650000
	NW_019265623.1:350000-400000		h1tg000037l:2400000-2450000
	NW_019266118.1:200000-250000		h1tg000037l:2450000-2500000
Arbequina	GWHAOPM00000008:19200000-19250000	Haplotype 2	h2tg000012l:6750000-6800000
	GWHAOPM00000010:29900000-29950000		h2tg000019l:13450000-13500000
	GWHAOPM00000010:47300000-47350000		h2tg000038l:1450000-1500000
	GWHAOPM00000012:46000000-46050000		h2tg000049l:900000-950000
	GWHAOPM00000018:17500000-17550000		
	GWHAOPM00000018:18600000-18650000		
	GWHAOPM00000018:19500000-19550000		
	GWHAOPM00000018:19600000-19650000		
	GWHAOPM00000018:22600000-22650000		

Table S4. List of Oleaceae genomic resources used for phylogenetics analyses.

Family	Tribe	Subtribe	Species	Floral phenotype	Biosample/ voucher no/ collection no	Reference	Data type
Oleaceae	Myxopyreae		*Dimetra craibiana* Kerr	?	S. Sudee et al. 2014 (K)		Genome skimming
			Nyctanthes aculeata Craib	?	A.F.G. Kerr 13501 (K)		Genome skimming
	Fontanesieae		*Fontanesia fortunei* Carrière	H,	T. Joßberger 1829 (BONN)		Genome skimming
	Forsythieae		*Abeliophyllum distichum* Nakai	?	M.W. Chase 3881 (K)		Genome skimming
			Forsythia europaea Degen & Bald.	B	C1225 (P; P03868531)		Genome skimming
			Forsythia suspensa (Thunb.) Vahl	?	SAMN26179488	PRJNA809471	Genome
			Forsythia x *intermedia* Zabel	B	G. Besnard 01-2023 (MPU)		Genome skimming
	Jasmineae		*Chrysojasminum bignoniaceum* (Wall. ex G.Don) Banfi	H, SC	G. Besnard 01-2020 (MPU); EDB collection	This study	Genome skimming, transcriptome
			Chrysojasminum fruticans (L.) Banfi	B	CEFE-S6 (CEFE collection)	This study	Genome, transcriptome
			Jasminum annamense Wernham	L	J. Munzinger 106 (P; P00246867)		Genome skimming
			Jasminum didymum G.Forst.	?	L. Barrabé et al. 1312 (P…)		Genome skimming

		Jasminum glaucum (L.f.) Aiton	B	P. Goldblatt et al. 12632 (K)		Genome skimming
		Jasminum mesnyi Hance	B	Liu Zheng-Yu 15432 (P; P03868496)		Genome skimming
		Jasminum pauciflorum Benth.	L	D. Bilivogui 193 (P; P00853672)		Genome skimming
		Jasminum sambac (L.) Aiton	?	SAMN18826951	PRJNA723725	Genome
		Menodora heterophylla Moric. ex DC.	H, ?	R. McVaugh 7744 (P…)		Genome skimming
		Menodora robusta (Benth.) A.Gray	H, ?	M.N. Correa 1166 (P, …)		Genome skimming
Oleeae	Schreberinae	*Schrebera swietinioides* Roxb.	H, ?	W.S. Kurz 2312 (K)		Genome skimming
		Schrebera obconica (Knobl.) Hong-Wa & Besnard	H, SC	C. Mas 98 (P, …)		Genome skimming
		Schrebera trichoclada Welw.	B	P.J. Greenway 15328 (MO)		Genome skimming
	Ligustrinae	*Ligustrum ovalifolium* Hassk.	H, G1	G. Besnard 01-2018 (MPU); CEFE collection		Genome skimming
		Syringa oblata Lindl.	H, ?	SAMC355507	PRJCA005008	Genome
		Syringa persica L.	H, ?	cv. ‘Laciniata’		Genome skimming
	Fraxininae	*Fraxinus ornus* L.	H, ?	SAMEA6544800	PRJEB20151	Genome
		Fraxinus excelsior Marshall	H, ?	SAMN18253904	PRJNA713541	Genome

	Fraxinus quadrangulata Michx.	H, ?	SAMEA6544806	PRJEB20151	Genome
	Chionanthus macrobotrys (Merr.) Kiew	H, ?	S. Collenette 905 (K)		Genome skimming
	Noronhia emarginata (Lam.) Poir.	H, SC	Ste Rose 1 (CEFE collection)		Genome skimming
	Olea europaea L. subsp. *europaea*	H, G1	SAMN05943011	PRJNA350614	Genome
	Olea europaea subsp. *laperrinei* (Batt. & Trab.) Cif.	H, G1	Adjelella 9_S4 (CEFE collection)	This study	Genome
	Olea welwitschii (Knobl.) Gilg & Schellenb.	H, G1	Kakamega 1 (CEFE collection)		Genome
	Osmanthus fragrans Lour.	H, ?	SAMN16954136	PRJNA679852	Genome
Carlemanniaceae	*Silvianthus bracteatus* Hook.f.	?	SAMN15205544	SRR12009658	Transcriptome
Scrophulariaceae	*Buddleja alternifolia* Maxim.		SAMN13022353	PRJNA577174	Genome
Pedaliaceae	*Sesamum indicum* L.		SAMN02981519	PRJNA268358	Genome
Lamiaceae	*Salvia hispanica* L.		SAMN27737700	PRJNA830713	Genome

Table S5. Synonymous divergence (dS) and sequencing depth ratio between *S*-locus genes and closest paralogs. Only values for species where BZR1-S and G2BD-S were both detected are reported.

Species	BZR1	G2BD-I-1	G2BD-I-1-O.A	G2BD-I-1-O.B	Depth ratio DSI/non-DSI
O. europaea	0.50	-	0.405	0.413	-
C. fruticans	0.48	0.58	-	-	-
N. aculeata	0.56	0.48	-	-	4.19
J. nudiflorum	0.54	0.56	-	-	1.93

J. glaucum	0.54	0.54	-	-	1.96
S. oblata	0.63	-	0.72	NA	-
S. trichoclada	0.43	-	0.42	NA	1.54

Table S6. Genomic location of the *S*-locus genes and their sister paralogs. Only species with a chromosome-level assembly available are presented.

Species	BZR1-S	G2BD-S	Distance (Mb)	BZR1	G2BD-I-1	G2BD-I-1-O.A	G2BD-I-1-O.B	Distance (Mb)
O. europaea	Chr18	Chr18	0.23	Chr21	-	Chr15	Chr21	0.96
O. fragrans	x	x	-	Chr22	-	Chr10	Chr22	0.85
F. excelsior	x	x	-	Chr2	-	Chr22	Chr2	0.75
S. oblata	Chr9	Chr9	0.14	Chr22	-	Chr23	NA	NA
J. sambac	x	x	-	Chr11	Chr11	-	-	1.10
C. fruticans	Chr2	Chr2	0.40	Chr11	Chr11	-	-	2.50
F. suspensa	Chr1	x	-	unanchored	Chr1	-	-	NA
S. indicum	x	x	-	Chr12	Chr12	-	-	0.40
B. alterniflora	x	x	-	Chr3	Chr3	-	-	0.76
S. hispanica	x	x	-	Chr5	Chr5	-	-	0.40

