



HAL
open science

Insertion of a mMoshan transposable element in PpLMI1, is associated with the absence or globose phenotype of extrafloral nectaries in peach [*Prunus persica* (L.) Batsch

Patrick Lambert, Carole Confolent, Laure Heurtevin, Naïma Dlalah, Véronique Signoret, Bénédicte Quilot-Turion, Thierry Pascal

► To cite this version:

Patrick Lambert, Carole Confolent, Laure Heurtevin, Naïma Dlalah, Véronique Signoret, et al.. Insertion of a mMoshan transposable element in PpLMI1, is associated with the absence or globose phenotype of extrafloral nectaries in peach [*Prunus persica* (L.) Batsch. *Horticulture research*, 2022, 9, pp.1-13. 10.1093/hortre/uhab044 . hal-03539501

HAL Id: hal-03539501

<https://hal.inrae.fr/hal-03539501v1>

Submitted on 23 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Title: Insertion of a *mMoshan* transposable element in *PpLMII*, is associated with the absence and globose phenotype of extrafloral nectaries in peach [*Prunus persica* (L.) Batsch]

Running title : a *mMoshan* is associated with the absence of EFNs in peach

Authors :

*Patrick Lambert¹

Carole Confolent^{1, 2}

Laure Heurtevin¹

Naïma Dlalah¹

Véronique Signoret¹

Bénédicte Quilot-Turion¹

Thierry Pascal¹

Affiliations

¹INRAE, GAFL, Montfavet, F-84143, FRANCE

²INRAE, UMR GDEC, Clermont-Ferrand, F-63100, France

Authors email addresses

patrick.lambert.ugaf@inrae.fr

carole.confolent@inrae.fr

laure.heurtevin@inrae.fr

naima.dlalah@inrae.fr

veronique.signoret@inrae.fr

benedicte.quilot-turion@inrae.fr

thierry.pascal84@orange.fr (previous address: thierry.pascal@inrae.fr)

Author for correspondence

Patrick Lambert (Telephone +33 432 722 815; Fax +33 432722 702)

1 **Abstract**

2 Most commercial peach [*Prunus persica* (L.) Batsch] cultivars have leaves with extrafloral nectaries
3 (EFNs). Breeders have selected this character over time, as they observed that the eglandular phenotype
4 resulted in high susceptibility to peach powdery mildew, a major disease of peach trees. EFNs are
5 controlled by a Mendelian locus (*E*), mapped on chromosome 7. However, the genetic factor underlying
6 *E* was unknown. In order to address this point, we developed a mapping population of 833 individuals
7 derived from the selfing of ‘Malo Konare’, a Bulgarian peach cultivar, heterozygous for the trait. This
8 progeny was used to investigate the *E*-locus region, along with additional resources including peach
9 genomic resequencing data, and 271 individuals from various origins used for validation. High-
10 resolution mapping delimited a 40.6 kbp interval including the *E*-locus and four genes. Moreover, three
11 double-recombinants allowed identifying *Prupe.7G121100*, a *LMII-like* homeodomain leucine zipper
12 (HD-Zip) transcription factor, as a likely candidate for the trait. By comparing peach genomic
13 resequencing data from individuals with contrasted phenotypes, a MITE-like transposable element of
14 the hAT superfamily (*mMoshan*) was identified in the third exon of *Prupe.7G121100*. It was associated
15 with the absence and globose phenotype of EFNs. The insertion of the transposon was positively
16 correlated with enhanced expression of *Prupe.7G121100*. Furthermore, a PCR marker designed from
17 the sequence-variants, allowed to properly assign the phenotypes of all the individuals studied. These
18 findings provide valuable information on the genetic control of a trait poorly known so far although
19 selected for a long time in peach.

20 **Introduction**

21 Most peach [*Prunus persica* (L.) Batsch] cultivars have extrafloral nectaries (EFNs), or leaf-glands, on
22 the leaf petioles, stipules, or margins^{1,2}. EFNs are nectar-secreting glands, physically apart from the
23 flowers. They have been observed on a vast diversity of species spanning over 93 families and 332
24 genera^{3,4}. EFNs are mainly known for providing plants with indirect defense against herbivores and
25 fungi, by attracting beneficial predatory arthropods, predominantly ants, and fungivorous mites with
26 their sugary secretions^{3,5,6}. EFNs can enhance plant–mite mutualisms by increasing mite abundance in

27 domatia, indirectly decreasing pathogen load. Therefore, predatory mites keep leaves free of
28 microscopic herbivores while fungivorous mites clean the leaves of detrimental fungi⁷. EFNs may thus
29 increase the potential of EFN-bearing peach cultivars to be protected from damaging organisms by
30 naturally-occurring biological control agents^{5,6}. From an extensive study of the main varieties of the
31 peach, Gregory¹ observed that, for the great majority, gland shapes were well defined and that, in many
32 cases, their shape could serve to separate groups of varieties. Indeed, gland shape was generally
33 homogenous on typical shoots, although some cultivars could exhibit mixed glands. This author
34 identified four main types of leaves, those with reniform (kidney-shape) glands, those with globose
35 glands, glandless leaves and leaves having indistinctive glands. He reported that glands varied in number
36 over the leaves of a same tree and were smaller on the leaf-margin than on the petiole. Gregory¹ also
37 observed that reniform glands were associated with single crenate leaf-margins, whereas leaf-margins
38 were doubly and deeply serrated in eglandular individuals. In the past, fruit breeding programs had
39 inadvertently produced peach cultivars with glandless leaves, yet without determining the effects on
40 either natural enemies or herbivorous pests^{2,8}. Mathews et al.⁵ however, comparing glandular and
41 eglandular peach trees derived from the selfing of the cultivar 'Lovell', observed that those trees with
42 EFNs harbored significantly fewer herbivores than trees without EFNs. The latter also experienced
43 lower growth and fruit production. Earlier, empirical observations showed that the absence of EFNs in
44 peach cultivars resulted in high susceptibility to peach powdery mildew (PPM), one of the major
45 diseases of the peach^{9,10}. Additionally, in wild grape, Weber et al.⁷ demonstrated that adding foliar sugar
46 to plant leaves increased the number of mutualistic mites inhabiting leaf domatia, and this was negatively
47 correlated with the extent of the establishment of grape powdery mildew, a fungal disease similar to
48 PPM. PPM is caused by *Podosphaera pannosa* var. *persicae*¹¹, a member of the Ascomycete fungi,
49 which can be responsible for serious damages in peach orchards. Indeed, the disease may induce necrosis
50 and malformation resulting in unmarketable fruits, premature drop and shoot stunting¹². For this reason,
51 eglandular peach seedlings were systematically discarded during the selection process of most of the
52 breeding programs. The Mendelian inheritance of the leaf-gland phenotype was first described by
53 Connors¹³. The trait has an incomplete dominance, the absence being recessive and the globose shape
54 of the nectaries representing the heterozygous phenotype. Further studies allowed to map the trait on a

