Genetic mapping of sex and self-incompatibility determinants in the androdioecious plant *Phillyrea angustifolia*

Authors:
Amélie Carré[1], Sophie Gallina[1], Sylvain Santoni[2], Philippe Vernet[1], Cécile Godé[1], Vincent Castric[1], Pierre Saumitou-Laprade[1#]

Affiliations :
[1] CNRS, Univ. Lille, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France
[2] UMR DIAPC - Diversité et adaptation des plantes cultivées

[#] **Corresponding author**: Pierre.Saumitou-Laprade@univ-lille.fr

Abstract

The diversity of mating and sexual systems in Angiosperms is spectacular, but the factors driving their evolution remain poorly understood. In plants of the Oleaceae family, an unusual self-incompatibility (SI) system has been discovered recently, whereby only two distinct homomorphic SI specificities segregate stably. To understand the role of this peculiar SI system in preventing or promoting the diversity of sexual phenotypes observed across the family, an essential first step is to characterize the genetic architecture of these two traits. Here, we developed a high-density genetic map of the androdioecious shrub *P. angustifolia* based on a F1 cross between a hermaphrodite and a male parent with distinct SI genotypes. Using a double restriction-site associated digestion (ddRAD) sequencing approach, we obtained reliable genotypes for 196 offspring and their two parents at 10,388 markers. The resulting map comprises 23 linkage groups totaling 1,855.13 cM on the sex-averaged map. We found strong signals of association for the sex and SI phenotypes, that were each associated with a unique set of markers on linkage group 12 and 18 respectively, demonstrating inheritance of these traits as single, independent, mendelian factors. The *P. angustifolia* linkage map shows robust synteny to the olive tree genome overall. Two of the six markers strictly associated with SI in *P. angustifolia* have strong similarity with a recently identified 741kb chromosomal region fully linked to the SI phenotype on chromosome 18 of the olive tree genome, providing strong cross-validation support. The SI locus stands out as being markedly more rearranged, while the sex locus has remained relatively more collinear. This *P. angustifolia* linkage map will be a useful resource to investigate the various ways by which the sex and SI determination systems have co-evolved in the broader phylogenetic context of the Oleaceae family.

Introduction

Sexual reproduction is strikingly diverse across Angiosperms, both in terms of the proportion of autogamous vs. allogamous matings and in terms of the distribution of male and female sexual functions within and among individuals (BARRETT 1998; SAKAI AND WELLER 1999; DIGGLE *et al.* 2011). The conditions under which this diversity could arise under apparently similar ecological conditions and have evolved rapidly -sometimes even within the same family- have been a topic of intense interest in evolutionary biology (BARRETT 1998). The control of self-fertilization and the delicate balance between its costs and benefits is considered to be a central force driving this diversity. Avoidance of self-fertilization is sometimes associated with observable phenotypic variations among reciprocally compatible partners. These variations can be morphological (e.g. distyly) or temporal (e.g. protandry, protogyny in the case of heterodichogamy), but in many cases the flowers show no obvious morphological or phenological variation, and self-fertilization avoidance relies on so-called "homomorphic" self-incompatibility (SI) systems. These systems are defined as the inability of fertile hermaphrodite plants to produce zygotes through self-fertilization (LUNDQVIST 1956; DE NETTANCOURT 1977), and typically rely on the segregation of a finite number of recognition "specificities" whereby matings between individuals expressing cognate specificities are not successful at producing zygotes. At the genetic level, the SI specificities most commonly segregate as a single multi-allelic mendelian locus, the S locus. This locus contains at least two genes, one encoding the male determinant expressed in pollen and the other encoding the female determinant expressed in pistils, with the male specificity sometimes determined by a series of tandemly arranged paralogs (KUBO *et al.* 2015). The male and female determinants are both highly polymorphic and tightly linked, being inherited as a single non-recombining genetic unit. In cases where the molecular mechanisms controlling SI could be studied in detail, they were found to be remarkably diverse, illustrating their independent evolutionary origins across the flowering plants (IWANO AND TAKAYAMA 2012). Beyond the diversity of the molecular functions employed, SI systems also can differ in their genetic architecture.

In the Poaceae family for example, two independent loci (named S and Z) control SI (Yang, et al., 2008). In other cases, the alternate allelic specificities can be determined by presence-absence variants rather than nucleotide sequence variants of a given gene, such as *e.g.* in *Primula vulgaris*, where one of the two reproductive phenotypes is hemizygous rather than heterozygous for the SI locus (LI *et al.* 2016).

In spite of this diversity of molecular mechanisms and genetic architectures, a common feature of SI phenotypes is that they are all expected to evolve under negative frequency-dependent selection, a form of natural selection favoring the long-term maintenance of high levels of allelic diversity (WRIGHT 1939). Accordingly, large numbers of distinct SI alleles are commonly observed to segregate within natural and cultivated SI species (reviewed in CASTRIC AND VEKEMANS 2004). There are notable exceptions to this general rule, however, and in some species only two SI specificities seem to segregate stably. Most often in such diallelic SI (DSI) systems, the two SI specificities are in perfect association with morphologically distinguishable floral phenotypes. In distylous species, for instance, two floral morphs called “pin” and “thrum” coexist (BARRETT 1992; BARRETT 2019). In each morph, the anthers and stigma are spatially separated within the flowers, but located at corresponding, reciprocal positions between the two morphs. Additional morphological differences exist, with S-morph flowers producing fewer but larger pollen grains than L-morph flowers (DULBERGER 1992). These morphological differences are believed to enhance the selfing avoidance conferred by the SI system but also to increase both male and female fitnesses (BARRETT 1990; BARRETT 2002; KELLER *et al.* 2014), although it is not clear which of SI or floral morphs became established in the first place (CHARLESWORTH AND CHARLESWORTH 1979).

The Oleacea family is another intriguing exception, where a DSI was recently found to be shared across the entire family (VERNET *et al.* 2016). In this family of trees, the basal genera *Jasminum* ($2n = 2x = 26$), *Fontanesia* ($2n = 2x = 26$) and *Forsythia* ($2n = 2x = 28$) are all heterostylous and therefore seem to possess a heteromorphic DSI system; in *Jasminum fruticans* self- and within-morph fertilization are unsuccessful (DOMMÉE *et al.* 1992). The ancestral heterostyly gave rise to species with hermaphrodite (e.g.

86 *Ligustrum vulgare*, *Olea europaea*), androdioecious (e.g. *P. angustifolia*, *Fraxinus ornus*), polygamous (e.g.
87 *F. excelsior*) and even dioecious (e.g. *F. chinensis*) sexual systems, possibly in association with a doubling
88 of the number of chromosomes ($2n = 2x = 46$ in the Oleeae tribe). Detailed crossing experiments in one of
89 these species, *P. angustifolia,* revealed the existence of a previously unsuspected homomorphic DSI. In
90 this androdioecious species (i.e. in which male and hermaphrodite individuals coexist in the same
91 populations), hermaphrodite individuals form two morphologically indistinguishable groups of SI
92 specificities that are reciprocally compatible but incompatible within groups, whereas males show
93 compatibility with hermaphrodites of both groups (SAUMITOU-LAPRADE *et al.* 2010). This "universal"
94 compatibility of males offsets the reproductive disadvantage they suffer from lack of their female function,
95 such that the existence of the DSI provides a powerful explanation to the long-standing evolutionary puzzle
96 represented by the maintenance of high frequencies of males in this species (PANNELL AND KORBECKA 2010;
97 SAUMITOU-LAPRADE *et al.* 2010; BILLIARD *et al.* 2015; PANNELL AND VOILLEMOT 2015). Extension of the stigma
98 test developed in *P. angustifolia* to more distant species including *L. vulgaris* (DE CAUWER *et al.* 2020), *F.*
99 *ornus* (VERNET *et al.* 2016) and *O. europaea* (SAUMITOU-LAPRADE *et al.* 2017; DUPIN *et al.* 2020) demonstrated
100 that all species exhibited some form of the DSI, but with no consistent association with floral morphology.
101 Cross-species pollination experiments even showed that pollen from *P. angustifolia* is able to trigger a
102 robust SI response on *O. europaea* and the more distant *F. ornus* and *F. excelsior* stigmas (the reciprocal is
103 also true). This opens the question of whether the homomorphic DSI is identical by descend (orthologous)
104 across the Oleeae tribe, even in the face of the variety of sexual polymorphisms present in the different
105 species. More broadly, its link with the heteromorphic DSI in the ancestral diploid, largely heterostylous,
106 species remains to be established (BARRETT 2019). Understanding the causes of the long-term maintenance
107 of this SI system and exploring its consequences on the evolution of sexual systems in hermaphrodite,
108 androdioecious, polygamous or dioecious species of the family represents an important goal. The case of
109 *P. angustifolia* is particularly interesting because it is one of the rare instances where separate sexes and

110 the DSI system can be studied in a single species, and therefore lies at the crossroads of processes that
111 may be obscured in lineages where they are unfolding independently (ROSAS AND DOMÍNGUEZ 2009; BARRETT
112 2010; LIU *et al.* 2012; BILLIARD *et al.* 2015)