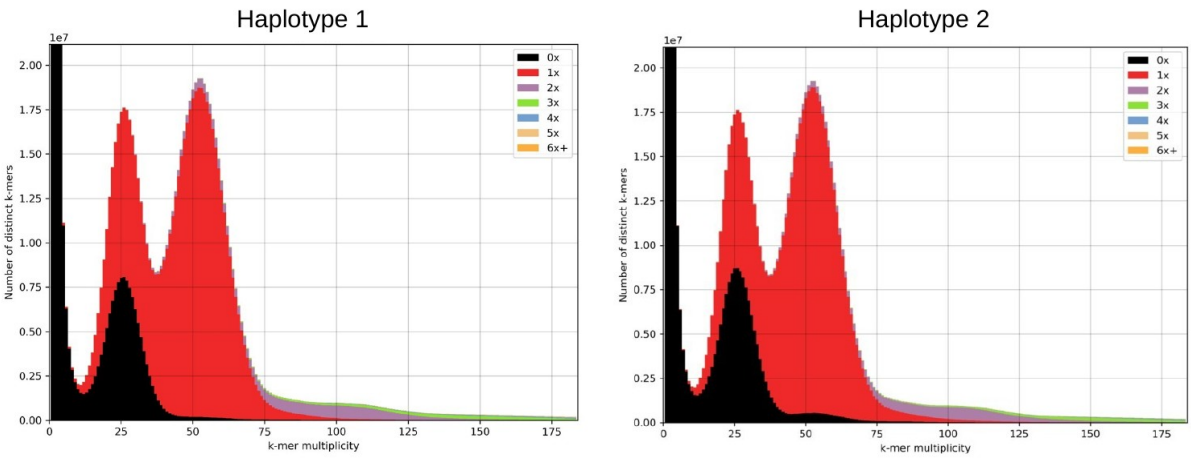


Figure S1. *k*-mer spectra of the two Saharan olive haplotype assemblies. Red curves correspond to *k*-mer present in one copy in the assembly (the two peaks at 25× and 50× reflect the high heterozygosity of the genome). Black curve represents *k*-mer in reads that were not included in the assembly, either erroneous (at the extreme boundary of the plot) or because of allele separation (peak half the size of the heterozygous peak). Other colors indicate higher *k*-mer multiplicity probably due to ancient whole-genome duplications.

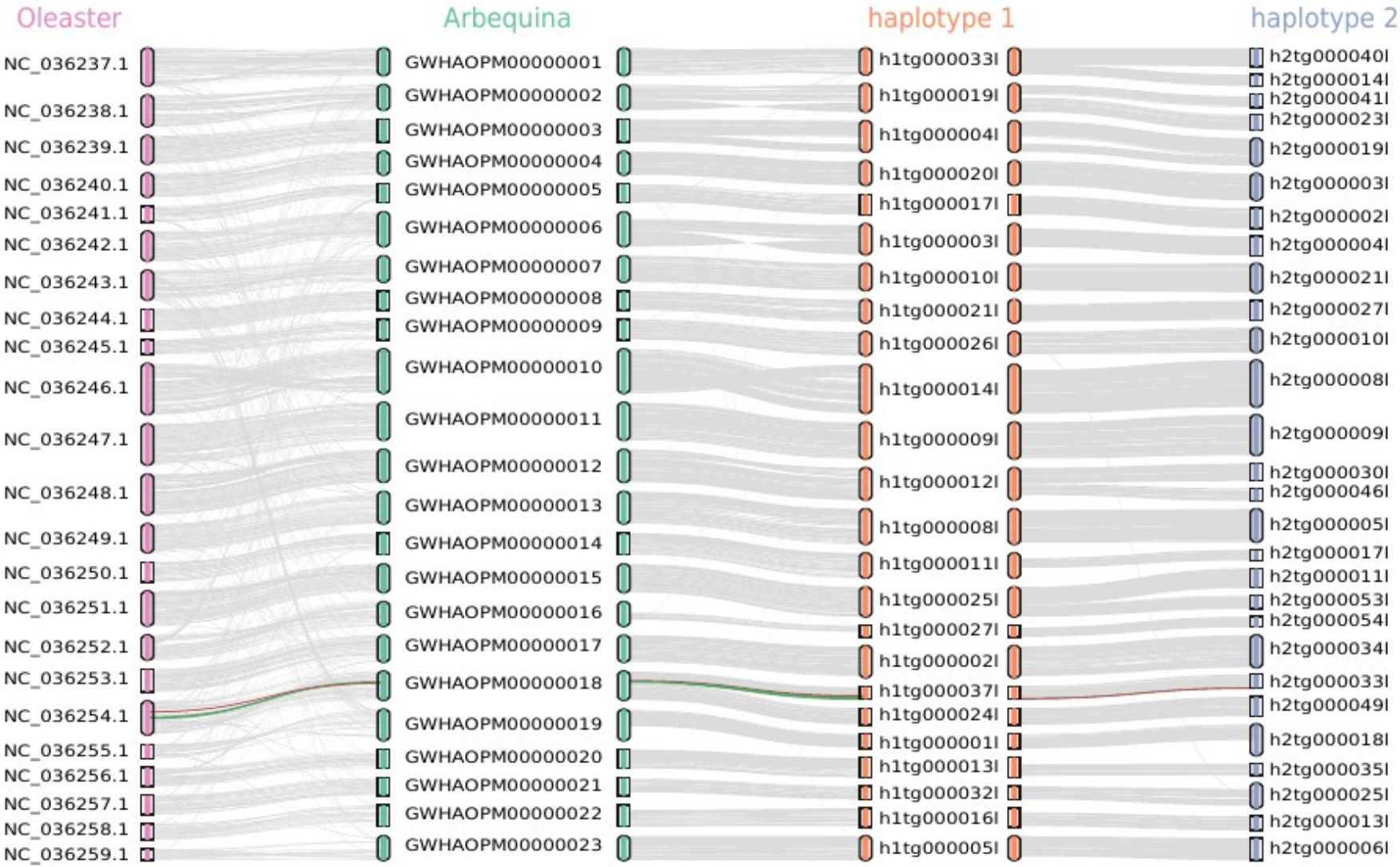


Figure S2. Macrosynteny between four olive genome assemblies. Lines connect syntenic blocks in different assemblies. Red lines correspond to the DSI locus according to Mariotti *et al.* (2020), green lines to the newly identified SI-linked region, which has no homolog in the 'haplotype 2' assembly. Only the largest anchored scaffolds are shown for each assembly. 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.* (2020), but 8 Mb apart in the oleaster assembly.