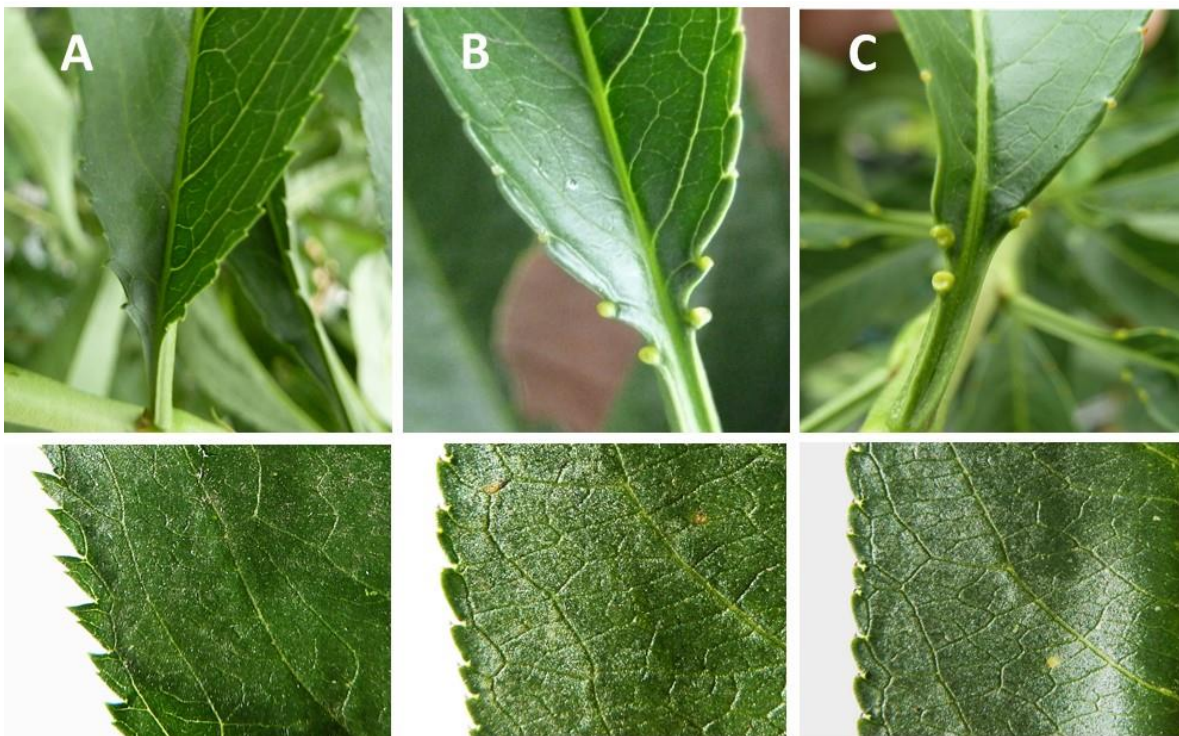
55 single locus (*E*) on chromosome 7 of the peach^{14,15}, but without identifying any factor responsible for
56 the trait and its variations. Furthermore, this same region was found associated with a minor Quantitative
57 Trait Loci (QTL) for resistance to PPM in *P. ferganensis*^{14,16}. These various studies contributed to
58 provide evidence that EFNs might play a role in lowering PPM incidence and could be of most interest
59 for limiting the populations of some classes of detrimental herbivores and fungi in the trees. Therefore,
60 further investigations deserved to be conducted to identify the factor underlying the *E* locus. For this
61 reason, in-depth study of the *E* locus was carried out, in the frame of our breeding program for resistance
62 to pests and diseases in peach. The main objectives of the current work were to develop a high-resolution
63 map of the *E* locus, then investigate the underlying genomic region in order to identify the factor
64 involved in the variation of the leaf-gland phenotype as well as its possible link to susceptibility to PPM,
65 apart from the indirect defense to fungi provided by EFNs. Then, accessorially, develop PCR marker(s)
66 to facilitate early selection of glandular seedlings. With this aim, a large mapping population of 833
67 individuals, referred to as 5392², was developed from the selfing of ‘Malo Konare’ (clone S5392), a
68 canning peach cultivar with globose leaf-glands, from Bulgarian origin¹⁷. This cultivar was selected as
69 it was heterozygous for the trait and part of our breeding program for resistance. Peach genomic
70 resequencing data from contrasted cultivars were used for in-depth investigation of the *E*-locus region.
71 Additional resources including offspring derived from another cross, as well as a collection of contrasted
72 cultivars from various origins were used to support our findings. The outcomes of this study will provide
73 valuable information on a trait little studied in peach so far and more widely in *Prunus* species.
74 Furthermore, they would benefit our breeding program aimed at developing multi-resistant elite peach
75 cultivars.

76 **Results**

77 **Phenotypic evaluation**

78 Seven hundred and seventy-nine progenies out of the initial 833 of the 5392² were observed over two
79 years, among which 197 from the initial population. Two hundred and two (26%) were eglandular, 382
80 (49%) globose and 195 (25%) reniform. This distribution was in agreement with the (1:2:1) segregation
81 ratio expected for a Mendelian trait in this type of population ($\chi^2 = 0.41$). Regarding the BC2, 62

82 individuals had globose leaf-glands and 60 had reniform leaf-glands. As regards the collection, 112
83 cultivars and the two wild peach relatives were scored reniform, 30 globose, four eglandular and one
84 indeterminate (Table S1). For *Prunus kansuensis* S1429, a few phenotypic differences with the other
85 accessions were observed: leaves were homogenous but EFNs were included in the margin of the lower
86 part of the leaf-blade instead of the upper ridge of the petiole. Regarding PER2.3N#1 (S7314), a possible
87 triploid scored indeterminate, no regular leaf-gland was noticeable but a number of small picks on the
88 petiole, close to the leaf-blade. Finally, with respect to leaf-margins, a close association was observed
89 between deeply serrated leaves and the eglandular phenotype in the population 5392². Eglandular
90 individuals had sharp doubly well-defined leaf serrations contrary to those with globose or reniform
91 glands, which had leaves with rounded, shorter crenellations. Crenellations were generally slightly more
92 pronounced in globose individuals (Fig. 1). Regarding the collection, the same association was observed
93 for the four eglandular accessions as compared to the others.