113 A first step toward a better understanding of the role of the DSI in promoting the sexual diversity
114 in Oleaceae is to characterize and compare the genetic architecture of the SI and sexual phenotypes. At
115 this stage, however, the genomic resources for most of these non-model species remain limited. In this
116 context, the recent sequencing efforts (UNVER *et al.* 2017; JIMÉNEZ-RUIZ *et al.* 2020) and the genetic mapping
117 of the SI locus in a biparental population segregating for SI groups in *Olea europaea* (MARIOTTI *et al.* 2020)
118 represent major breakthroughs in the search for the SI locus in Oleaceae. They have narrowed down the
119 SI locus to an interval of 5.4cM corresponding to a region of approximately 300kb, but it is currently
120 unknown whether the same region is controlling SI in other species. In *P. angustifolia,* based on a series of
121 genetic analysis of progenies from controlled crosses, Billiard *et al.* (2015) proposed a fairly simple genetic
122 model, where sex and SI are controlled by two independently segregating diallelic loci. Under this model,
123 sex would be determined by the "M" locus at which a dominant allele *M* codes for the male phenotype
124 (*i.e. M is a* female-sterility mutation leading e.g. to arrested development of the stigma) and a recessive
125 allele *m* codes for the hermaphrodite phenotype. The S locus would encode the SI system and comprise a
126 dominant allele *S2* and a recessive allele *S1.* The model thus hypothesizes that hermaphrodites are
127 homozygous *mm* at the sex locus, and fall into two groups of SI specificities, named H_a and H_b carrying the
128 *S1S1* and *S1S2* genotypes at the S locus, respectively (their complete genotypes would thus be *mmS1S1*
129 and *mmS1S2* respectively). The model also hypothesizes three male genotypes (M_a: *mMS1S1*, M_b:
130 *mMS1S2*, and M_c*: mMS2S2*). In addition, Billiard *et al.* (2015) showed that, while males are compatible
131 with all hermaphrodites, offspring segregation varies according to which group of hermaphrodite they
132 sire: the progeny of H_a hermaphrodites pollinated by males systematically segregates both
133 hermaphrodites and males with a consistent but slight departure from 1:1 ratio, while that of H_b

134 hermaphrodites pollinated by the very same males systematically consists of male individuals only. This
135 observation suggests a pleiotropic effect of the *M* allele, conferring not only female sterility and universal
136 pollen compatibility, but also a complete male-biased sex-ratio distortion when crossed with one of the
137 two groups of hermaphrodites and a more subtle departure from 1:1 ratio when crossed with the other
138 group of hermaphrodites. The latter departure, however, was observed on small progeny arrays only, and
139 its magnitude thus comes with considerable uncertainty.

140 In this study, we developed a high-density genetic map for the non-model tree *P. angustifolia* using
141 a ddRAD sequencing approach and used it to address three main questions related to the evolution of its
142 peculiar reproductive system. First, are the SI and sex phenotypes in *P. angustifolia* encoded by just two
143 independent loci, as predicted by the most likely segregation model of Billiard *et al.* (2015)? Second, which
144 specific genomic regions are associated with the SI and sex loci, and what segregation model do the SI and
145 sex-associated loci follow (i.e. which of the males or hermaphrodites, and which of the two SI phenotypes
146 are homozygous *vs.* heterozygous at either loci, or are these phenotypes under the control of hemizygous
147 genomic regions?). Third, what is the level of synteny between our *P. angustifolia* genetic map and the
148 recently published Olive tree genome (UNVER *et al.* 2017; MARIOTTI *et al.* 2020), both globally and
149 specifically at the SI and sex-associated loci?

150

151 ## Material and Methods

152 ### Experimental cross and cartography population

153 In order to get both the SI group and the sexual phenotype (males vs hermaphrodites) to segregate
154 in a single progeny array, a single maternal and a single paternal plant were chosen among the progenies
155 of the controlled crosses produced by (BILLIARD *et al.* 2015). Briefly, a H_a maternal tree (named 01.N-25,
156 with putative genotype mmS1S1) was chosen in the progeny of a (H_a x M_a) cross. It was crossed in March
157 2012 to a M_b father (named 13.A-06, putative genotype mM S1S2) chosen in the progeny of a (H_a x M_c)
158 cross, following the protocol of Saumitou-Laprade *et al.* (2010). Both trees were maintained at the

159 experimental garden of the "Plateforme des Terrains d'Expérience du LabEx CeMEB," (CEFE, CNRS) in
160 Montpellier, France. F1 seeds were collected in September 2012 and germinated in the greenhouse of the
161 "Plateforme Serre, cultures et terrains expérimentaux," at the University of Lille (France). Seedling
162 paternity was verified with two highly polymorphic microsatellite markers (VASSILIADIS *et al.* 2002), and
163 1,064 plants with confirmed paternity were installed in May 2013 on the experimental garden of the
164 "Plateforme des Terrains d'Expérience du LabEx CeMEB," (CEFE, CNRS) in Montpellier. Sex was visually
165 determined based on the absence of stigma for 1,021 F1 individuals during their first flowering season in
166 2016 and 2017. 21 progenies did not flower and 22 died during the test period. The hermaphrodite
167 individuals were assigned to an SI group using the stigma test previously described in Saumitou-Laprade
168 *et al.* (2010; SAUMITOU-LAPRADE *et al.* 2017).
169

170 **DNA extraction, library preparation and sequencing**

171 In 2015, *i.e.* the year before sex phenotyping and stigma test were performed, 204 offspring were
172 random selected for genomic library preparation and genotyping. Briefly, DNA from parents and progenies
173 were extracted from 100 mg of frozen young leaves with the Chemagic DNA Plant Kit (Perkin Elmer
174 Chemagen, Baesweller, DE, Part # CMG-194), according to the manufacturer's instructions. The protocol
175 was adapted to the use of the KingFisher Flex™ (Thermo Fisher Scientific, Waltham, MA, USA) automated
176 DNA purification workstation. The extracted DNA was quantified using a Qubit fluorometer (Thermo Fisher
177 Scientific, Illkirch, France). Genome complexity was reduced by Double Digestion Restriction Associated
178 DNA sequencing (ddRAD seq) (PETERSON *et al.* 2012) using two restriction enzymes: *PstI*, a rare-cutting
179 restriction enzyme sensitive to methylation, and *MseI*, a common-cutting restriction enzyme. The libraries
180 were constructed at the INRAE - AGAP facilities (Montpellier, France). Next-generation sequencing was
181 performed in a 150-bp paired-ends-read mode using three lanes on a HiSeq3000 sequencer (Illumina, San
182 Diego, CA, USA) at the Get-Plage core facility (Genotoul platform, INRAE Toulouse, France).

183

184 **GBS data analysis and linkage mapping**

185 Illumina sequences were quality filtered with the *process_radtags* program of Stacks v2.3 (CATCHEN
186 *et al.* 2011) and then aligned to a *de novo* catalog created from parents and progeny of the cross. According
187 to the Rochette & Catchen protocol (ROCHETTE AND CATCHEN 2017) some parameters are tested to assembly
188 and alignment steps. The following Stacks parameters were used: a minimum stack depth of 3 (–m=3), at
189 most two nucleotide differences when merging stacks into loci (–M=2), and at most two nucleotide
190 differences between loci when building the catalog (–n=2). Both parental and all offspring FASTA files were
191 aligned to the *de novo* catalog using Bowtie2 v2.2.6 (LANGMEAD AND SALZBERG 2012), the option 'end-to-
192 end' and 'sensitive' were used for the alignment. At this step, one .bam file was obtained per individual to
193 construct the linkage map with Lep-MAP3 (RASTAS 2017). A custom python script was used, after remove
194 SPNs markers with reads cover <5, the script combines samtools v1.3.1 and the script *pileupParser2.awk*
195 (limit1=5) to convert .bam files to the good Lep-Map3 format. We used the *ParentCall2* module of Lep-
196 MAP3 to select loci with reliable parental genotypes by considering genotype information on parents and
197 offspring. The *Filtering2* module was then used to remove non-informative and distorted markers
198 (dataTolerance = 0.0000001). The module *SeparateChromosomes2* assigned markers to linkage groups
199 (LGs), after test, where the logarithm of odds score (LodLimit) varied from 10 to 50 in steps of 5 then from
200 20 to 30 in steps of 1 and the minimum number of SNP markers (sizeLimit) per linkage group from 50 to
201 500 in steps of 50 for each of the LodLimit. The two parameters, lodLimit = 27 and sizeLimit = 250, were
202 chosen as the best parameters to obtain the 23 linkage (as expected in members of the Oleoideae
203 subfamily; (WALLANDER AND ALBERT 2000). A custom python script removed loci with SNPs markers mapped
204 on two or more different linkage groups. The last module *OrderMarkers2* ordered the markers within each
205 LG. To consider the slight stochastic variation in marker distances between executions, the module was
206 run three times on each linkage group, first separately for the meiosis that took place in each parent
207 (sexAveraged = 0) and then averaged between the two parents (sexAveraged = 1). To produce the most