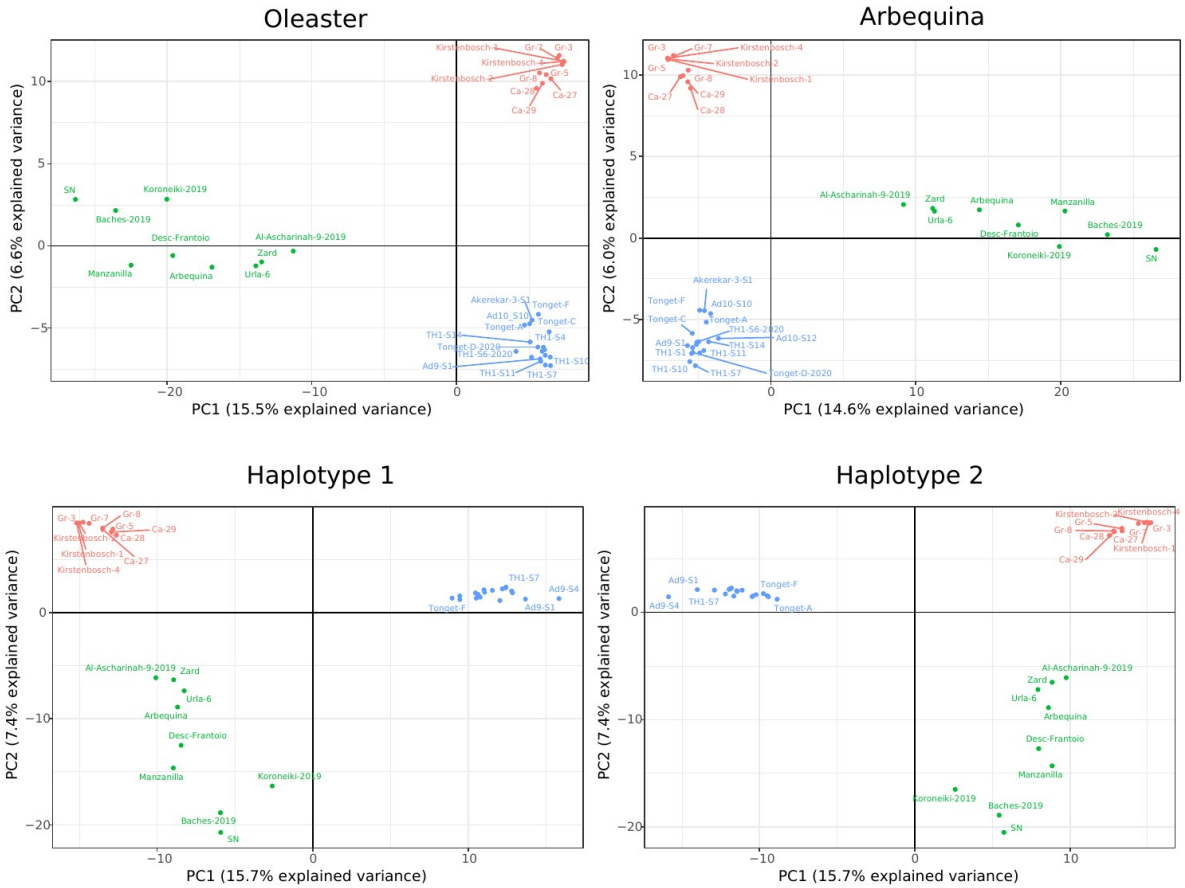


Figure S3. PCA component analysis plot of sequenced olive individuals based on bi-allelic nuclear SNPs called on four different genome assemblies. Each subspecies is figured in a different color. Blue = *Olea e. europaea*, green = *O. e. laperrinei*, orange = *O. e. africana*. Number of SNPs used for each genome: 'Oleaster' 43,525 / 'Arbequina' 41,508 / 'Haplotype 1' 48,623 / 'Haplotype 2' 48,623.

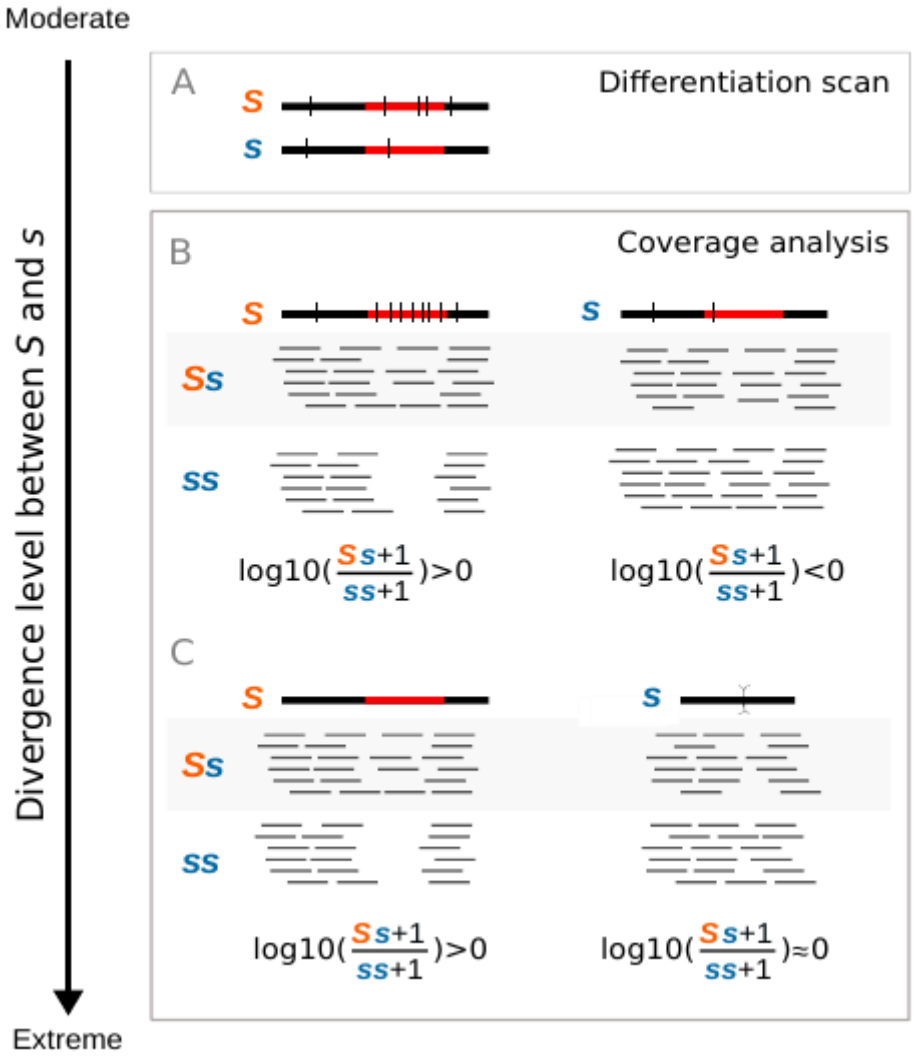


Figure S4. Potential genomic footprints of the *S*-locus depending on its architecture. A) Due to the impossibility of recombination, mutations accumulate on the ever heterozygous *S* allele. This accumulation is detectable using differentiation scans as the mapping success is not affected by this moderate level of divergence. B and C) The divergence between the two alleles causes mapping success difference at the *S*-locus. Both groups will however have equal mapping rates along the black part. Depth of sequencing coverage analysis will detect this change. Different profiles can be expected depending on the reference allele used for mapping and whether the two alleles are still homomorphic (B) or not (C). B) Reads from s allele do not map to the *S* reference leading to a reduced coverage at the locus in *ss* individuals. On the other hand, half of the reads from *Ss* individuals (corresponding to the *S* allele) will map successfully at the locus. The coverage is yet reduced compared to the flanking regions, as the other half of the sequencing reads (corresponding to the s allele) will not map. The coverage ratio between *Ss* and ss individuals will be greater than 0. Conversely, if the reference is *s*, unsuccessful mapping of *S* reads will lead to a reduced coverage at the locus in *Ss* individuals, compared to ss individuals, as well as compared to the flanking regions. The coverage ratio between *Ss* and *ss* individuals will be smaller than 0. C) In case of heteromorphic *S*-locus, the coverage profile will be similar to the one in B when considering *S* as a reference. If the reference is *s*, a coverage reduction in *Ss* individuals will appear at the border of the indel sequence due to the misalignment of reads mapping the junction. However, it will be hyper-localized. The coverage ratio between *Ss* and *ss* individuals will then be roughly equal.