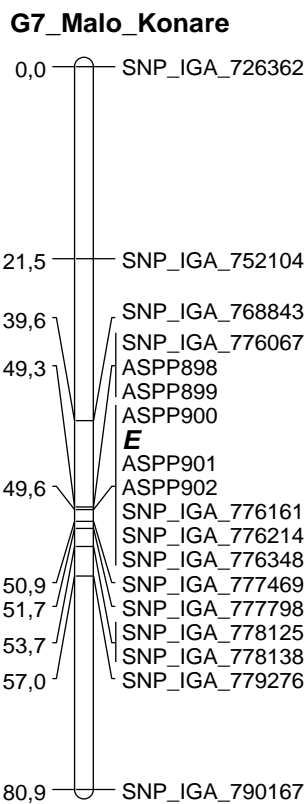


94
95 **Fig. 1 Photographs of the three types of leaves observed in the population 5392².** Top photos show
96 the three different phenotypes observed for the EFNs. Bottom photos show the leaf-margins associated
97 with each of the above phenotypes. (A) Eglandular ‘S10215’, (B) Globose ‘Malo Konare’ (C), Reniform
98 ‘S10216’.

100 **Genetic map of linkage group 7 and high-resolution mapping of the *E* locus**

101 The map of G7 derived from the 212 initial individuals covered a total genetic distance of 80.9 cM (Fig.
102 2) spanning a physical distance of 19,892,186 bp (88.85% of chromosome 7).

103



104

105 **Fig. 2 Genetic map of linkage group 7** of ‘Malo Konare’ developed from the initial mapping
106 population of 212 individuals. The EFN locus (***E***) is in bold and in italics. Genetic distances are in
107 centiMorgan (cM).

108

109 The map was composed of 18 SNPs among which six, including ASPP900, collocated with the *E* locus,
110 at 49.6 cM, spanning a physical distance of 107.8 kbp. The physical distance between the SNPs on either
111 side of the *E*-locus region (SNP_IGA_776067 and SNP_IGA_777469) was 665.7 kbp for a genetic
112 distance of 1.6 cM. No significant deviation of marker segregation was observed ($P < 0.05$). In addition

113 to the above individuals, 567 individuals from the 5392² were genotyped with SNPs included in the *E*-
 114 locus region as well as in the two flanking loci. Forty-eight recombinants were observed in the above
 115 interval, among which twelve between SNP_IGA_776067 and SNP_IGA_776214 (70.5 kbp) and three
 116 double-recombinants between ASPP899 and ASPP901 (Table 1), the latter delimiting an interval of 10.7
 117 kbp including ASPP900 and the *E* locus.

118

119 **Table 1. Recombinant individuals observed in the region of 70.5 kbp between SNP_IGA_776067**
 120 **and SNP_IGA_776214**

Marker on G7	Position on the peach genome sequence v2.0	Predicted gene including the SNP/indel	S5392	S10215	S10216	S5392 ² ₋₅₄	(S5392 ² ₋₆₄)	(S5392 ² ₋₁₂₄)	(S5392 ² ₋₁₃₀)	(S5392 ² ₋₁₅₀)	(S5392 ² ₋₄₂₃)	(S5392 ² ₋₅₂₂)	(S5392 ² ₋₅₃₄)	(S5392 ² ₋₅₄₈)	(S5392 ² ₋₅₇₃)	(S5392 ² ₋₇₉₇)	(S5392 ² ₋₈₁₁)
SNP_IGA_776067	Pp07:14,414,202	Prupe.7G120700	h	b	a	h	b	h	h	b	a	b	h	b	h	b	b
ASPP898	Pp07:14,426,651	Prupe.7G120900	h	b	a	a	b	h	h	b	a	b	b	a	h	b	b
ASPP899	Pp07:14,428,469	Prupe.7G121000	h	b	a	a	b	h	h	b	a	b	b	a	h	b	b
ASPP900	Pp07:14,437,331	Prupe.7G121100	h	b	a	a	h	h	a	b	a	h	b	a	h	h	h
<i>E</i>	-	-	h	b	a	a	h	h	a	b	a	h	b	a	h	h	h
ASPP901	Pp07:14,439,211	Prupe.7G121200	h	b	a	a	b	h	h	b	a	b	b	a	h	h	h
ASPP902	Pp07:14,458,031	-	h	b	a	a	b	h	h	b	a	b	b	a	h	h	h
SNP_IGA_776161	Pp07:14,469,094	Prupe.7G121500	h	b	a	a	b	h	h	h	a	b	b	a	b	h	h
SNP_IGA_776214	Pp07:14,484,739	Prupe.7G121800	h	b	a	a	b	a	h	h	h	b	b	a	b	h	h

S5392 globose, S10215 eglandular, and S10216 reniform haplotypes, are before the twelve recombinant individuals; *a* homozygous reniform, *b* homozygous eglandular, *h* heterozygous; breakpoints are highlighted in grey color.

121

122 *In silico* analysis

123 Based on the above results, investigations were firstly carried out in the region of 10.7 kbp then extended
 124 to the 70.5-kbp genomic region between SNP_IGA_776067 and SNP_IGA_776214 (Positions
 125 Pp07:14,414,202 to Pp07:14,484,739 respectively), for gene and variant discovery. Twelve predicted
 126 genes (Table 2) retrieved from the Genome Database for Rosaceae
 127 (www.rosaceae.org/species/prunus_persica/genome_v2.0.a1) were identified, among which three genes
 128 (*Prupe.7G121000*, *Prupe.7G121100* and *Prupe.7G121200*) were located in the ASPP899-ASPP901
 129 interval.

131 **Table 2 Predicted genes observed in the 70.5-kbp genomic region comprised between**
 132 **SNP_IGA_776067 and SNP_IGA_776214**

Annotated gene	Position on Peach v2.0	Swissprot description/match	TAIR description/match
Prupe.7G120700	Pp07:14410619..14416964	Pyruvate kinase, cytosolic isozyme (<i>Glycine max</i>) /Q42806	Pyruvate kinase family protein/ AT3G52990.1
Prupe.7G120800	Pp07:14417865..14420155	Uncharacterized	Sequence-specific DNA binding transcription factors/ AT3G10040.1
Prupe.7G120900	Pp07:14425004..144427458	Uncharacterized	Glycoside hydrolase family 28 protein/ AT2G33160.1
Prupe.7G121000	Pp07:14428432..14431463	F-box protein PP2-A15 (<i>Arabidopsis thaliana</i>) /Q9LF92	Phloem protein 2-A15/ AT3G53000.1
Prupe.7G121100	Pp07:14436305..14437630	Putative homeobox-leucine zipper protein ATHB-51 (<i>Arabidopsis thaliana</i>) /Q9LZR0	Homeobox 51/ AT5G03790.1
Prupe.7G121200	Pp07:14438623..14440853	60S ribosomal protein L24 (<i>Prunus avium</i>) /Q9FUL4	Ribosomal protein L24e family protein/ AT3G53020.1
Prupe.7G121300	Pp07:14441874..14445148	Protein kinase dsk1 (<i>Schizosaccharomyces pombe</i>) /P36616	Ser/arg-rich protein kinase 4/ AT3G53030.1
Prupe.7G121400	Pp07:144459229..14449576	Uncharacterized	Sulfite exporter TauE/SafE family protein 4/ AT2G36630.1
Prupe.7G121500	Pp07:14467931..14469921	Embryonic protein DC-8 (<i>Daucus carota</i>) /P20075	Embryonic cell protein 63/ AT2G36640.1
Prupe.7G121600	Pp07:14471118..14474276	Probable glycosyltransferase At5g03795 (<i>Arabidopsis thaliana</i>) /Q9FFN2	Exostosin family protein/ AT5G03795.1
Prupe.7G121700	Pp07:14480000..14482717	Pentatricopeptide repeat-containing protein At5g03800 (<i>Arabidopsis thaliana</i>) /Q9FFN1	Pentatricopeptide repeat (PPR) superfamily protein/ AT5G03800.1
Prupe.7G121800	Pp07:14483790..14494599	E3 ubiquitin-protein ligase UPL7 (<i>Arabidopsis thaliana</i>) /Q9SCQ2	Ubiquitin-protein ligase 7/ AT3G53090.2