208 likely final father and mother specific maps and a final sex-averaged maps (De-Kayne and Feulner 2018),
209 we kept for each map the order of markers that had the highest likelihoods for each linkage group.
210
211 **Sex and SI locus identification**
212 To identify the sex-determination system in *P. angustifolia* we considered two possible genetic
213 models. First, a "XY" male heterogametic system, where males are heterozygous or hemizygous (XY) and
214 hermaphodites are homozygous (XX). Second, a "ZW" hermaphrodite heterogametic system, where
215 hermaphodites are heterozygous or hemizygous (ZW) and males are homozygous (ZZ). We applied the
216 same logic to the SI determination system, as segregation patterns (Billiard et al. 2015) suggested that SI
217 possibly also has a heterogametic determination system, with homozygous H_a and heterozygous H_b. In the
218 same way as for sex, it is therefore possible to test the different models (XY, ZW or hemigous) to determine
219 which SNPs are linked to the SI phenotypes.
220 Based on this approach, we identified sex-linked and SI-linked markers on the genetic map by
221 employing SEX-DETector, a probabilistic model based on maximum-likelihood inference (Muyle *et al.*
222 2016). Briefly, a new alignment of reads from each individual on the loci used to construct the linkage map
223 was done with bwa (Li and Durbin 2009). This new alignment has the advantage of retrieving more SNPs
224 than used by LepMap3, as SNPs considered as non-informative by LepMap3 can still be informative to
225 distinguish among sex- or SI-determination systems by SEX-DETector. The alignment was analyzed using
226 Reads2snp (default tool for SEX-DETector) (Tsagkogeorga *et al.* 2012). For each phenotype (H_a vs. H_b and
227 males vs. hermaphrodites), SEX-DETector was run for both a XY and a ZW model with the following
228 parameters: -detail, -L, -SEM, -thr 0.8, -E 0.05. For each run, SEX-DETector also calculates the probability
229 for X (or Z)-hemizygous segregation in the heterozygous haplotypes. To compensate for the heterogeneity
230 between the number of males (83) and hermaphrodites (113) in our progeny array, each model was tested
231 three times with sub-samples of 83 hermaphrodites obtained by randomly drawing from the 113

232 individuals. We retained SNPs with a ≥80% probability of following an XY (or ZW) segregation pattern, with
233 a minimum of 50% of genotyped individuals and less that 5% of the individuals departing from this model
234 (due to either genotyping error or crossing-over).

235

236 ### Synteny analysis with the olive tree

237 To study synteny, we used Basic Local Alignment Search Tool (BLAST) to find regions of local
238 similarity between the *P. angustifolia* ddRADseq loci in the linkage map and the *Olea europea var. sylvestris*
239 genome assembly (UNVER *et al.* 2017). This assembly is composed of 23 main chromosomes and a series
240 of 41,233 unanchored scaffolds for a total of 1,142,316,613 bp. Only loci with a unique hit with at least
241 85% identity over a minimum of 110 bp were selected for the synteny analysis. Synteny relationships were
242 visualized with *circos-0.69-6* (KRZYWINSKI *et al.* 2009). Synteny between linkage groups of *P. angustifolia*
243 and the main 23 *O. europea* chromosomes was established based on the number of markers with a
244 significant BLAST hit. At a finer scale, we also examined synteny with the smaller unanchored scaffolds of
245 the assembly, as the history of rearrangement and allo-tetraploidization is likely to have disrupted synteny.

246

247 ## Results

248 ### Phenotyping progenies for sex and SI groups

249 As expected, our cartography population segregated for sex and SI phenotypes, providing a
250 powerful resource to genetically map these two traits. Among the 1,021 F1 individuals that flowered
251 during the two seasons of phenotyping, we scored 619 hermaphrodites and 402 males, revealing a biased
252 sex ratio in favor of hermaphrodites (khi^2= 1.28×10^{-11}). Stigma tests were successfully performed on 613
253 hermaphrodites (6 individuals flowered too late to be included in a stigma test), revealing 316 H_a and 297
254 H_b, i.e. an equilibrated segregation of the two SI phenotypes (khi^2= 0.27). The random subsample of 204
255 F1 progenies chosen before the first flowering season for ddRAD-seq analysis (see below) followed similar

256 phenotypic proportions. 196 of the 204 progenies ended up flowering, revealing 83 males and 113
257 hermaphrodites, among which 60 belonged to the H_a group and 53 to the H_b group.
258
259 **Linkage mapping**
260 The two parents and the 196 offspring that had flowered were successfully genotyped using a
261 ddRAD-seq approach. Our stringent filtering procedure identified 11,070 loci composed of 17,096 SNP
262 markers as being informative for Lep-MAP3. With a LOD score of 27, 10,388 loci composed of 15,814 SNPs
263 were assigned to, and arranged within, 23 linkage groups in both sex-averaged and sex-specific maps
264 (Table 1).
265 The linkage groups of the mother map were on average larger (84.45 cM) than the linkage groups
266 of the father map (77.84 cM) and varied from 54.14 cM to 133.40 cM and from 56.06 cM to 131.66 cM
267 respectively (Table 1). The total map lengths were 1,855.13 cM, 1,790.33 cM and 1,942.30 cM in the sex-
268 averaged, male and female maps, respectively. The length of the linkage groups varied from 59.10 cM to
269 129.70 cM in the sex-averaged map, with an average of 687 SNPs markers per linkage group (Table 1).
270
271 **Sex and SI loci identification**
272 When comparing H_a and H_b, among the 38,998 SNPs analyzed by SEX-DETector, 496 have a
273 probability of following an XY pattern ≥0.80. We then applied two stringent filters by retaining only SNPs
274 that had been genotyped for more than 50% of the offspring (n=211), and for which less than 5% of the
275 offspring departed from the expected genotype under a XY model (n=23). Six of these 23 SNPs, distributed
276 in 4 loci, followed a segregation pattern strictly consistent with a XY model. These four loci are tightly
277 clustered on the linkage map and define a region of 1.230 cM on linkage group 18 (LG18, Figure 1). Relaxing
278 the stringency or our thresholds, this region also contains five loci that strictly follow an XY segregation
279 but with less than 50% of offsprings successfully genotyped, as well as six loci with autosomal inheritance.
280 Using the same filtering scheme, none of the SNPs was found to follow a ZW pattern. This provides

evidence that this region on LG18 is associated with the SI phenotypes, with a determination system akin to a XY system where H_b have the heterogametic genotype.

For the comparison of male and hermaphrodites, an average of 44,565 SNPs were analyzed by SEX-DETector across the three subsamples, among which an average of 438 have a probability of following an XY pattern ≥0.80. We applied the same set of stringent filters and retained an average of 171 SNPs having been genotyped for at least 50% of the offspring, among which 41 have less than 5% of the offspring departing from the expected genotype under a XY model and are shared across the three subsets. Thirty two of these SNPs followed a segregation pattern strictly consistent with a XY model. These 32 markers, corresponding to 8 loci, are distributed along a region of 2.216 cM on linkage group 12 (LG12, Figure 1). Relaxing the stringency or our thresholds, this region also contains five loci that strictly follow an XY segregation pattern but with less than 50% of offspring successfully genotyped, as well as 17 loci consistent with autosomal inheritance. Again, no SNP was found to follow a ZW pattern. This provides evidence that this independent region on LG12 is associated with sex, with a determination system akin to a XY system where males have the heterogametic genotype.

Synteny analysis with the olive tree

About half (49%) of the 10,388 *P. angustifolia* loci used for the genetic map had a significant BLAST hit on the olive tree genome. Overall, the relative position of these hits was highly concordant with the structure of the linkage map. Indeed, the vast majority (79.7%) of loci belonging to a given linkage group found hits on the same olive tree chromosome. Loci that did not follow this general pattern did not find hits in particularly clustered regions on other chromosomes, suggesting either small rearrangements or mapping/assembly errors at the scale of individual loci. The order of loci within the linkage groups was also well conserved with only limited evidence for rearrangements (Figure 2, Figure 3), suggesting that the two genomes have remained largely collinear overall.

306 We then specifically inspected synteny between the linkage groups carrying either the sex or the
307 SI locus and the olive tree genome. Synteny was good for LG12, the linkage group containing the markers
308 associated with the sex phenotype. Among the 645 loci of LG12, 365 find homology in the olive tree. Eighty
309 eight percent had their best hits on the same chromosome of the olive tree (chromosome 12 per our
310 numbering of the linkage groups), and the order of markers was largely conserved along this chromosome.
311 Six loci contained in the region associated with sex on LG12 had hits on a single small 1,940,009bp region
312 on chromosome 12. This chromosomal interval contains 81 annotated genes in the olive tree genome. In
313 addition, eight loci in the sex region had their best hits on a series of five smaller scaffolds (Sca393,
314 Sca1196, Sca1264, Sca32932, Sca969) that could not be reliably anchored in the main olive tree assembly
315 but may nevertheless also contain candidates for sex determination. Collectively, these scaffolds represent
316 1.849.345bp of sequence in the olive tree genome and contain 47 annotated genes.
317 Synteny was markedly poorer for markers on LG18, the linkage group containing the markers
318 associated with the SI specificity phenotypes. Of the 440 loci on LG18, 203 have BLAST hits on the olive
319 tree genome. Although a large proportion (89%) had their best hits on chromosome 18, the order of hits
320 along that chromosome suggested a large number of rearrangements. This more rearranged order was
321 also observed for the six markers that were strictly associated with SI in *P. angustifolia*. Two of them had
322 hits on a single region of 741,403bp on the olive tree genome. This region contains 32 annotated genes
323 and contains two markers that were previously found to be genetically associated with SI directly in the
324 olive tree by Mariotti *et al.* (2020). Three markers more loosely associated to SI in *P. angustifolia* found
325 hits on a more distant region on chromosome 18 (19,284,909-19,758,630Mb). The three other strongly
326 associated markers all had hits on scaffold 269, which contains 11 annotated genes and represents
327 545.128bp. Nine other loci strongly or loosely associated with SI had hits on a series of seven other
328 unanchored scaffolds (Sca1199, Sca1200, Sca1287, Sca1579, Sca213, Sca327, Sca502) that collectively
329 represent 78 annotated genes and 2.539.637bp.