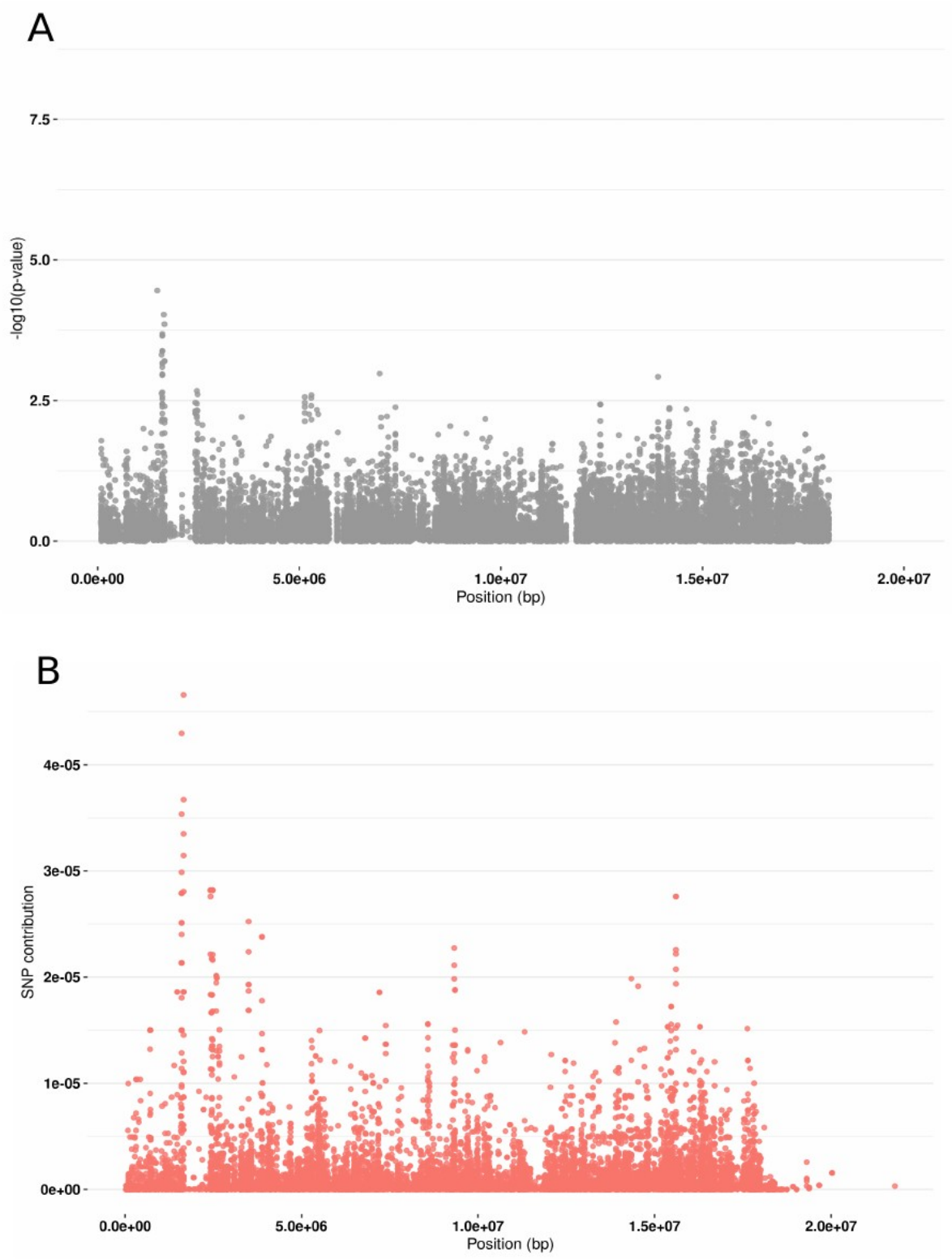
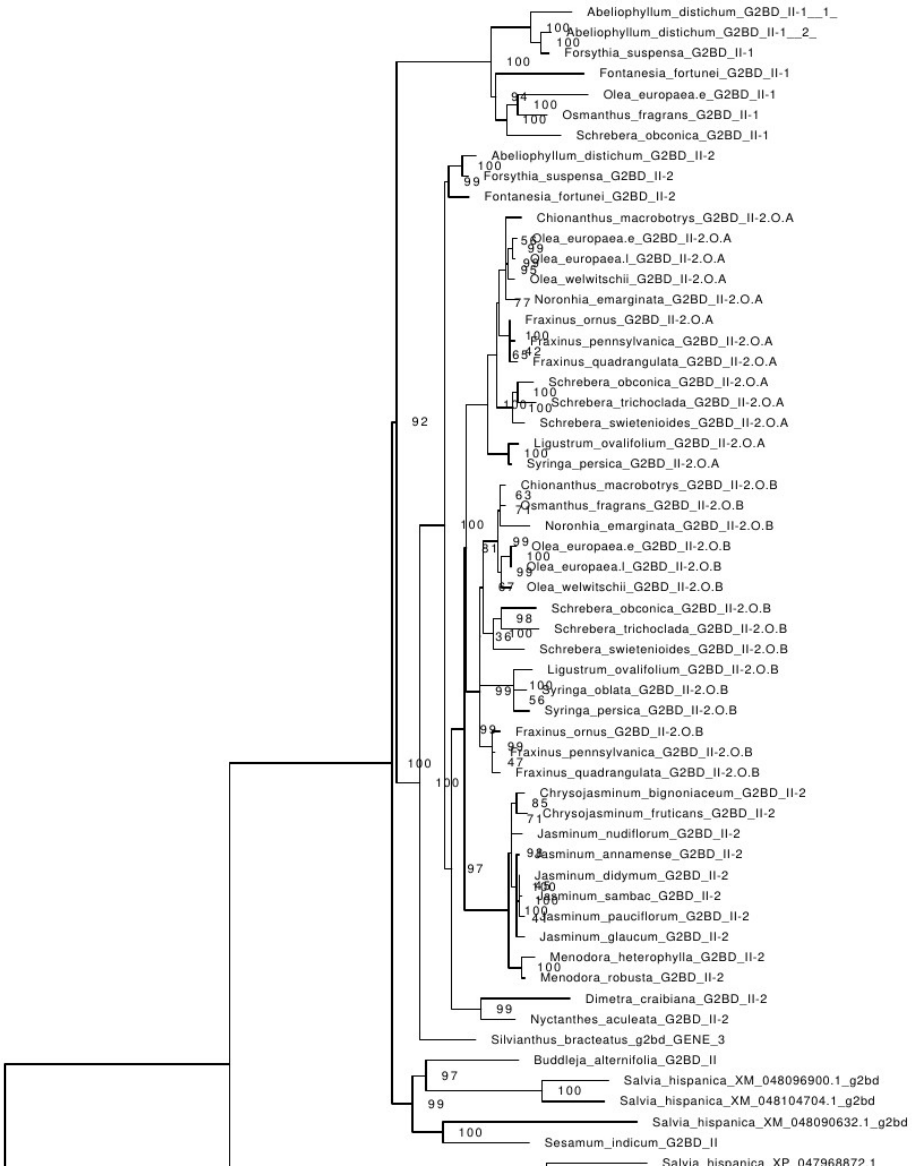
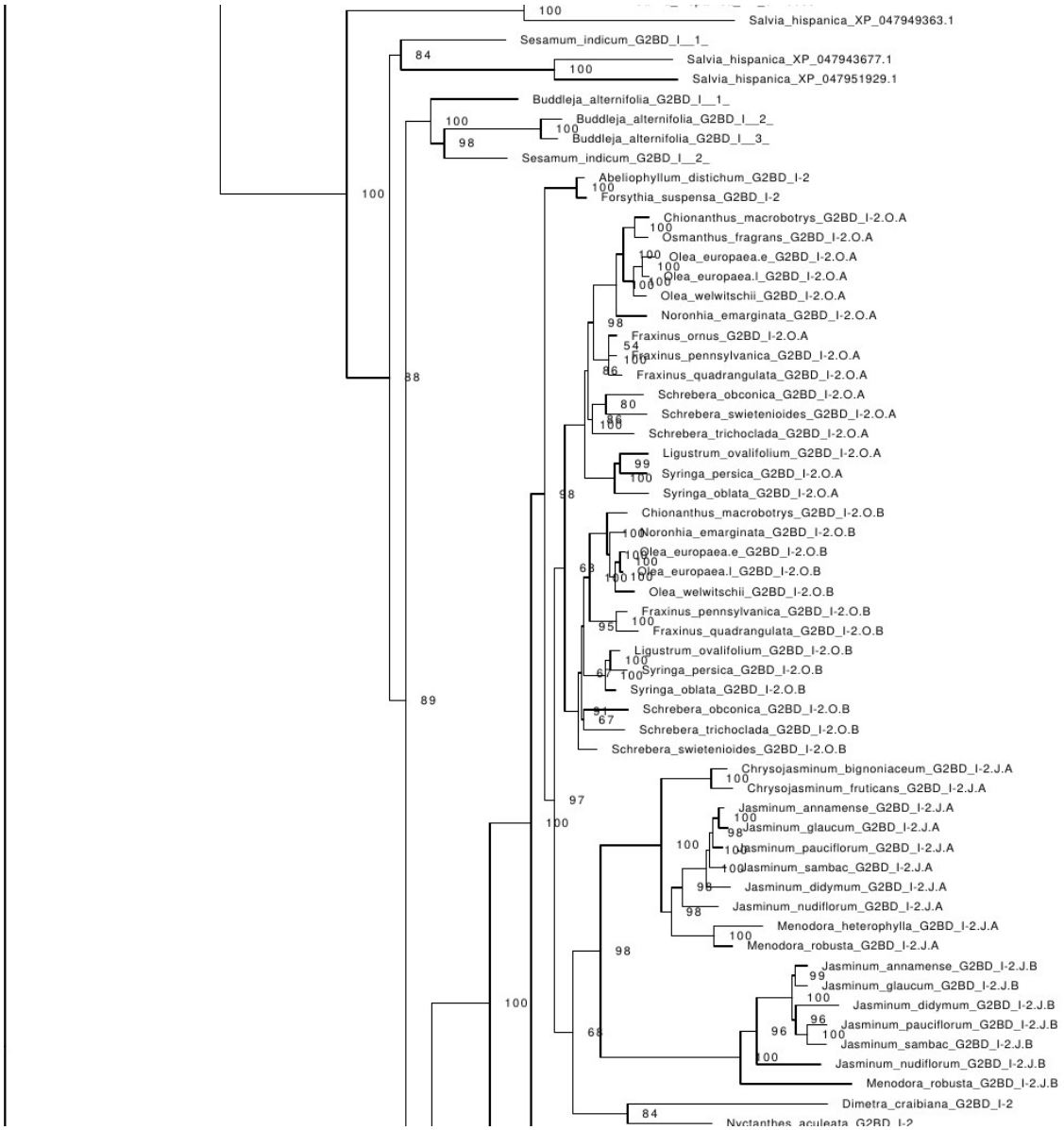