134 Reads from the eglandular ‘S10215’, the reniform ‘S10216’, ‘Summergrand’ and ‘Pamirskij
135 5’, as well as the globose ‘Zephyr’ and ‘Malo Konare’ were aligned onto Peach v2.0.a1 derived
136 from the reniform peach *cv.* Lovell (Plov2-2N) and compared. A total of two hundred and
137 seventy-seven variants between ‘S10215’ and ‘S10216’ and heterozygous in ‘Malo Konare’,
138 were identified among which six SNPs and an indel in the ASPP899-ASPP901 region (Table
139 S2). However, no relationship was observed between any of the variants and the trait, except
140 for the indel, which clearly differentiated eglandular, reniform and globose accessions. For the
141 other 276 variants, the eglandular ‘S10215’ had the same haplotype as the reniform
142 ‘Summergrand’ and ‘Lovell’ (Plov2-2N), as well as the globose ‘Zephyr’. In contrast, the
143 reniform ‘S10216’ was highly similar to ‘Pamirskij 5’ (reniform), except for an 11-kbp region
144 upstream of the indel, for which ‘Pamirskij 5’ had the same haplotype as the above four other
145 accessions (Table S2). The indel was located in *Prupe.7G121100*, a gene annotated as putative
146 homeobox-leucine zipper protein ATHB-51 (Table 2). According to Gene Ontology,
147 *Prupe.7G121100*, has a DNA-binding transcription factor activity and is involved in bract
148 formation and leaf morphogenesis. Based on these findings, 25 primers (Table S3) were
149 developed from consensus regions between ‘S10215’ and ‘S10216’ in order to sequence the
150 interval encompassing *Prupe.7G121100*, as well as the 100-bp gap which remained in the peach
151 genome sequence reference⁴⁴ (Peach v2.0.a1) immediately upstream of the CG (position
152 Pp07:14436205..14436304). The sequencing of the gap region resulted in sequences 9-fold longer than
153 expected (905 bp and 903 bp for S10215 and S10216 respectively), therefore impacting coordinates
154 downstream (Fig. S1). Sequence comparison allowed identifying a 590-bp insertion in the last
155 coding DNA sequence (CDS) of *Prupe.7G121100*, in the eglandular ‘S10215’, as well as two
156 additional polymorphisms due to differences in the number of CT repeats in two SSRs present in the
157 gap region (Fig. S1). The 590-bp insertion was located between positions Pp07:14437331 and
158 Pp07:14437332, disrupting the initial reading frame (Fig. S1). BLASTN search against NCBI
159 database allowed finding a high similar hit (98% of identity) with an insertion fragment of 588

160 bp, upstream of the start codon of a chalcone isomerase (CHI) gene of peach (Sequence ID:
161 KF990613.1). This insertion was identified as a MITE-like *Moshan* (*mMoshan*) transposable
162 element of the hAT superfamily. BLASTN search with the sequence inserted in
163 *Prupe.7G121100*, against Peach v2.0.a1, returned 91 additional highly-similar hits (> 95%
164 identity) spanning all the chromosomes, all starting from the third 5' nucleotide of the inserted
165 element. The most similar (100% of identity from the third 5' nucleotide) was located on
166 chromosome 5 (Pp05:16,628,569-16,629,156), 460 bp upstream of the start codon of
167 *Prupe.5G208500*, a homolog of AGL8 (agamous-like 8) transcription factor. Nevertheless, a
168 fine analysis of the insertion sequence highlighted some differences with the other transposable
169 elements. The 92 above transposons had terminal inverted repeats (TIRs) composed of 13 or
170 14 complementary nucleotides and 8-nucleotide target site duplication (TSD). Regarding the
171 insertion, the TIR in 3' was composed of 13 nucleotides identical to that of 90 of the 92
172 transposons. However, only 10 nucleotides of the 5' TIR were complementary with those of
173 the 3' TIR (Fig. S1). In addition, no direct repeat sequence was observed at the target insertion
174 site and therefore no TSD. A likely hypothesis is a deletion in the original sequence
175 (GACGAGCCTAGGGGTGGGCAC) where "GACGAGCC" was the TSD, the deleted motif
176 "CGAGCCTAGG" and the original 5' TIR started with the motif "TAGGG".

177

178 ***Analysis with FGENESH***

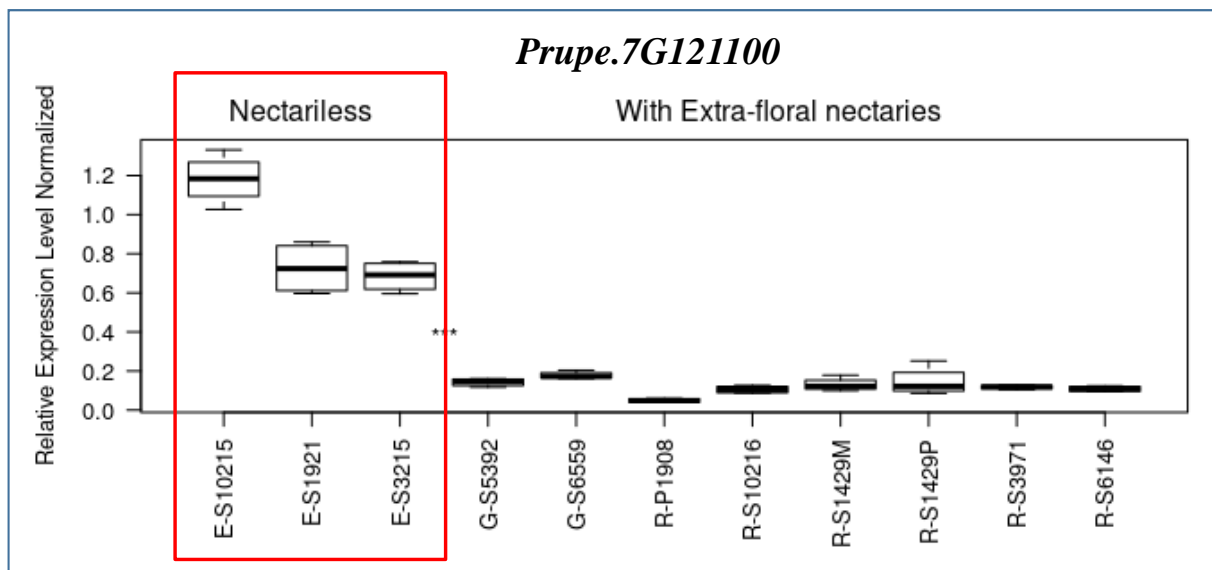
179 The analysis with FGENESH was performed for both variants, using the genomic sequence of
180 *Prupe.7G121100* supplemented by the sequence of the gap region, and various dicot plant species as
181 models, among which *P. persica*, *A. thaliana*, *M. domestica* and *L. esculentum*. One single
182 prediction was obtained with the reniform sequence (Fig. S2). The addition of the gap-region sequence
183 led to a primary transcript composed of three CDS instead of two, the initial start codon being
184 replaced by another one 431 bp upstream. Queries of GDR_RefTransV1 and NCBI database using the