Discussion

Until now, studies have mostly relied on theoretical or limited genetic segregation analyses to investigate the evolution of sexual and SI phenotypes in *P. angustifolia* (VASSILIADIS *et al.* 2002; SAUMITOU-LAPRADE *et al.* 2010; HUSSE *et al.* 2013; BILLIARD *et al.* 2015). In this study, we created the first genetic map of the androdioecious species *P. angustifolia* and identified the genomic regions associated with these two important reproductive phenotypes. The linkage map we obtained shows strong overall synteny with the olive tree genome, and reveals that sex and SI phenotypes segregate independently from one another, and are each strongly associated with a different genomic region (in LG18 and LG12, respectively).

The SI linked markers on LG18 are syntenic with the genomic interval recently identified by Mariotti *et al.* (2020) in the olive tree, providing strong reciprocal support that the determinants of SI are indeed located in this region. Interestingly, we observed a series of shorter scaffolds that could not previously be anchored in the main assembly of the olive tree genome but are syntenic to genetic markers that are strictly linked to SI in *P. angustifolia.* These unanchored scaffolds provide a more complete set of genomic sequences that will be important to consider in the perspective of identifying the (currently elusive) molecular determinants of SI in these two species. We note that poor assembly of the S-locus region was indeed expected given the considerable levels of structural rearrangements typically observed in SI- and more generally in the mating type-determining regions (GOUBET *et al.* 2012; BADOUIN *et al.* 2015), making *P. angustifolia* a useful resource to map the SI locus in the economically important species *O. europeae*.

Our observations also provide direct support to the hypothesis that the determinants of SI have remained at the same genomic position at least since the two lineages diverged, 30 to 40 Myrs ago (BESNARD *et al.* 2009; OLOFSSON *et al.* 2019). Stability of the genomic location of SI genes has been observed in some Brassicaceae species, where the *SRK-SCR* system maps at orthologous positions in the Arabidopsis

354 and Capsella genuses (GUO *et al.* 2011). In more distant Brassicaceae species, however, the SI system is
355 found at different genomic locations, such as in Brassica and Leavenworthia. In the former, the molecular
356 determinants have remained orthologous (also a series of *SRK-SCR* pairs, (IWANO *et al.* 2014), but not in
357 the latter, where SI seems to have evolved d*e novo* from exaptation of a pair of paralogous genes (CHANTHA
358 *et al.* 2013; CHANTHA *et al.* 2017). Together with the fact that *P. angustifolia* pollen is able to trigger a robust
359 SI response on *O. europaeae* stigmas (SAUMITOU-LAPRADE *et al.* 2017), the synteny we observed provides
360 strong support to the hypothesis that the *P. angustifolia* and *O. europaeae* SI systems are orthologous.

361 Several approaches could now be used to refine the mapping location in *P. angustifolia,* and
362 ultimately zero in on the molecular determinants of SI. A first possibility would require fine-mapping using
363 larger offspring arrays, starting from our cross for which only a fraction of all phenotyped individuals were
364 genotyped. Beyond the analysis of this controlled cross, evaluating whether the association of the SI
365 phenotype still holds for markers within a larger set of accessions from diverse natural populations will
366 constitute a powerful fine-mapping approach. Since the SI phenotypes seem to be functionally
367 orthologous across the Oleeae tribe (VERNET *et al.* 2016), the approach could, in principle, be extended to
368 more distant SI species of the family like *L. vulgare* or *F. ornus*. Identification of sequences whose
369 association holds over these considerable time scales would represent excellent corroborative evidence
370 to validate putative SI candidates. In parallel, an RNA-sequencing approach could be used to identify
371 transcripts specific to the alternate SI phenotypes.

372 While comparison to the closely related *O. europeae* genome is a useful approach for the mapping
373 of SI in *P. angustifolia*, it is *a priori* of limited use for mapping the sex-determining region, since the olive
374 tree lineage has been entirely hermaphroditic for at least 32.22 Myrs (confidence interval: 28-36 Myrs)
375 (FigS1 in OLOFSSON *et al.* 2019). Detailed exploration of the genomic region in the olive tree that is
376 orthologous to the markers associated with sex in *P. angustifolia* is however interesting, as it may either
377 have anciently played a role in sex determination and subsequently lost it, or alternatively it may contain

378 quiescent sex-determining genes that happen to have been activated along the *P. angustifolia* specifically.
379 At a broader scale, mapping and eventually characterizing the sex locus in other androdioecious species
380 such as e.g. *F. ornus* could indicate whether the different instances of androdioecy in the family represent
381 either orthologous phenotypes or independent evolutionary emergences.

382 Identifying the molecular mechanisms of the genes controlling SI and sex and tracing their
383 evolution in a phylogenetic context would prove extremely useful. First, it could help understand the
384 strong functional pleiotropy between sex and SI phenotypes, whereby males express universal SI
385 compatibility (SAUMITOU-LAPRADE *et al.* 2010). In other words, males are able to transmit the SI specificities
386 they inherited from their parents, but they do not express them themselves even though their pollen is
387 fully functional. This intriguing feature of the SI system was key to solve the puzzle of why *P. angustifolia*
388 maintains unusually high frequencies of males in natural populations (HUSSE *et al.* 2013), but the question
389 of how being a male prevents expression of the SI phenotype is an open question. A possibility is that the
390 *M* allele of the sex locus contains a gene interacting negatively either with the pollen SI determinant itself
391 or with a gene of the downstream response cascade. Identifying the molecular basis of this epistasis will
392 be an interesting next step. Second, another intriguing feature of the system is segregation distortion, that
393 is observed at several levels. Billiard *et al.* (2015) observed complete segregation bias in favor of male
394 offspring of H_b hermaphrodites sired by males. Here, by phenotyping >1,000 offspring of a H_a
395 hermaphrodite sired by a M_b male, we confirmed that this cross also entails a departure from Mendelian
396 segregation, this time in favor of hermaphrodites, albeit of a lesser magnitude. Although the generality of
397 this observation still remains to be determined by careful examination of the other possible crosses (H_a
398 hermaphrodites x M_a and M_c males), it is clear that segregation distortion is a general feature of this
399 system, as was already observed in other sex determination systems causing departures from equal sex
400 ratios (e.g. KOZIELSKA *et al.* 2010). Beyond the identification of the mechanisms by which the distortions

401 arise, pinpointing the evolutionary conditions leading to their emergence will be key to understanding the
402 role they may have played in the evolution of this reproductive system.

403 More broadly, while sex and mating types are fully aligned in many species across the tree of life
404 and cannot be distinguished, the question of when and how sex and mating types evolve separately raises
405 several questions. The evolution of anisogamy (and hence, sexual differentiation) has been linked to that
406 of mating types (CHARLESWORTH 1978). In volvocine algae for instance, the mating-type locus in isogamous
407 species is orthologous to the pair of U/V sex chromosomes in anisogamous/oogamous species, suggesting
408 that the sex-determination system derives from the mating-type determination system (GENG *et al.* 2014).
409 From this perspective the Oleaceae family is an interesting model system, where a SI system is ancestral,
410 and in which some species have evolved sexual specialization that is aligned with the two SI phenotypes
411 (e.g. in the polygamous *F. excelsior* males belong to the H_a SI group and can only mate with hermaphrodites
412 or females of the H_b group, and the sexual system of *F. excelsior* can be viewed as subdioecy (SAUMITOU-
413 LAPRADE *et al.* 2018). In other species, sexual specialization is disjoint from SI and led to the differentiation
414 of males and hermaphrodites (e.g. in the androdiecious *P. angustifolia* and probably *F. ornus*, the male
415 determinant is genetically independent from the SI locus but fully linked to a genetic determinant causing
416 the epistatic effect over SI (BILLIARD *et al.* 2015; VERNET *et al.* 2016). Yet other species have remained
417 perfectly hermaphrodites and have no trace of sexual differentiation whatsoever (*O. europeae*).
418 Understanding why some species have followed one evolutionary trajectory while others have followed
419 another will be an exciting avenue for future research.

420

421 ## Author Contributions

422 All authors contributed to the study presented in this paper. PS-L and PV developed, designed and oversaw
423 the study; they coordinated the cross and carried out the phenotyping and stigma tests. CG performed the
424 seedling paternity analysis. SS performed DNA extraction, library preparation and organized sequencing.

AC and SG constructed the data analysis pipeline and AC, SS, PS-L and VC interpreted the results and wrote the manuscript.