Figure S5. Manhattan plots depicting the association between SNP genotype and compatibility group for olive chromosome 18. Analyses were conducted on the SNP set generated for the 37 individuals mapped on the Saharan olive genome. A) Genome-wide association results showing peaks of phenotype-associated SNPs surrounding the hemizygous region (1.6-2.3 Mbp) B) Discriminant analysis of principal components results revealing SNPs strongly contributing to compatibility groups separation surrounding the hemizygous region (1.6-2.3 Mbp).


Abeliophyllum_distichum_G2BD_II-1__1_
Abeliophyllum_distichum_G2BD_II-1__2_
Forsythia_suspensa_G2BD_II-1
Fontanesia_fortunei_G2BD_II-1
Olea_europaea.e_G2BD_II-1
Osmanthus_fragrans_G2BD_II-1
Schrebera_obconica_G2BD_II-1
Abeliophyllum_distichum_G2BD_II-2
Forsythia_suspensa_G2BD_II-2
Fontanesia_fortunei_G2BD_II-2
Chionanthus_macrobotrys_G2BD_II-2.O.A
Olea_europaea.e_G2BD_II-2.O.A
Olea_europaea.l_G2BD_II-2.O.A
Olea_welwitschii_G2BD_II-2.O.A
Noronhia_emarginata_G2BD_II-2.O.A
Fraxinus_ornus_G2BD_II-2.O.A
Fraxinus_pennsylvanica_G2BD_II-2.O.A
Fraxinus_quadrangulata_G2BD_II-2.O.A
Schrebera_obconica_G2BD_II-2.O.A
Schrebera_trichoclada_G2BD_II-2.O.A
Schrebera_swietenioides_G2BD_II-2.O.A
Ligustrum_ovalifolium_G2BD_II-2.O.A
Syringa_persica_G2BD_II-2.O.A
Chionanthus_macrobotrys_G2BD_II-2.O.B
Osmanthus_fragrans_G2BD_II-2.O.B
Noronhia_emarginata_G2BD_II-2.O.B
Olea_europaea.e_G2BD_II-2.O.B
Olea_europaea.l_G2BD_II-2.O.B
Olea_welwitschii_G2BD_II-2.O.B
Schrebera_obconica_G2BD_II-2.O.B
Schrebera_trichoclada_G2BD_II-2.O.B
Schrebera_swietenioides_G2BD_II-2.O.B
Ligustrum_ovalifolium_G2BD_II-2.O.B
Syringa_oblata_G2BD_II-2.O.B
Syringa_persica_G2BD_II-2.O.B
Fraxinus_ornus_G2BD_II-2.O.B
Fraxinus_pennsylvanica_G2BD_II-2.O.B
Fraxinus_quadrangulata_G2BD_II-2.O.B
Chrysojasminum_bignoniaceum_G2BD_II-2
Chrysojasminum_fruticans_G2BD_II-2
Jasminum_nudiflorum_G2BD_II-2
Jasminum_annamense_G2BD_II-2
Jasminum_didymum_G2BD_II-2
Jasminum_sambac_G2BD_II-2
Jasminum_pauciflorum_G2BD_II-2
Jasminum_glaucum_G2BD_II-2
Menodora_heterophylla_G2BD_II-2
Menodora_robusta_G2BD_II-2
Dimetra_craibiana_G2BD_II-2
Nyctanthes_aculeata_G2BD_II-2
Silvianthus_bracteatus_g2bd_GENE_3
Buddleja_alternifolia_G2BD_II
Salvia_hispanica_XM_048096900.1_g2bd
Salvia_hispanica_XM_048104704.1_g2bd
Salvia_hispanica_XM_048090632.1_g2bd
Sesamum_indicum_G2BD_II


Salvia_hispanica_XP_047949363.1
Sesamum_indicum_G2BD_I__1_
Salvia_hispanica_XP_047943677.1
Salvia_hispanica_XP_047951929.1
Buddleja_alternifolia_G2BD_I__1_
Buddleja_alternifolia_G2BD_I__2_
Buddleja_alternifolia_G2BD_I__3_
Sesamum_indicum_G2BD_I__2_
Abeliophyllum_distichum_G2BD_I-2
Forsythia_suspensa_G2BD_I-2
Chionanthus_macrobotrys_G2BD_I-2.O.A
Osmanthus_fragrans_G2BD_I-2.O.A
Olea_europaea.e_G2BD_I-2.O.A
Olea_europaea.l_G2BD_I-2.O.A
Olea_welwitschii_G2BD_I-2.O.A
Noronhia_emarginata_G2BD_I-2.O.A
Fraxinus_ornus_G2BD_I-2.O.A
Fraxinus_pennsylvanica_G2BD_I-2.O.A
Fraxinus_quadrangulata_G2BD_I-2.O.A
Schrebera_obconica_G2BD_I-2.O.A
Schrebera_swietenioides_G2BD_I-2.O.A
Schrebera_trichoclada_G2BD_I-2.O.A
Ligustrum_ovalifolium_G2BD_I-2.O.A
Syringa_persica_G2BD_I-2.O.A
Syringa_oblata_G2BD_I-2.O.A
Chionanthus_macrobotrys_G2BD_I-2.O.B
Noronhia_emarginata_G2BD_I-2.O.B
Olea_europaea.e_G2BD_I-2.O.B
Olea_europaea.l_G2BD_I-2.O.B
Olea_welwitschii_G2BD_I-2.O.B
Fraxinus_pennsylvanica_G2BD_I-2.O.B
Fraxinus_quadrangulata_G2BD_I-2.O.B
Ligustrum_ovalifolium_G2BD_I-2.O.B
Syringa_persica_G2BD_I-2.O.B
Syringa_oblata_G2BD_I-2.O.B
Schrebera_obconica_G2BD_I-2.O.B
Schrebera_trichoclada_G2BD_I-2.O.B
Schrebera_swietenioides_G2BD_I-2.O.B
Chrysojasminum_bignoniaceum_G2BD_I-2.J.A
Chrysojasminum_fruticans_G2BD_I-2.J.A
Jasminum_annamense_G2BD_I-2.J.A
Jasminum_glaucum_G2BD_I-2.J.A
Jasminum_pauciflorum_G2BD_I-2.J.A
Jasminum_sambac_G2BD_I-2.J.A
Jasminum_didymum_G2BD_I-2.J.A
Jasminum_nudiflorum_G2BD_I-2.J.A
Menodora_heterophylla_G2BD_I-2.J.A
Menodora_robusta_G2BD_I-2.J.A
Jasminum_annamense_G2BD_I-2.J.B
Jasminum_glaucum_G2BD_I-2.J.B
Jasminum_didymum_G2BD_I-2.J.B
Jasminum_pauciflorum_G2BD_I-2.J.B
Jasminum_sambac_G2BD_I-2.J.B
Jasminum_nudiflorum_G2BD_I-2.J.B
Menodora_robusta_G2BD_I-2.J.B
Dimetra_craibiana_G2BD_I-2