185 sequence of the resulted transcript, validated the prediction, three transcripts
186 (*P.persica_gdr_reftransV1_0044698*, *P. dulcis* LOC117635596 and *P. avium* LOC 110754228) having
187 sequences highly similar with the predicted transcript (100%, 99% and 98% identity respectively). For
188 eglandular accessions, four different predictions were obtained, all of these including an additional CDS
189 in the 3' region. Differences between predictions were linked to the proportion of the transposon
190 included in the third CDS and the position of the fourth.

191 **Expression analysis of *Prupe.7G121100***

192 Relative expression levels of *Prupe.7G121100* in leaves were assessed in three eglandular, two globose
193 and three reniform cultivars, as well as two wild species, *P. davidiana* P1908 and *Prunus kansuensis*
194 S1429 (Fig. 4).

195



196

197 **Fig. 4 Relative expression of *Prupe.7G121100* in ten accessions contrasting for EFNs.** Expressions
198 were normalized with the constitutive genes *PpTEF2* and *PpRPL13*. Eglandular, globose and reniform
199 individuals are denoted by the letter E, G or R, before the accession number, respectively. The three
200 eglandular individuals are framed red. The expression of *P. kansuensis* is represented by two samples:
201 R-S1429M (leaf-margin) and R-S1429P (upper-petiole region).

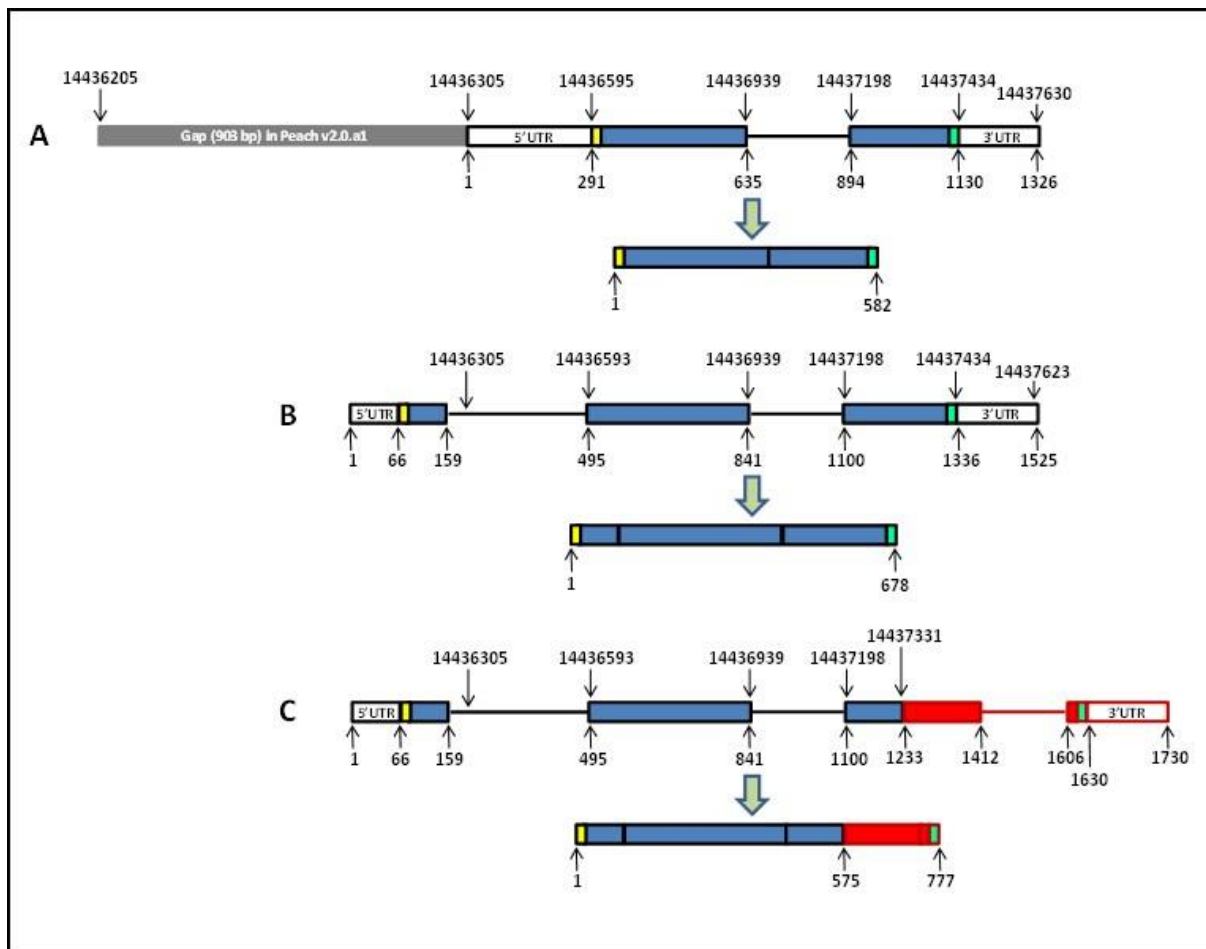
202

203 Contrary to our initial expectations, *Prupe.7G121100* had significantly higher expression levels ($p <$
204 0.001) in eglandular accessions, than in both reniform and globose individuals, either before or after
205 normalization with *PpTEF2* and *PpRPL13*. Normalized differential expressions were comprised
206 between 0.049 ± 0.0052 (mean \pm SE) and 0.1454 ± 0.03675 for reniform individuals, 0.1423 ± 0.01 and
207 0.1778 ± 0.0093 for the globose ones, and between 0.6846 ± 0.0394 and 1.1818 ± 0.0627 for eglandular
208 accessions (Fig. 4), thus showing a negative correlation between the expression level of *Prupe*
209 *7G121100* ($p < 0.001$) and the presence of EFNs. No significant difference was observed between the
210 two samples of *Prunus kansuensis* S1429 ($p < 0.01$). In comparison, differential expression values
211 before normalization, were comprised between 1 ± 0.10 (mean \pm SE) and 2.96 ± 0.75 for reniform
212 individuals, 2.89 ± 0.19 and 3.62 ± 0.20 for the globose ones, and between 13.93 ± 0.77 and $23.79 \pm$
213 0.99 for eglandular accessions.