Acknowledgements

We thank Sylvain Bertrand and Fantin Carpentier for technical help for the phenotyping and Jos Käfer for scientific discussion and help in applying Sex-DETector to our material. We thank Jacques Lepart†, Mathilde Dufay, Pierre Olivier Cheptou, Xavier Vekemans, Sylvain Billiard and Bénédicte Felter for scientific discussions. Field and laboratory work for phenotyping were done at the Platform Terrains d'Expériences (Labex CeMEB ANR-10-LABX-0004-CeMEB) of the Centre d'Ecologie Fonctionnelle et Evolutive (CEFE, CNRS) with the help of Thierry Mathieu and David Degueldre and at the platform Serres cultures et terrains expérimentaux of the Lille University with the help of Nathalie Faure and Angélique Bourceaux. We are grateful to Marie-Pierre Dubois for providing access to the microscopy facilities at the SMGE (Service des Marqueurs Génétiques en Ecologie) platform (CEFE). This work was funded by the French National Research Agency through the project 'TRANS' (ANR-11-BSV7- 013-03) and by a grant from the European Research Council (NOVEL project, grant #648321). A.C was supported by a doctoral grant from the French ministry of research. The authors also thank the Région Hauts-de-France, and the Ministère de l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for Regional Economic Development for their financial support. We also thank the HPC Computing Mésocentre of the University of Lille which provided us with the computing grid.

References

Badouin, H., M. E. Hood, J. Gouzy, G. Aguileta, S. Siguenza *et al.*, 2015 Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae. Genetics 200**:** 1275-1284.

Barrett, S. C., 2002 The evolution of plant sexual diversity. Nature Reviews Genetics 3**:** 274-284.

Barrett, S. C. H., 1990 The evolution and adaptive significance of heterostyly. Trends in Ecology & Evolution 5**:** 144-148.

Barrett, S. C. H., 1992 Heterostylous Genetic Polymorphisms: Model Systems for Evolutionary Analysis, pp. 1-29 in *Evolution and Function of Heterostyly*, edited by S. C. H. Barrett. Springer Berlin Heidelberg, Berlin, Heidelberg.

Barrett, S. C. H., 1998 The evolution of mating strategies in flowering plants. Trends in Plant Science 3**:** 335-341.

Barrett, S. C. H., 2010 Understanding plant reproductive diversity. Philosophical Transactions of the Royal Society B: Biological Sciences 365**:** 99-109.

Barrett, S. C. H., 2019 'A most complex marriage arrangement': recent advances on heterostyly and unresolved questions. New Phytologist 224**:** 1051-1067.

Besnard, G., R. Rubio de Casas, P.-A. Christin and P. Vargas, 2009 Phylogenetics of Olea (Oleaceae) based on plastid and nuclear ribosomal DNA sequences: Tertiary climatic shifts and lineage differentiation times. Annals of Botany 104**:** 143-160.

Billiard, S., L. Husse, P. Lepercq, C. Gode, A. Bourceaux *et al.*, 2015 Selfish male-determining element favors the transition from hermaphroditism to androdioecy. Evolution 69**:** 683-693.

Castric, V., and X. Vekemans, 2004 Plant self-incompatibility in natural populations: a critical assesment of recent theoretical and empirical advances. Molecular Ecology 13**:** 2873-2889.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait, 2011 Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda, Md.) 1**:** 171-182.

Chantha, S.-C., A. C. Herman, V. Castric, X. Vekemans, W. Marande *et al.*, 2017 The unusual S locus of Leavenworthia is composed of two sets of paralogous loci. New Phytologist 216**:** 1247-1255.

Chantha, S.-C., A. C. Herman, A. E. Platts, X. Vekemans and D. J. Schoen, 2013 Secondary Evolution of a Self-Incompatibility Locus in the Brassicaceae Genus Leavenworthia. PLOS Biology 11**:** e1001560.

Charlesworth, B., 1978 The population genetics of anisogamy. Journal of Theoretical Biology 73**:** 347-357.

Charlesworth, D., and B. Charlesworth, 1979 A Model for the Evolution of Distyly. The American Naturalist 114**:** 467-498.

De-Kayne, R., and P. G. D. Feulner, 2018 A European Whitefish Linkage Map and Its Implications for Understanding Genome-Wide Synteny Between Salmonids Following Whole Genome Duplication. G3 (Bethesda) 8**:** 3745-3755.

De Cauwer, I., P. Vernet, S. Billiard, C. Godé, A. Bourceaux *et al.*, 2020 Widespread coexistence of self-compatible and self-incompatible phenotypes in a diallelic self-incompatibility system in *Ligustrum vulgare* (Oleaceae). bioRxiv**:** 2020.2003.2026.009399.

De Nettancourt, D., 1977 *Incompatibility in Angiosperms*. Springer-Verlag, Berlin, Heidelberg et New York.

Diggle, P. K., V. S. Di Stilio, A. R. Gschwend, E. M. Golenberg, R. C. Moore *et al.*, 2011 Multiple developmental processes underlie sex differentiation in angiosperms. Trends in Genetics 27**:** 368-376.

Dommée, B., J. D. Thompson and F. Cristini, 1992 Distylie chez Jasminum fruticans L.: hypothèse de la pollinisation optimale basée sur les variations de l'écologie intraflorale. Bulletin de la Société Botanique de France. Lettres Botaniques 139**:** 223-234.

Dulberger, R., 1992 Floral Polymorphisms and Their Functional Significance in the Heterostylous Syndrome, pp. 41-84 in *Evolution and Function of Heterostyly*, edited by S. C. H. Barrett. Springer Berlin Heidelberg, Berlin, Heidelberg.

Dupin, J., P. Raimondeau, C. Hong-Wa, S. Manzi, M. Gaudeul *et al.*, 2020 Resolving the Phylogeny of the Olive Family (Oleaceae): Confronting Information from Organellar and Nuclear Genomes. Genes 11**:** 1508.

Geng, S., P. De Hoff and J. G. Umen, 2014 Evolution of sexes from an ancestral mating-type specification pathway. PLoS biology 12**:** e1001904-e1001904.

Goubet, P. M., H. Bergès, A. Bellec, E. Prat, N. Helmstetter *et al.*, 2012 Contrasted Patterns of Molecular Evolution in Dominant and Recessive Self-Incompatibility Haplotypes in Arabidopsis. PLOS Genetics 8**:** e1002495.

Guo, Y.-L., X. Zhao, C. Lanz and D. Weigel, 2011 Evolution of the S-locus region in Arabidopsis relatives. Plant physiology 157**:** 937-946.

Husse, L., S. Billiard, J. Lepart, P. Vernet and P. Saumitou-Laprade, 2013 A one-locus model of androdioecy with two homomorphic self-incompatibility groups: expected vs. observed male frequencies. Journal of evolutionary biology 26**:** 1269-1280.

Iwano, M., K. Ito, H. Shimosato-Asano, K.-S. Lai and S. Takayama, 2014 Self-Incompatibility in the Brassicaceae, pp. 245-254 in *Sexual Reproduction in Animals and Plants*, edited by H. Sawada, N. Inoue and M. Iwano. Springer Japan, Tokyo.

Iwano, M., and S. Takayama, 2012 Self/non-self discriminition in angiosperm self-incompatibility. Current Opinion in Plant Biology 15**:** 78-83.

Jiménez-Ruiz, J., J. A. Ramírez-Tejero, N. Fernández-Pozo, M. d. l. O. Leyva-Pérez, H. Yan *et al.*, 2020 Transposon activation is a major driver in the genome evolution of cultivated olive trees (Olea europaea L.). The Plant Genome 13**:** e20010.

Keller, B., J. D. Thomson and E. Conti, 2014 Heterostyly promotes disassortative pollination and reduces sexual interference in Darwin's primroses: evidence from experimental studies. Functional Ecology 28**:** 1413-1425.

Kozielska, M., F. J. Weissing, L. W. Beukeboom and I. Pen, 2010 Segregation distortion and the evolution of sex-determining mechanisms. Heredity 104**:** 100-112.

Krzywinski, M. I., J. E. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: An information aesthetic for comparative genomics. Genome Research.

Kubo, K.-i., T. Paape, M. Hatakeyama, T. Entani, A. Takara *et al.*, 2015 Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. Nature Plants 1**:** 1-9.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature methods 9**:** 357-359.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25**:** 1754-1760.

Li, J., J. M. Cocker, J. Wright, M. A. Webster, M. McMullan *et al.*, 2016 Genetic architecture and evolution of the S locus supergene in Primula vulgaris. Nature plants 2**:** 16188.

Liu, Y., Z. Luo, X. Wu, X. Bai and D. Zhang, 2012 Functional dioecy in Morinda parvifolia (Rubiaceae), a species with stigma-height dimorphism. Plant Systematics and Evolution 298**:** 775-785.

Lundqvist, A., 1956 Self-incompatibility in rye. Hereditas 42**:** 293-348.

Mariotti, R., S. Pandolfi, I. De Cauwer, P. Saumitou-Laprade, P. Vernet *et al.*, 2020 Diallelic self-incompatibility is the main determinant of fertilization patterns in olive orchards. Evolutionary Applications n/a.

Muyle, A., J. Kafer, N. Zemp, S. Mousset, F. Picard *et al.*, 2016 SEX-DETector: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. Genome Biol Evol 8**:** 2530-2543.