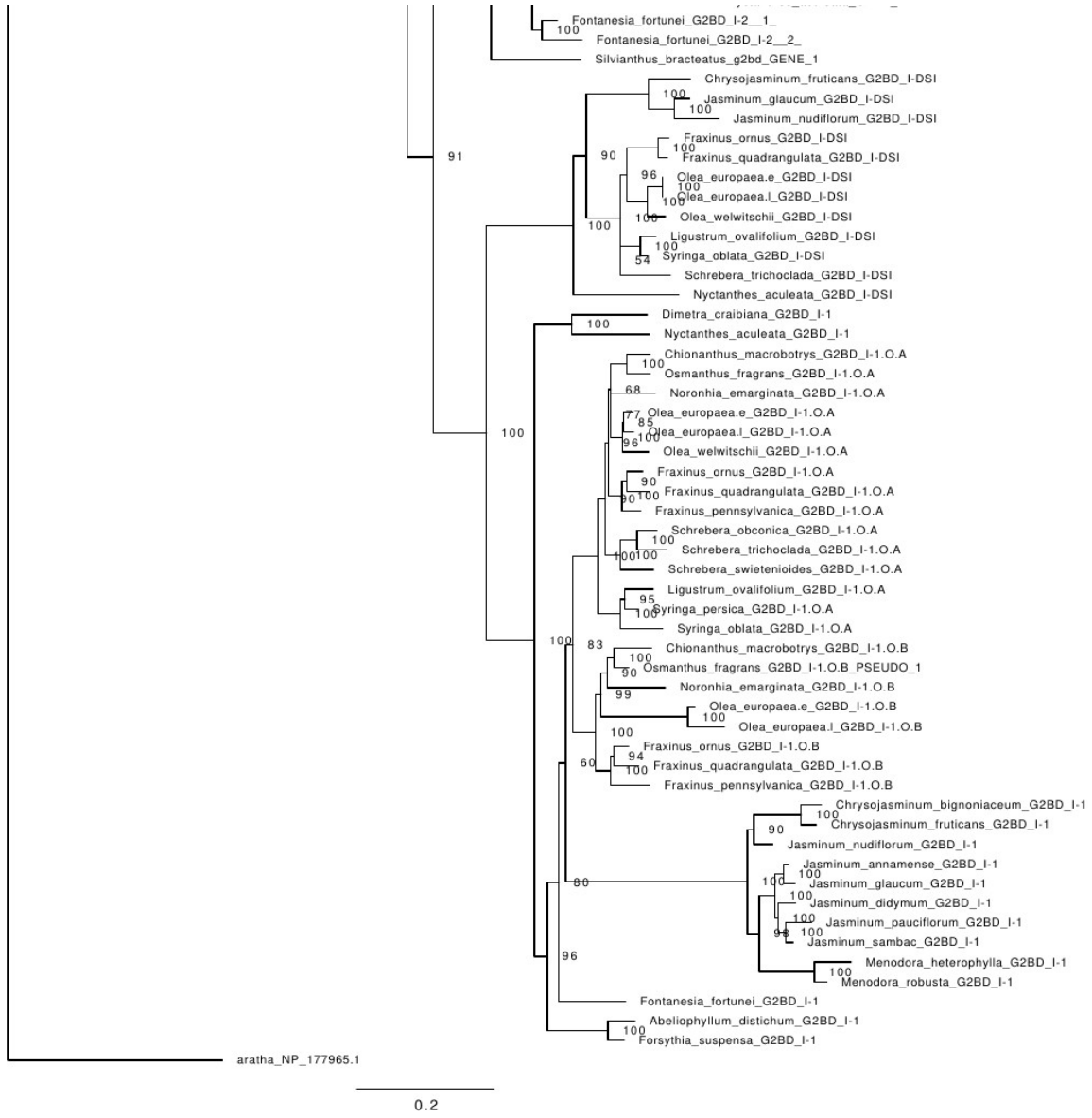


Figure S6. Full representation of the maximum-likelihood phylogenetic tree of the G2BD gene family. *Arabidopsis thaliana* was used as an outgroup to root the tree.

Supplementary Dataset 1. DNA segments of *S* and *s* alleles at the two insertion sites of the hemizygous DSI region [(**A**) Junction DSI-A, and (**B**) Junction DSI-B]. Sequences were isolated from two *Ss* individuals belonging to subspecies *europaea* and *laperrinei*. Location of primers is represented by frames, and the microsatellite motifs of DSI-A-*s* are highlighted in green. In bold, the parts that are shared by *S* and *s* alleles, with the insertion site highlighted in grey.

A) **Junction DSI-A**

>OP779748 - Allele S_*europaea* [Talambote; DSI-A: allele_S_298]
GGCCTCCACTTACCTTGGCATTCAGGGGTTCCATCCACGATTTGAGGAGATCAAAGAATTAATAGAGAGGATGAA
AATTACTCCTGCACAAGTTGCTGAAGAACTAATGAAGAGTGATGATCCTCATACTTCTCTTGAAGGTCTAGTTAA
TTTTTTGAAGTGCAGAAGGATCGAAAGCAACCTATCAAAGGATAAAGGAAACGAAGAAAATGAAGTCGAGAATAT
AAAAAGACTTCCATCTGACAGTGATATTGAAGTCGTGACAGACTAAAATTTATGTGAATTTGATCATGCTTATTA
TGATTTTAAATTAGAATTTGAAATTGTGGTTCTTGATTAAAGATCCAAGACTCTACCTATTGAACAAACAATGAG
CTTGTATTTCTTAATATAAAACTATTTCTAGAATTTTGTCTCATTACAGTAATTTCTAGAGAAGAAGAGCGACCT

CTATCACCATGTTATTTATTATCTCCATCAGTTGTACGTGAAATAATTGGAGTGGAAGGAGCATGACATATGCTA
CTAAAAAATAGCTTTTCAGTTAAGAAAAATTATTATCCGTTGGTAAAAGTAATTTACTTTTAACTATGATAATCA
GCAACGGATTTTTCGTTGGTAACAATGATATCCTTTTACCAATGAATATTCTGTGTCTAATGTAACCGTTTAACA
TTTGAATCTTTCACACGTACC

>OP779749 - Allele S_*laperrinei* [Adjelella_9_S4; DSI-A: allele_S_332]
GGCCTCCACTTACCTTGGCATTCAGGGGTTCCATCCGCGATTTGAGGAGATCAAAGAATTAATAGAGATTATGAA
AATTACTCCTGCACAAGTTGCTGAAGAACTAATGAAGAGTGATGATCCTCATACTTCTCTTGAAGGTCTAGTTAA
TTTTTTGAAGTGCAGAAGGATCGAAAGCAACCTATCAAAGGATAAAGGAAACGAAGAAAATGAAGTCGAGAATAA
AAAAAGACTTCCATCTGACAGTGATATTGAAGTCGTGACAGACTAAAATTTATGTGAATTTGATCATGCTTATTA
TGATTTTAAATTAGAATTTGAAATTGCGGTTCTTGATTAAAGATCCAAGACTCTACCTATCGAACAAACAATGAG
CTTGTATTTCTTAATATAAAACTATTTCTAGAATTTTGTCTCATTACAGTAATTTCTAGAGAAGAAGAGTGACCT
CTATCACCATGTTATTTATTATCTCCATCAGTTGTACGTGAAATAATTGGAGTGGAAGGAGCATGACATACGCTA
CTAAAAAATAGCTTTTCAGCTAAGGAAAATTACCAACGGATACAATCCGTCGGTAATTAGAAACGAATTATCCAT
TGGTAAAAGTAATTTACTTTTAACTATGATAATCAGCAACGAATTTTTCATTGGTAACAATGATATCCTTTTACC
AATGAATATTCTGTGTCTAATGTAACCGTTTAACATTTGAATCTTTCACACGTACC