214 **3' RACE PCR and comparison of the alleles of the transcript**

215 Nested PCRs based on sense primers associated with the AUAP antisense primer gave single amplicons
216 for the reniform accessions only, whereas those carried out on eglandular accessions produced mixtures
217 of amplicons of different sizes (smears). In contrast, those carried out using the antisense primers
218 developed from each of the four predictions, (Table S6) gave the expected results, with amplicons
219 present or absent according to the prediction considered. Sequences derived from the amplicons confirm
220 the insertion of the 179 first nucleotides of the transposon after position 134 of the third exon of the
221 initial transcript, as well as the presence of a fourth exon in the eglandular accessions (Fig. S1 and Fig.3).
222 This confirms that prediction #3, which includes *P. persica* in the model, is the only valid (Fig. S2).
223 Regarding the reniform accessions, the transcript was as expected. The size of the eglandular transcript
224 was 99 nucleotides longer than that of the reniform one (777 and 678 nucleotides respectively) resulting
225 in a larger predicted protein (258 and 225 amino acids respectively). Moreover, major changes were
226 observed in the eglandular transcript : 33 amino acids of the 3' end were replaced by 65 others in the
227 eglandular transcript (Fig. S2).

228



229

230 **Fig. 3 Diagrams of *PpLMI1*.** Spliced transcripts are displayed below their respective primary transcripts
 231 **(A)** Truncated *PpLMI1* (Prupe 7.G121100) as annotated in Peach v2.0.a1. CDSs are shown as blue
 232 rectangles, 5' and 3'-UTR as white rectangles, introns as black lines. The start and stop codons are
 233 shown as yellow and green rectangles respectively. Upper coordinates represent current positions on
 234 Peach v2.0.a1 (Pp07) although they are no longer relevant as the gap upstream of Prupe 7.G121100 is
 235 longer than indicated. Lower coordinates represent the distance from the first nucleotide of the 5'-
 236 UTR. **(B)** *PpLMI1* as observed in reniform individuals. **(C)** *PpLMI1* as observed in eglandular individuals.
 237 Regions not shared with the reniform transcript are shown in red and correspond to transposon
 238 segments (coding sequences, intron and 5'-UTR).

239

240 Genotyping with the ASPP900 marker

241 One thousand and fifty individuals in total were genotyped with the ASPP900 marker, among which
 242 two hundred and seventy-one individuals used for validation, including 149 accessions (Table S1). For
 243 all of them except 'S7314', genotypes were consistent with phenotypes and globose individuals could
 244 also be clearly differentiated from reniform ones. 'S7314' was considered as a possible triploid derived
 245 from the eglandular 'Prosser 2.1N'; however, it was genotyped as globose (heterozygous) and
 246 phenotyped as undifferentiated. These discrepancies do not question the efficiency of ASPP900, but are

247 rather due to its peculiar genotype. In addition, this raises doubts on the single parental origin of this
248 accession.

249 **Discussion**

250 The aim of our study was to identify the genomic factor responsible for the presence/absence of EFNs
251 in peach, a Mendelian trait previously mapped on chromosome 7 (ref. 14, 15), but little studied so far.
252 The fine-mapping approach allowed delimiting the trait to an interval of 10.7 kbp between positions
253 Pp07:14,428,469 and Pp07:14,439,211. Micheletti et al.¹⁸, using the ISPC 9K SNP peach array¹⁹ and a
254 collection comprising 750 reniform and 190 globose accessions, identified a single SNP,
255 SNP_IGA_776161 (position Pp07:14,469,094) as associated to the leaf-gland type. This association was
256 not fully congruent with our observations as ‘S10215’ (eglandular), ‘Zephyr’ (globose),
257 ‘Summergrand’, ‘Rubira’ and the peach genome reference derived from ‘PLov2-2N’ (reniform) had
258 the same allele combination (C/C) for this SNP, whereas ‘S10216’ and ‘Pamirskij 5’, both reniform
259 were T/T. Nevertheless, taking into account the limited number of SNPs on the array corresponding to
260 the trait region, as well as possible misclassification of some individuals of the collection, the results of
261 the two studies were convergent. Coupling the results of the fine-mapping approach with the comparison
262 of the genomic sequences of accessions contrasting for the trait, then allowed to clearly identifying a
263 single candidate gene, *Prupe.7G121100*, among the three genes included in the above interval, and more
264 broadly, among the 12 genes comprised in the longer 70.5-kbp genomic region encompassing the latter.
265 Indeed, *Prupe.7G121100* has two variants: the regular one, associated with the presence of EFNs and
266 homozygous in reniform individuals, and a second one including a 590-bp insertion homozygous in
267 eglandular individuals. This insertion was identified as a MITE-like transposable element of the hAT
268 superfamily, termed as *mMoshan*²⁰. Miniature inverted-repeat transposable elements (MITEs) are non-
269 autonomous class II transposable elements. They are considered a major driving force for generating
270 allelic diversity in plant genomes²¹. MITES account for 3.89% of the peach genome²² with 0.16% for
271 the 491 *Moshan* elements identified. *Moshan* elements are unique to Rosaceae and the *mMoshan* class
272 is predominant with 432 elements²⁰. Interestingly, two of these *mMoshan* elements generated no obvious
273 target site duplication, as the element inserted in *Prupe.7G121100*, suggesting that these three elements