Olofsson, J. K., I. Cantera, C. Van de Paer, C. Hong-Wa, L. Zedane *et al.*, 2019 Phylogenomics using low-depth whole genome sequencing: A case study with the olive tribe. Molecular Ecology Resources 19**:** 877-892.

Pannell, J. R., and G. Korbecka, 2010 Mating-System Evolution: Rise of the Irresistible Males. Current Biology 20**:** R482-R484.

Pannell, J. R., and M. Voillemot, 2015 Plant mating systems: female sterility in the driver's seat. Current Biology 25**:** R511-514.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher and H. E. Hoekstra, 2012 Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7**:** e37135.

Rastas, P., 2017 Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33**:** 3726-3732.

Rochette, N. C., and J. M. Catchen, 2017 Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols 12**:** 2640-2659.

Rosas, F., and C. A. DomÍNguez, 2009 Male sterility, fitness gain curves and the evolution of gender specialization from distyly in Erythroxylum havanense. Journal of Evolutionary Biology 22**:** 50-59.

Sakai, A. K., and S. G. Weller, 1999 Gender and Sexual Dimorphism in Flowering Plants: A review of Terminology, Biogeographic Patterns, Ecological Correlates, and Phylogenetic Approaches, pp. 1-31 in *Gender and Sexual Dimorphism in Flowering Plants*, edited by M. A. Geber, T. E. Dawson and L. F. Delph. Springer Berlin Heidelberg, Berlin, Heidelberg.

Saumitou-Laprade, P., P. Vernet, A. Dowkiw, S. Bertrand, S. Billiard *et al.*, 2018 Polygamy or subdioecy? The impact of diallelic self-incompatibility on the sexual system in Fraxinus excelsior (Oleaceae). Proceedings. Biological sciences 285**:** 20180004.

Saumitou-Laprade, P., P. Vernet, C. Vassiliadis, Y. Hoareau, G. de Magny *et al.*, 2010 A self-incompatibility system explains high male frequencies in an androdioecious plant. Science 327**:** 1648-1650.

Saumitou-Laprade, P., P. Vernet, X. Vekemans, S. Billiard, S. Gallina *et al.*, 2017 Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. Evolutionary Applications**:** 1-14.

Tsagkogeorga, G., V. Cahais and N. Galtier, 2012 The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate Ciona intestinalis. Genome Biol Evol 4**:** 740-749.

Unver, T., Z. Wu, L. Sterck, M. Turktas, R. Lohaus *et al.*, 2017 Genome of wild olive and the evolution of oil biosynthesis. Proceedings of the National Academy of Sciences of the United States of America 114**:** E9413-e9422.

Vassiliadis, C., P. Saumitou-Laprade, J. Lepart and F. Viard, 2002 High male reproductive success of hermaphrodites in the androdioecious Phillyrea angustifolia. Evolution 56**:** 1362-1373.

Vernet, P., P. Lepercq, S. Billiard, A. Bourceaux, J. Lepart *et al.*, 2016 Evidence for the long-term maintenance of a rare self-incompatibility system in Oleaceae. New Phytologist 210**:** 1408-1417.

Wallander, E., and V. A. Albert, 2000 Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data. American Journal of Botany 87**:** 1827-1841.

Wright, S., 1939 The Distribution of Self-Sterility Alleles in Populations. Genetics 24**:** 538-552.

Table 1. Comparison of the sex-averaged, male and female linkage maps.

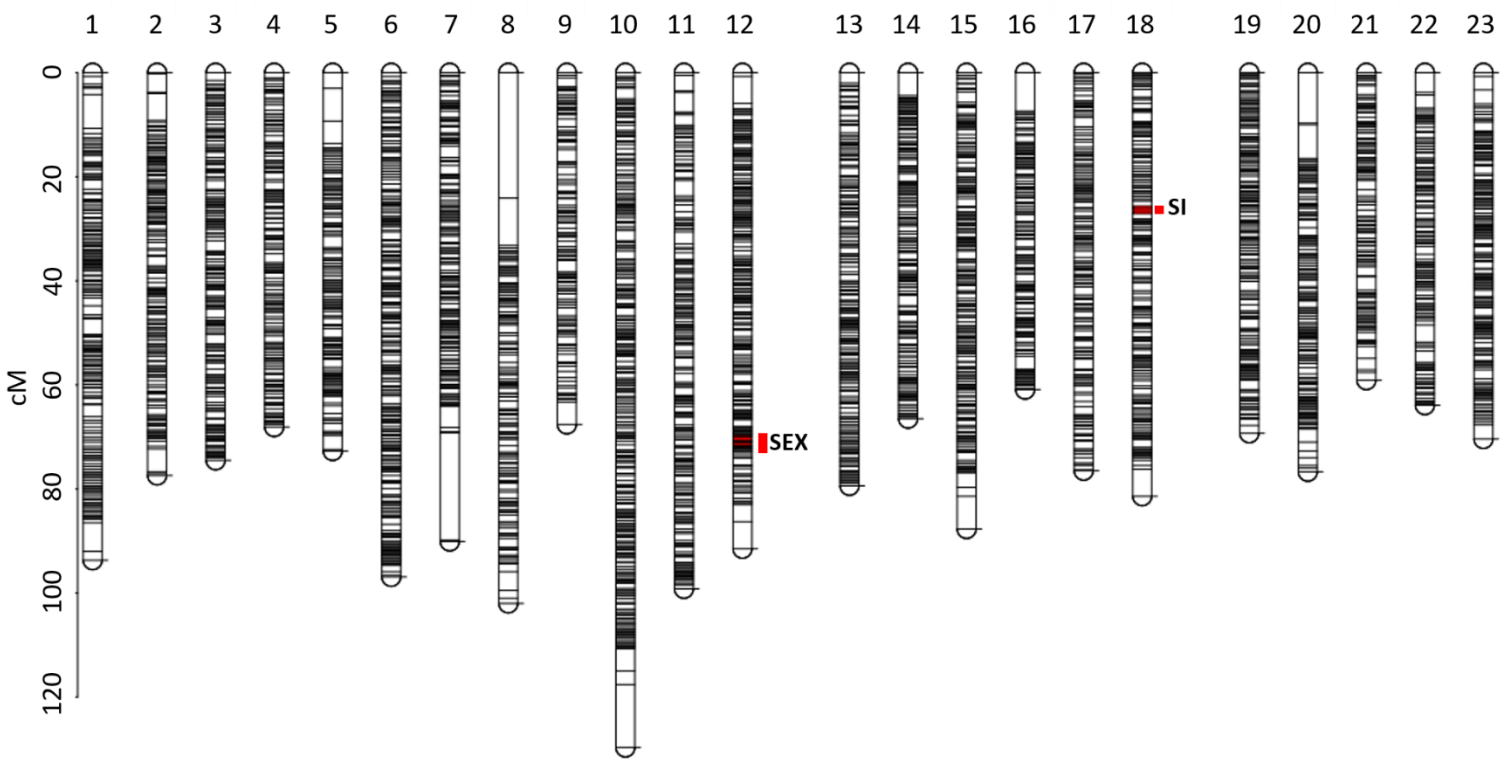
		Sex-averaged map			Paternal map			Maternal map		
Linkage group	**Number of SNPs**	**LG length (cM)**	**SNPs/cM**	**average intermarker distance**	**LG length (cM)**	**SNPs/cM**	**average intermarker distance**	**LG length (cM)**	**SNPs/cM**	**average intermarker distance**
1	854	93.71	9.11	0.11	79.30	10.77	0.09	96.64	8.84	0.11
2	633	77.39	8.18	0.12	76.65	8.26	0.12	68.73	9.21	0.11
3	676	74.50	9.07	0.11	61.94	10.91	0.09	85.42	7.91	0.13
4	535	68.07	7.86	0.13	71.63	7.47	0.13	69.82	7.66	0.13
5	502	72.68	6.91	0.14	60.85	8.25	0.12	85.66	5.86	0.17
6	877	96.89	9.05	0.11	90.81	9.66	0.10	103.63	8.46	0.12
7	609	90.09	6.76	0.15	68.93	8.84	0.11	64.99	9.37	0.11
8	486	101.96	4.77	0.21	103.42	4.70	0.21	119.25	4.08	0.25
9	408	67.63	6.03	0.17	56.06	7.28	0.14	80.15	5.09	0.20
10	1365	129.70	10.52	0.10	131.66	10.37	0.10	133.40	10.23	0.10
11	793	99.24	7.99	0.13	80.40	9.86	0.10	108.84	7.29	0.14
12	973	91.48	10.64	0.09	93.50	10.41	0.10	92.63	10.50	0.10
13	849	79.40	10.69	0.09	77.29	10.98	0.09	80.77	10.51	0.10
14	566	66.47	8.51	0.12	79.06	7.16	0.14	72.88	7.77	0.13
15	750	87.70	8.55	0.12	82.98	9.04	0.11	96.01	7.81	0.13
16	591	60.92	9.70	0.10	56.93	10.38	0.10	69.56	8.50	0.12
17	613	76.52	8.01	0.12	70.58	8.69	0.12	83.94	7.30	0.14
18	660	81.38	8.11	0.12	84.64	7.80	0.13	95.40	6.92	0.14
19	806	69.29	11.63	0.09	91.10	8.85	0.11	79.49	10.14	0.10
20	547	76.71	7.13	0.14	70.07	7.81	0.13	66.91	8.18	0.12
21	479	59.10	8.11	0.12	69.49	6.89	0.15	54.14	8.85	0.11
22	550	63.87	8.61	0.12	60.80	9.05	0.11	67.15	8.19	0.12
23	690	70.43	9.80	0.10	72.26	9.55	0.10	66.91	10.31	0.10
Average	687.48	80.66	8.51	0.12	77.84	8.82	0.12	84.45	8.22	0.13