>OP779750 - Allele s_*europaea* [Talambote; DSI-A: allele_s_446]
GGCCTCCACTTACCTTGGCATTCAGGGGTTCCATCCGCGATTTGAGGAGATCAAAGAATTAATAGAGAGTATGAA
AATTACTCCTGCACAAGTTGCTGAAGAACTAATGAAGAGTGATGATCCTCAAACTTCTCTTGAAGGTCTAGTTAA
TTTTTTGAAGTGCAGAAGGATCGAAAGCAACATATCAAAAGATAAAGGAAACGAAGAAAATGAAGTCGAGAATAT
AAAAAGACTTCCATCTGACAGTGATGTTGAAGGCGTGACAGACTAAAATTTATGTGAATTCGATCATGCTTATTA
TGATTTTAAATTAGAATTTGAAATTGCAGTTCTTGATTAAAGATCGAAGACTCTACCTATTGAACAAACAATGAG
CTTGTATTTCTTAATATAAAACTATTTCTAGGATTTGTCTCATTACGTGCTTTCTTCATTGTAAAATGTAGCGCC
CCGAAAATTTAATAAAATAAAACTATTTGTGCTACTTGTTATAGTTTCATATTTTTGGAAAGTTAATGTACTAGG
GTAAAATTGCGTTATTGGGTCAAATATGCAGATTTCGATAGGACGAAGATAAGTATAACAATATTGAGATTGTTG
GATAGCTCGTAATTAATTATGTATATTAAGTTGGAGAAAAAGTTAATTCGGATAAACTAAACCTAAATTATTGAA
GTTTAAAGTTGTCAATAAATATTATTTTCTTAGTGATTATTTATTTATTTATTTATTATTATTAATTTTTT
TAATTTGTGAAAGTGGGCAGCAACGTGGAATTGTTGGGATGAAATTAAAATGGGTTGCTTTATTTTGTACTAGAT
AGAAGTTTTGGGTAAATTGGACTTTCACTTTCTATAAGACACTTGTCCTATATTTTCATTTATGAAACCAATTTC
CCCAAAGAAAAATTCAGATTTTCTTTCACCATATAGTATAAGTTTATAAACACAGTAGCATGCTGAA

>OP779751 - Allele s_*laperrinei* [Adjelella_9_S4; DSI-A: allele_s_442]
GGCCTCCACTTACCTTGGCATTCAGGGGTTCCATCCGCGATTTGAGGAGATCAAAGAATTAATAGAGAGTATGAA
AATTACTCCTGCACAAGTTGCTGAAGAACTAATGAAGAGTGATGATCCTCAAACTTCTCTTGAAGGTCTAGTTAA
TTTTTTGAAGTGCAGAAGGATCGAAAGCAACATATCAAAGGATAAAGGAAACGAAGAAAATGAAGTCGAGAATAT
AAAAAGACTTCCATCTGACAGTGATGTTGAAGGCGTGACAGACTAAAATTTATGTGAATTCGATCATGCTTATTA
TGATTTTAAATTAGAATTTGAAATTGCAGTTCTTGATTAAAGATCGAAGACTCTACCTATTGAACAAACAATGAG
CTTGTATTTCTTAATATAAAACTATTTCTAGGATTTGTCTCATTACGTGCTTTCTTCATTGTAAAATGTAGCGCC
CCGAAAATTTAATAAAATAAAACTATTTGTGCTACTTGTTATAGTTTCATATTTTTGGAAAGTTAATGTACTAGG
GTAAAATTGCGTTATTGGGTCAAATATGCAGATTTCGATAGGACGAAGATAAGTATAACAATATTGAGCTTGTTG
GATAGCTCGTAATTAATTATGTATATTAAGTTGGAGAAAAAGTTAATTCGGATAAACTAAACCTAAATTATTGAA
GTTTAAAGTTGTCAATAAATATTATTTTCTTAGTGATTATTTATTTATTTATTTATTATTATTAATTTTTTTAAT
TTGTGAAAGTGGGCAGCAACGTGGAATTGTTGGGATGAAATTAAAATGGGTTGCTTTATTTTGTACTAGATAGAA
GTTTTGGGTAAATTGGACTTTCACTTTCTATAAGACACTTGTCCTATATTTTCATTTATGAAACCAATTTCCCCA
AAGAAAAATTCAGATTTTCTTTCACCATATAGTATAAGTTTATAAACAGAGTAGCATGCTGAA

B) **Junction DSI-B**

>OP779752 - Allele S_*europaea* [Talambote; DSI-B: allele_S_185]
TTAAATTGATGCGATGACCAGTCACAGTTTTCTAATGAAAAGAGTGCACATCATAGAGTCATATTGTTTCAAATT
GAATGTGAGTATTCAAAAATCTACATTACGAAAAGTCATGACCTTTTTACAGGTGCAATATAGACCATTAAATAT
AAAATTATGTGATTTAATAATCCTCATTAATTTATGTCAAAAGATCCAATCATCCCAAAAATACAATGGTAAATC
TATAGATAGGAACCCCTAATTCCATATTTACTCTAAAATCAGTTCAAGAAATTTTGGTTTCTAAATTCTCTTATT
TTTTCAAAATAATCTGTTTTTTATTATTTTAGTGAAATTAAATTCCATCCAATTATTATTAGGAATAAAATTCAG
ACCGAACAATTTTAATTCGTTCCTATCCTATTAGTGAAGAATACTATCATCACTTAAATTTGCTGAGAAAAGTAT
TTTCTTGCTTTAATAATTTTGAAAGTGTGTTGTATTTTAGGAGACAAACATCCATTTAATATGTTAAGACCTGTG
GGCAAAATTTGTCTTAAGAAGCTGTACACTCACATACGTCATTTTATTTCATGTCCATCTAATTTTAGCGTTTCA

GTAAATCCCATTTTGTGTACATTTTGATGCACATATGTTGATTTCGATAATTTGAAACTGTATCCAATATTTTGC
ATCAGTGAATCACGATTTTATCATATTTTTATGCATGATTTTTTTATTTATCCAACAATTTTAACACAGTGGCAT
CCAATGATTGCACATAGAGTCTCTCAAGACTTGTTTATTTGGGGATAGTGACTCTTATGACTACTACTAGAGAGA
CACACATGCATCGTACGTGTAGCGTAGATATAAAAATAAATTATTTGTGTTAAGCATTAGTTACTTCAAATTTAA
ACATATGTTGCTAAATGTTTTAAGGCAACAACATTGCAAAAATGATGTATTCAGAACCATACACAACAAACAATT
AGGGATCAAATTCTTGATAAAAAAAAAAAAAAAAAANTGTAATATTACTTTGTGAATGAGACCGTCATCGATTAT
TGAGCTGCTTATTTTCTTAAATATTTCATACAGTGCCTCAACTTCAGTTACGGTGACTGCATGACAATGTTTTCA
ATGAAACTGCATTAGTCCATTCCTTGTCTAGGATC

>OP779753 - Allele S_*laperrinei* [Adjelella_9_S4; DSI-B: allele_S_185]
TTAAATTGATGCGATGACCAGTCACAGTTTTCTAATGAAAAGAGTGCACATCATAGAGTCATATTGTTTCAAATT
GAATGTGAGTATTCAAAAATCTACATTACGAAAAGTCATGACCTTTTTACAGGTGCAATTTAGACCATTAAATAT
AAAATTATGTGATTTAATAATCCTCATTAATTTATGTCAAAAGATCCAATCATCCCAAAAATACAATCGTAAATC
TATAGATAGGAACCCCTAATTCCATATTTACTCTAAAATCAGTTCAAGAAATTTTGGTTTCTAAATTCTCTTATT
TTTTCAAAATAATCTGTTTTTTATTATTTTAGTGAAATTAAATTCCATCCAATTATTATTAGAAATAAAATTCAG
ACCGAACAATTTTAATTCGTTCCTATCCTATTAGTGAAGAATACTATCATCACTTAAATTTGCTGACAAAAGTAT
TTTCTTGCTTTAATAATTTTGAAAGTGTGTTGTATTTTAGGAGACAAACATCCATTTAATATGTTAAGACCTGTG
GGCAAAATTTGTCTTAAGAAGCTGTACACTCACATACGTCATTTTATTTCGTGTCCATCTAATTTTAGCGTTTCA
GTAAATCCCATTTTGTGTACATTTTGATGCACATATGTTGATTTCGATAATTTGAAACTGTATCCAATATTTTGC
ATCAGTGAATCACGATTTTATCATATTTTTATGCATGATTTTTTTATTTATCCAACAATTTTAACACAGTGGCAT
CCAATGATTGCACACAGAGTCTCTCAAGACTTGTTTATTTGGGGATAGTGACTCTTATGACTACTACTAGAGAGA
CACACATGCATCGTACGTGTAGCGTAGATATAAAAATAAATTATTTGTGTTAAGCATTAGTTACTTCAAATTTAA
ACATATGTTGCTAAATGTTTTAAGGCAACAACATTGCAAAAATGATGTATTCAGAACCATACACAACAAATAATT
AGGGATCAAATTCTTGATAAAAAAAAAAAATGTAATATTACTTTGTGAATGAGACCGTCATCGATTATTGGGCTG
CTTATTTTCTTAAATATTTCATACAGTGCCTCAACTTCAGTTACGGTGACTGCATGACAATGTTTTCAATGAAAC
TGCATTAGTCCATTCCTTGTCTAGGATC