274 were atypical. Wang et al.²⁰ identified 29 *mMoshan* which were inserted in genes, among which 14 in
275 exons. The 29 genes were distributed over all the chromosomes but none in chromosome 7. The
276 *mMoshan* in *Prupe.7G121100* was not detected probably because Peach genome v2.0.a1 was derived
277 from the reniform double haploid ‘Lovell’ Plov2-2N and therefore does not include the inserted element.
278 This author observed that genes including *mMoshan* elements showed relatively lower expression levels
279 compared with genes lacking these elements and this was consistent with previous studies on MITES²³.
280 However, this was not the case in our study since *Prupe.7G121100* demonstrated enhanced expression
281 in eglandular individuals compared to that in globose and reniform ones, which were quite similar.
282 *mMoshan* elements contain several *cis*-regulatory elements such as MYB and WRKY binding sites in
283 the first third of the sequence, which could be involved in upregulation of the transcription²⁰. However,
284 when we take into account the minor differences in gene expression observed between reniform and
285 globose individuals, and the similarity of the phenotypes of their leaf-margins, as well as the incomplete
286 dominance of the trait, this seems not correlated. It would be therefore interesting to investigate
287 possible functional differences between the two alleles This point needs a dedicated approach.
288 *Prupe.7G121100* was annotated as putative homeobox-leucine zipper protein ATHB-51, a member of
289 the class I (HD-Zip I) superfamily of transcription factors. (HD-Zip) proteins are unique to plants. They
290 include the peculiar combination of a DNA-binding homeodomain (HD) and an adjacent Leucine zipper
291 (Zip) motif, which mediates protein-dimer formation²⁴. Saddic et al.²⁵ identified ATHB-51 as a meristem
292 identity regulator and named it LATE MERISTEM IDENTITY (*LMII*) based on its regulation
293 functions; accordingly, we will further refer to *Prupe.7G121100* as *PpLMII*. These authors showed that
294 *LMII* was a direct target of LEAFY (*LFY*), a central meristem identity regulator in *Arabidopsis thaliana*
295 as well as a direct upstream activator of a second meristem identity regulator, the MADS-box
296 transcription factor CAULIFLOWER (*CAL*). *LMII* acts together with *LFY* to induce *CAL* expression,
297 the interaction between these three genes corresponding to a feed-forward loop transcriptional network
298 motif²⁶. *LMII* thus belongs to the complex of genes including others transcription factors, such as
299 APETALA1 (*API*), involved in the meristem identity switch leading to flower formation^{25,27}.
300 Interestingly, the *mMoshan* transposable element identified on chromosome 5, with the highest

301 percentage of identity with that inserted in *PpLMII*, was in the promoter region of an homolog of AGL8
302 (agamous-like 8) transcription factor, a MADS-box negatively regulated by APETALA1, suggesting a
303 possible involvement of APETALA1 in the regulation of *PpLMII*. However, *LMII* has also additional
304 *LFY*-independent roles in leaf morphogenesis and bract formation²⁵. For instance, *LMII* regulates leaf
305 growth in *Arabidopsis thaliana*²⁸ as well as organ proportions such as stipules size, via an
306 endoreduplication-dependent trade-off, that limits tissue size and cell proliferation, through the
307 activation of the mitosis blocker *WEE1*²⁹. Moreover, modifications of *GhLMII-D1b*, one of its
308 homologues³⁰ were found to be responsible for the major leaf shapes in Upland cotton
309 (*Gossypium.hirsutum*). This designates *LMII-LIKE* genes (along with the *KNOXI* genes) as evolutionary
310 hotspots that have been recruited in angiosperms to modify leaf shape³¹. In this way, Chang et al.³²
311 demonstrated that *LMII-like* and *KNOXI* genes coordinately control leaf development in dicotyledons
312 and that different expression patterns of these two genes correspond to the formation of different leaf
313 marginal structures. The same way, loss of function of *CrLMII*, a likely ortholog of *LMII* was reported
314 to decrease leaf serration in *Capsella rubella*³³. This is in agreement with the results of our study, in
315 which increased leaf-margin serration was found strictly associated with the absence of EFNs,
316 concurrently with the higher expression of *PpLMII*. This relationship between leaf serration and absence
317 of EFNs was already reported in previous studies¹³. These findings thus contribute to confirm the
318 possible involvement of *PpLMII* in leaf-margin structures in peach and accordingly in the phenotype of
319 EFNs. Likewise, in cucumber (*Cucumis sativus*), *mict*, a class I HD-Zip factor which sequence had
320 52% of identity with *LMII*, regulates multicellular trichome development³⁴. Regarding the EFNs,
321 however, molecular genetic understanding of their formation is still underdeveloped. No study to date
322 is available in tree species and only a few ones have been published in annual plants. For instance, Hu
323 et al.³⁵ identified *GaNEC1*, a gene encoding a PB1 domain-containing protein, as positive regulator of
324 nectary formation in cotton, which silencing led to a smaller size of foliar nectary phenotype. However,
325 EFNs were located in the leaf midribs and their conformation was different than peach EFNs. Phenotypic
326 diversity usually results from diversity in the genetic organization, regulation and/or expression of
327 underlying developmental programs⁴. In the case of EFNs, such underlying programs have been poorly
328 investigated. The gene *CRABS CLAW (CRC)*, a YABBY transcription factor^{36,37}, appears to be an early-

329 functioning regulator of the development of both floral and extrafloral nectaries in core eudicots^{38,39}.
330 But while the location of floral nectaries may be determined by *CRC* along with several upstream MADS
331 box floral homeotic genes and other unknown regulatory genes³⁸, the development of EFNs may involve
332 the recruitment of different transcriptional control networks than those needed in floral nectaries³⁹. This
333 means that the program involved in EFN development may be closely associated with that of the EFN-
334 bearing organ⁴, the leaf, in the case of peach trees. As a result, the functional characteristics of *LMII*, its
335 involvement in leaf morphogenesis as well as in the meristem identity switch leading to flower
336 formation, suggest that this transcription factor might have a pivotal role in the regulation of different
337 characteristics of the leaf. This makes *PpLMII* a most likely candidate for the presence/absence of EFNs
338 in peach. Therefore, a plausible hypothesis is that functional modification of *PpLMII* associated with
339 the insertion of the HD-Zip I element might trigger endoreduplication. This would result in changes in
340 the cell-wall composition of the lamina as well as in leaf margins, through a developmental program
341 involving target-genes of *PpLMII*, leading notably to serrated leaves and the absence of EFN. In
342 addition, changes in cell-wall composition of the leaf-blade surface could thus make easier the
343 development of fungi, such as *Podosphaera pannosa*, on the leaves. Modification of the cell walls at
344 regions targeted by pathogen attack is a common response to infection and the inability to do so, or the
345 presence of weakened cell walls, might explain, at least in part, the susceptibility to pathogens⁴⁰. As a
346 result, these changes, along with the absence of the positive effects associated with the presence of
347 domitia-inhabiting mutualist mites, and fungivore mites attracted by EFN nectar, might be responsible
348 for enhanced susceptibility to PPM in glandular individuals, as compared to those with EFNs. Further
349 studies need however to be undertaken in order to assess our hypothesis.