601 **Figure 1.** *Phillyrea angustifolia* sex-averaged linkage map showing the grouping and position of 10,388

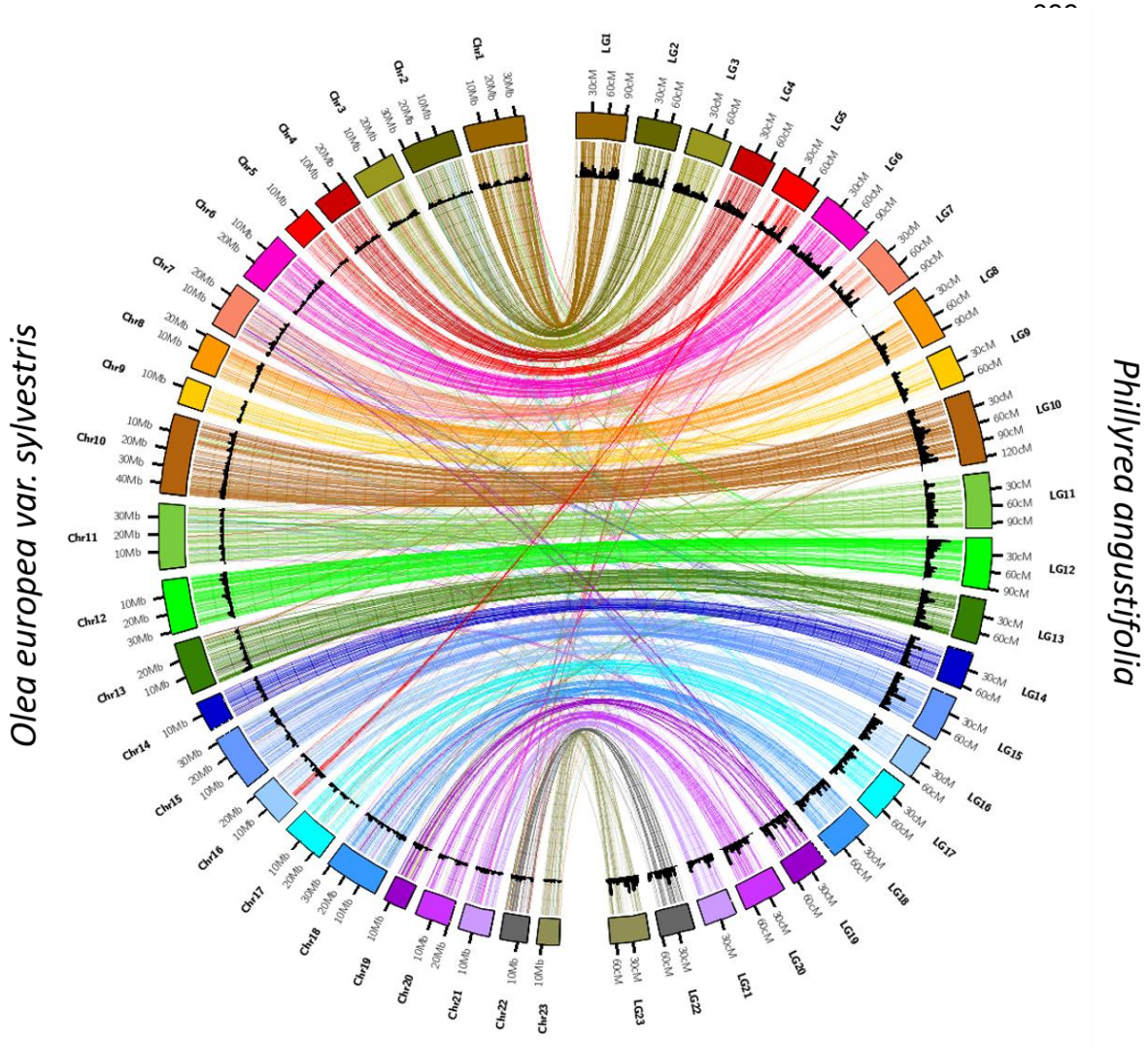
602 SNPs. The length of each of the 23 linkage groups is indicated by the vertical scale in cM. The markers

603 strictly linked to sex and self-incompatibility (SI) phenotypes are shown in red.

604



605 **Figure 2.** Synteny plot identifying homologous *P. angustifolia* linkage groups (LG, scale in cM) with olive
606 tree chromosomes (Chr, scale in Mb). Lines connect markers in the *P. angustifolia* linkage map with their
607 best BLAST hit in the *O. europea* genome. Variation of the density of loci in bins of 3.125cM along linkage
608 groups and 1 Mbp along chromosomes is shown in the inner circle as a black histogram.



618 **Figure 3.** Visualization of chromosome-scale synteny by comparing the location of markers along the *P.*

619 *angustifolia* linkage groups (LG, scale in cM) with the location of their best BLAST hit along the homologous

620 olive tree chromosome (Chr, scale in Mbp). The blue line represents the linear regression and the grey

621 area its 95% credible interval.