>**OP779754 - Allele s_*europaea* [Talambote; DSI-B: allele_s_191]**
TTAAATTTATGCGATGAGTAGTCACGATTTTTCAAATGAATATGGTGCATATCAAAGAGTCAAATTGCTTTGAAT
TGAATGTGAGTATTCAAAAATCAACATTATGAAAAGTCATGAAGTTTTTACAGGTATAGTCTAGACTATTAAATA
TAAAAATGTGTGATCTAATGATCCACATTAATTTTAGTCAAAAGATCCAATCATTCCAAAACGACACTCATAAAT
CTATAGATAAGAATCCCTAATTCTTTATTTCTTACGCCAAACAAGTTTAAGAAAGATTGATTTCCAAATTCTCTC
ATCTTTTTAGACTAATTTTTTTTTTTAATTATTCTAGTATAATTAAATTCCATCCGTATTATTACTAGAAATAAAA
TTCGGCCTGTTGGTGAAGATTACTATCATCAAATCACCTTTGATTTGCCGACAAAATTATTTTTGTGGTTCATTA
ATTTTGAAAGAAGCATTGTATTCTATTTTAGGAGACAAACATCCAATTGATTATATTAACACCTGTGCAGTGACC
AAAATCTGTCTTAAGGAGCTGTACATTTGCATACATTGTTTTATTTCATCTCCATCTAAATTTAATGTTCCAATA
AATCGCATTTTTGTGTACATTTGATGCACATATGTTGATTTCGATAATTTGAAACTTTATCCTATATTTTGTTTC
AATACACAATTTTATCATAGCTGTATATAAAAGATACAGTAGAAGATAATGGAGAATGGTATATTTCATTCATTC
TAATGGCAGTATTTATACAGGTATTCAGCATGTGTTAATCCTATAATTAAGCCCTATGATATAGGAGAAATTAAG
GGCATATTACGAGAAATATTACAGGAAAATATTTAGAGTAAACATCCTAAAATTTTGTCTGTCATCCTAGAATAT
TGCACATGATGATCATATTTTGTCATGCCCCCGCAAGATTATGCTTCTATCATGGATGCCAATCTTGGATCGATG
AAGAAGAAATGGTATGCGAGATAGAGCCTTGGTGAGTGAATCTGCCAGTTGGTCATTGGTGTGGACATGAGATAC
ACGTAGTCTGTTATTGAGAACTTGATCTCGAACAAAATGAAAGTCAATGTCGTCATGCTTCATGGAAGAATGGAA
GATCGGATTTTTGCTCAAGTAAGTAGCATCAATATTATCGCAGAAGATAGTTGGAATAGTGGATGAGTGAAACCC
C

>OP779755 - Allele s_*laperrinei* [Adjelella_9_S4; DSI-B: allele_s_191]
TTAAATTTATGCGATGAGTAGTCACGATTTTTCAAATGAATATGGTGCATATCAAAGAGTCAAATTGCTTTGAAT
TGAATGTGAGTATTCAAAAATCAACATTATGAAAAGTCATGAACTTTTTACAGGTATAGTCTAGACTATTAAATA
TAAAAATGTGTGATCTAATGATCCACATTAATTTTAGTCAAAAGATCCAATCATTCCAAAACGACACTCATAAAT
CTATAGATAAGAATCCCAAATTCTTTATTTCTTACGCCAAACAAGTTTAAGAAAGATTGATTTCCAAATTCTCTC
ATCTTTTTAGACTAATTTTTTTTTTTAATTATTCTAGTATAATTAAATTCCATCCGTATTATTACTAGAAATAAAA
TTCGGCCTGTTGGTGAAGATTACTATCATCAAATCAACTTTGATTTGCCGACAAAATTATTTTTGTGGTTCATTA
ATTTTGAAAGAAGCATTGTATTCTATTTTAGGAGACAAACATCCAATTGATTATATTAACACCTGTGCAGTGACC
AAAATCTGTCTTAAGGAGCTGTACATTTGCATACATTGTTTTATTTCATCTCCATCTAAATTTAATGTTCCAATA
AATCGCATTTTTGTGTACATTTGATGCACATATGTTGATTTCGATAATTTGAAACTTTATCCTATATTTTGTTTC
AATACACAATTTTATCATAGCTGTATATAAAAGATACAGTAGAAGATAATGGAGAATGGTATATTTCATTCATTC
TAATGGCAGTATTTATACAGGTATTCAGCATGTGTTAATCCTATAATTAAGCCCTATGATATAGGAGAAATTAAG

GGCATATTACGAGAAATATTACAGGAAAATATTTAGAGTAAACATCCTAAAATTTTGTCTGTCATCCTAGAATAT
TGCACATGATGATCATATTTTGTCATGCCCCCGCAAGATTATGCTTCTATCATGGATGCCAATCTTGGATCGATG
AAGAAGAAATGGTTTGCGAGATAGAGCCTTGGTGAGTGAATCTGCCAGTTGGTCATTGGTGTGGACATGAGATAC
ACGTAGTCTGTTATTGAGAACTTGATCTCGAACAAAATGAAAGTCAATGTCGTCATGCTTCATGGAAGAATGGAA
GACCGGATTTTTGCTCAAGTAAGTAGCATCAATATTATCACAGAAGATAGTTGGAATAGTGGATGAGTGAAACCC
C

(In separate pdf) Supplementary Dataset 2. Variation in coverage ratio between incompatibility groups along each anchored scaffold in four olive genome assemblies. 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 with RAD-seq are figured in a different color. Blue = *Olea e. europaea*, green = *O. e. laperrinei,* orange = *O. e. africana*. The four reference genomes (including the two haploid sequences of the Saharan olive) were scanned for variation in coverage ratio between the two incompatibility groups for each subspecies (Supplementary Dataset 2). 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, highlighted in grey on Figure 1; Mariotti *et al.,* 2020), 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 incompatibility groups regardless of the subspecies), did identify 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* (Carré *et al.*, 2021). We did not detect any windows shared across the three subspecies with the opposite profile (higher coverage in *ss* than in *Ss*). Similar results were observed using the 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 (Figure 1). This was supported by Wilcoxon tests, as we only detected nine 50-kb windows with significant differences in depth, 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) between 1.6 and 2.3 Mb (Figure 1). All 50-kb windows in this interval (15 windows) had significantly greater coverage in *Ss* individuals (see Figure 2 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 1) 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 incompatibility 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 S4A) or with the absence of this *S*-locus (Figure S4C). 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.

References

Besnard G, Cheptou PO, Debbaoui M, Lafont, P, Hugueny B, Dupin J, 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.

Besnard G, Rubio de Casas R, Christin PA, 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.

Carré A, Gallina S, Santoni S, Vernet P, Godé C, Castric V, Saumitou-Laprade P. 2021. Genetic mapping of sex and self-incompatibility determinants in the androdioecious plant *Phillyrea angustifolia*. *Peer Comm. J.* 1: e15.

Green PS. 2002. A revision of *Olea* L. (*Oleaceae*). *Kew Bull.* 57: 91–140.

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.

Mariotti R, Fornasiero A, Mousavi S, Cultrera NGM, 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 STS marker. *Front. Plant Sci.* 10: 1760.

Puyoou A, Gryta H, Fuchs AL, Blanchard P, Cheptou PO, Civeyrel L, Dufay M, Dupin J, Jargeat P, Lecompte E, *et al.* 2023. 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.*, doi: 10.1093/botlinnean/boad035.

Schuelke M. 2000. An economic method for the fluorescent labelling of PCR fragments. *Nat. Biotechnol.* 18: 233–234.