622

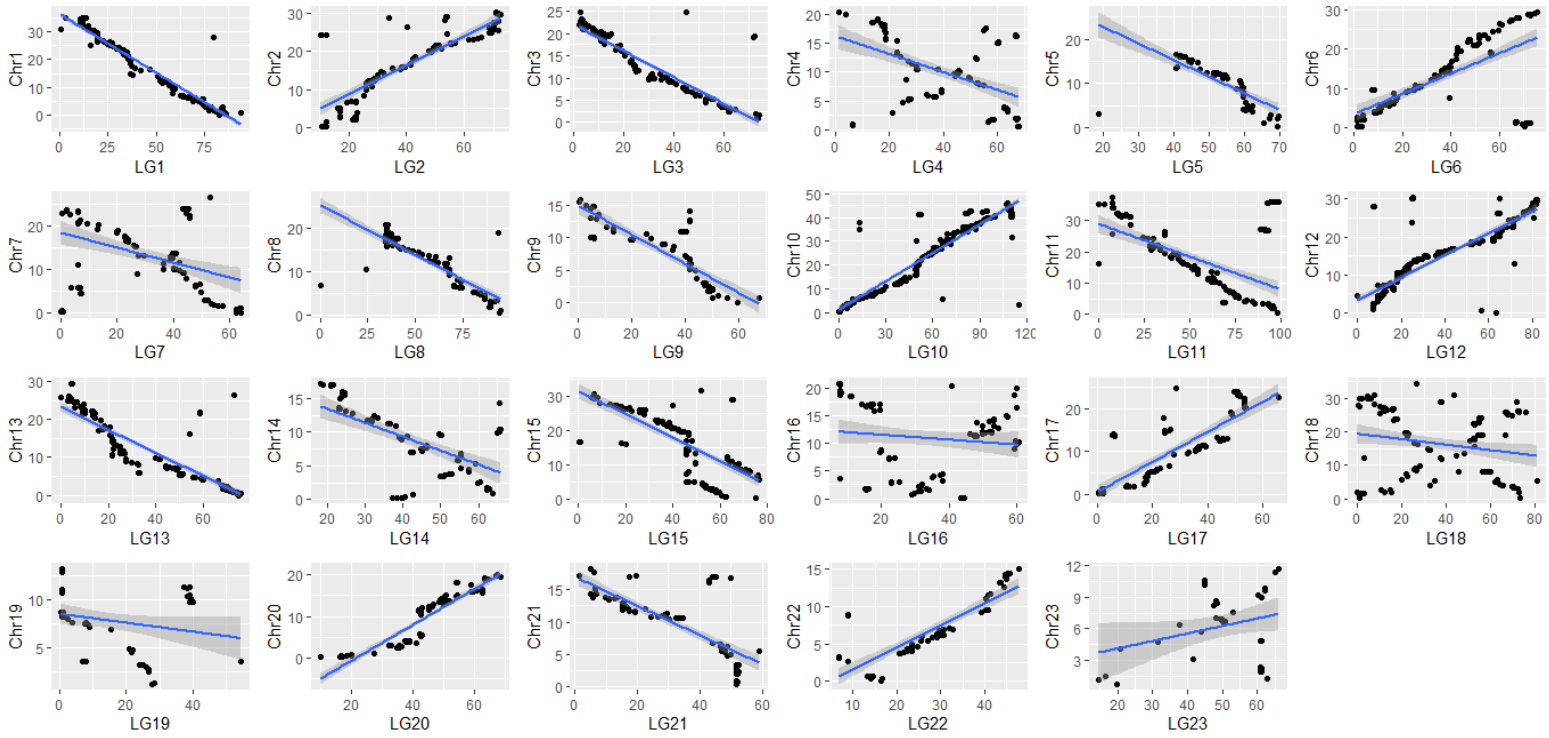
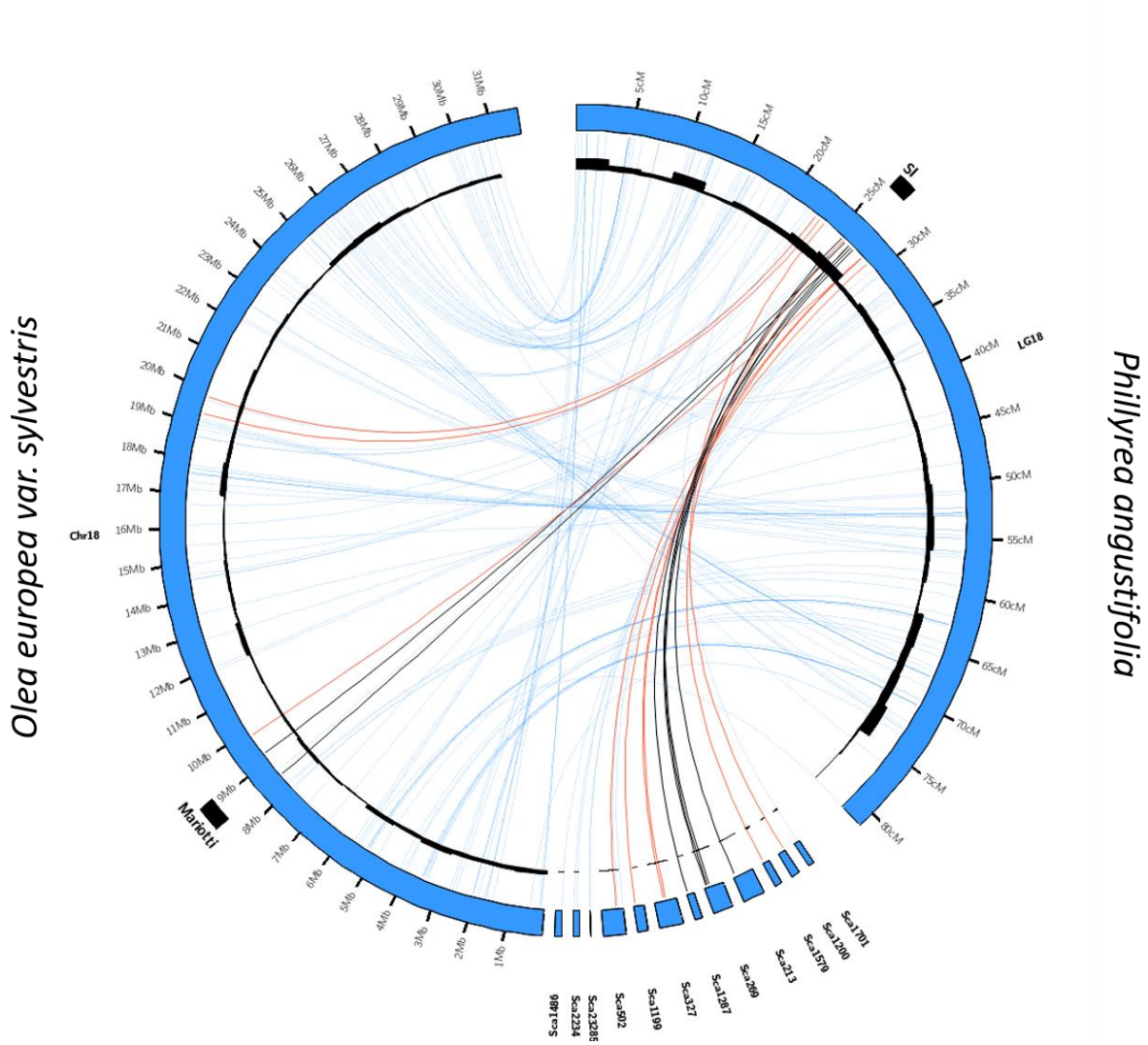
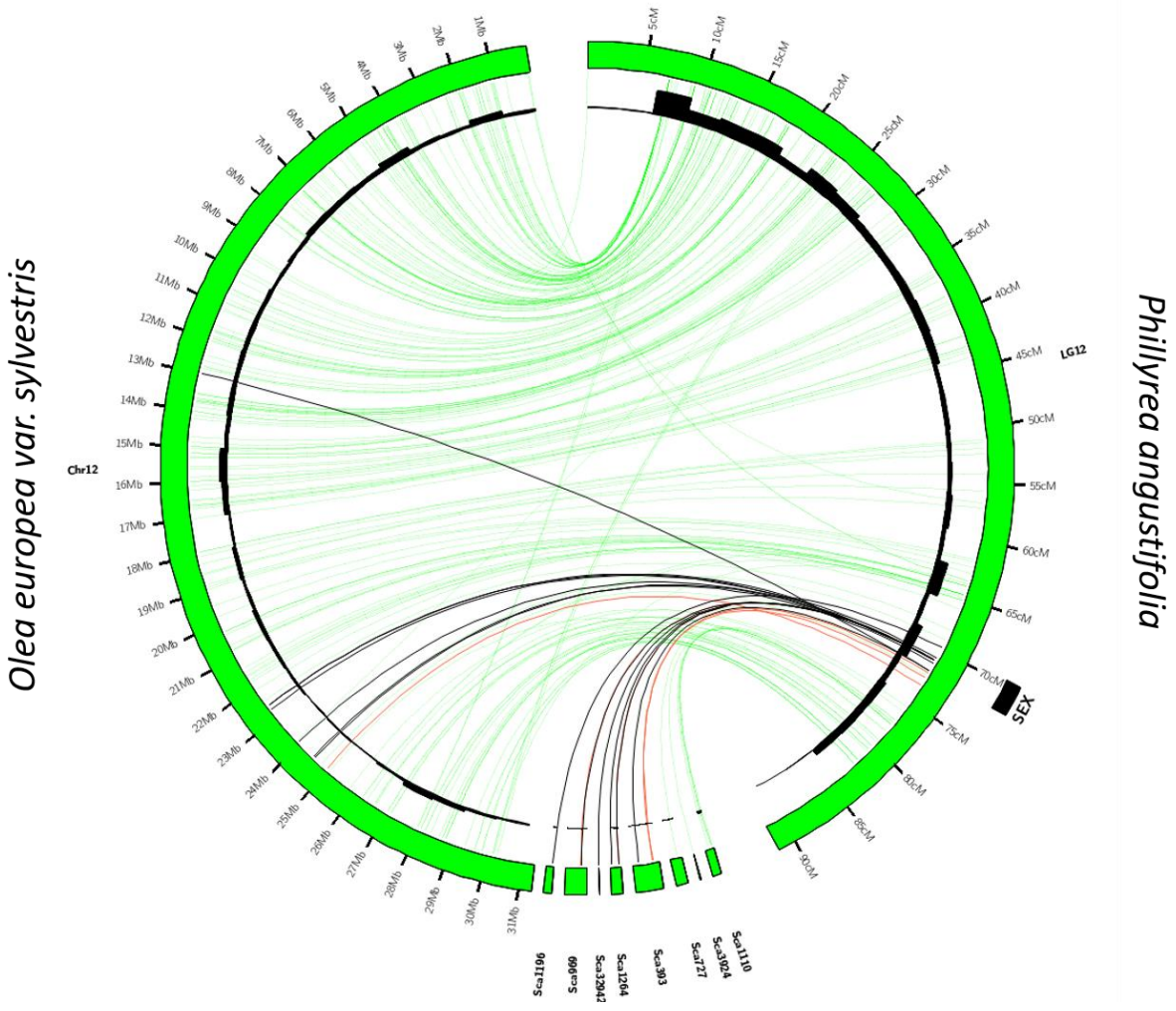


Figure 4. Synteny plot between the *P. angustifolia* linkage group 18 (scale in cM) and the olive tree chromosomes 18 and a series of unanchored scaffolds (scale in Mb). Lines connect markers in the *P. angustifolia* linkage map with their best BLAST hit in the *O. europea* genome. Blue lines correspond to markers with autosomal inheritance. Black lines correspond to markers which strictly cosegregate with SI phenotypes (H_a *vs.* Hb). Red lines correspond to markers with strong but partial (95%) association with SI. The region found to be genetically associated with SI in the olive tree by Mariotti *et al.* (2020) is shown by a black rectangle. Variation of the density of loci in bins of 3.125cM along linkage groups and 1 Mbp along chromosomes is shown in the inner circle as a black histogram.



632 **Figure 5.** Synteny plot between the *P. angustifolia* linkage group 12 (scale in cM) and the olive tree
633 chromosomes 12 and a series of unanchored scaffolds (scale in Mb). Lines connect markers in the *P.*
634 *angustifolia* linkage map with their best BLAST hit in the *O. europea* genome. Blue lines correspond to
635 markers with autosomal inheritance. Black lines correspond to markers which strictly cosegregate with sex
636 phenotypes (males *vs.* hermaphrodites). Red lines correspond to markers with strong but partial (95%)
637 association with sex. Variation of the density of loci in bins of 3.125cM along linkage groups and 1 Mbp
638 along chromosomes is shown in the inner circle as a black histogram.