350 **Conclusion**

351 In this study, we were interested in identifying the genetic factor responsible for the presence/absence
352 of EFNs in peach. In our knowledge, this is the first time that a molecular genetic approach has been
353 undertaken to clarify the genetic basis of this Mendelian trait, in peach and, more broadly in *Rosaceae*
354 perennial crops. Based on our results, *PpMLII* appears the most likely candidate gene for this character.
355 A comprehensive study of the genomic region including *PpMLII* study did not bring to light another

356 alternative candidate. In addition, its characteristics, regulation functions as a meristem identity
357 regulator as well as its role in leaf morphogenesis, make it highly plausible its involvement in the control
358 of the presence/absence of EFNs as well as its association, at some extent, with the variation of the
359 susceptibility to PPM, in link with cell-wall changes. However, this has to be further validated
360 functionally. Virus induced gene silencing (VIGS) method could be considered as a relevant approach,
361 as genetic transformation in peach is currently an obstacle. In addition, a broader study including the
362 expression of *PpMLII* in the meristem as well as that of genes interacting with *PpMLII* or target genes
363 such as *WEE1*, may be undertaken to elucidate the molecular interactions underlying this interesting
364 trait. This study is thus a first step. Nevertheless, in the short term and from a breeder perspective,
365 ASPP900 marker already allows differentiating the different phenotypes at the seedling level, and could
366 then be used in peach breeding programs.

367 **Material and methods**

368 **Plant material**

369 The initial mapping population included 212 individuals derived from the self-pollination of ‘Malo
370 Konare’ (clone S5392). ‘Malo Konare’ is a canning peach cultivar developed in 1984 at the Fruit-
371 growing Institute in Plovdiv (Bulgaria). It originated from the cross ‘Stoika’ × ‘New Jersey Cling 97’,
372 has globose leaf-glands and shows strong resistance to powdery mildew. ‘Stoika’ was for its part derived
373 from ‘House Kling’ and ‘Ferganskyi Zheltyi’ (1973), a clone of *Prunus ferganensis*. The population
374 was further extended to 833 individuals for the fine mapping of the leaf-gland region and identifying
375 recombinants. This population will be referred to as 5392². In addition, 271 individuals were used to
376 validate phenotype/genotype association in different genetic backgrounds: at first, 149 accessions with
377 contrasting leaf-gland phenotypes (Table S1), including 143 peach cultivars from various origins, two
378 accessions of wild species close to peach, *Prunus davidiana* (Carr.) and *Prunus kansuensis* (Koehne),
379 one accession of *Prunus ferganensis* (Kost. & Rjab.), two double haploids and a possible triploid; then
380 a sample of 122 individuals from a complex breeding population, referred to as BC2⁴¹. The latter was
381 derived from two successive crosses (F₁ and back-cross) including *Prunus davidiana* clone P1908 and
382 peach cv. ‘Summergrand’, followed by a final cross derived from a mixture of pollen of the back-cross

383 population and ‘Zephyr’ as maternal parent. These 271 individuals were planted in triplicate and grown
384 in three different places: greenhouse and tunnels for the cultivars, orchards and tunnels for the BC2. All
385 the individuals were conserved at the *Prunus* Biological Resource Center of INRAE in Montfavet,
386 except the two double haploids and the possible triploid that were conserved at the *Prunus-Juglans*
387 Biological Resource Center, Domaine des Jarres, 33210 Toulence.

388 **DNA isolation**

389 Samples of young leaves from each of the individuals were collected in the spring. Genomic DNA was
390 subsequently isolated using the Qiagen DNeasy 96 Kit (<https://www.qiagen.com>) according to the
391 manufacturer’s instructions. DNA of each sample was at first assessed for quality using a NanoDrop™
392 ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA USA) and then quantified using
393 Quant-iT™ Picogreen® reagent (Invitrogen Ltd.2, Paisley UK). Stock solutions of genomic DNAs were
394 then diluted to a final concentration of 40 ng/μl.

395 **Leaf-gland phenotyping**

396 Leaf-glands were observed over two years, on five to ten leaves from different parts of each of the trees
397 (progenies and cultivars). Individuals were classified under the three phenotypes encountered: reniform,
398 globose and eglandular (no leaf-gland observed). Those trees that were planted in triplicate were scored
399 individually. Leaf-margins were examined concurrently to EFNs, as an association between the
400 eglandular phenotype and deep leaf-serration was previously reported.

401 **Next Generation Sequencing of accessions**

402 Additionally to the reniform double-haploid peach reference ‘Lovell’ (PLov2-2N), which sequence is
403 available at the GDR (https://www.rosaceae.org/species/prunus_persica/genome_v2.0.a1), seven peach
404 accessions were used for genome comparison of the leaf-gland region: ‘Summergrand’, ‘Pamirskij 5’
405 and ‘Rubira’ (reniform), ‘Zephyr’ and ‘Malo Konare’ (globose) and two individuals derived from the
406 self-pollination of ‘Malo Konare’, 5392²_60 (eglandular) and 5392²_76 (reniform) renamed ‘S10215’
407 and ‘S10216’ respectively. These seven accessions were sequenced by MGX GenomiX (Montpellier,
408 France, <http://www.mgx.cnrs.fr>). In brief DNA libraries were prepared using the Nextera DNA Flex

409 Library preparation kit from Illumina (Illumina Inc. San Diego CA, USA) following recommendations
410 provided by the supplier. 125-bp paired-end sequencing was performed utilizing the Illumina HiSeq
411 2500 sequencing platform and the sequence by synthesis (SBS) technique. Base calling was performed
412 by the Real Time Analysis (RTA) software. Raw Illumina paired-end reads were subsequently trimmed
413 using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Potential contaminants
414 were investigated using FastQ Screen software (Babraham Institute) and Bowtie2 aligner ([http://bowtie-](http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)
415 [bio.sourceforge.net/bowtie2/index.shtml](http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)). Resulting reads were aligned onto Peach v2.0.a1 using
416 BWA-MEM (v0.7.12-r1039) and the Ppersica_298_v2.0.fa version. BAM files (*.sorted.bam and
417 *.sorted.bam.bai) were generated in order to visualize sequences under the Integrative Genomics Viewer
418 (IGV) tool⁴².

419 **Marker development and genotyping**

420 ‘Malo Konare’ has been genotyped earlier in the frame of the European project Fruitbreedomics¹⁸, using
421 the IPSC peach 9K SNP array v1¹⁹. Based on the available SNP dataset, a first set of heterozygous SNPs
422 was selected to develop the genetic map of linkage group 7 of ‘Malo Konare’ and insure a sufficient
423 coverage of the group. Genotyping was done using the PCR-based KASP™ (Kompetitive Allele
424 Specific PCR) method from LGC Biosearch technologies (<https://www.biosearchtech.com/>). Primer-
425 triplets (two competitive allele-specific forward primers and one common reverse primer for each
426 marker) were developed from the 60-bp genomic sequence available on either side of the SNPs
427 (https://www.rosaceae.org/species/rosaceae_family_genera/IRSC_SNP_array), using Primer3⁴³ under
428 the following primer-picking conditions: optimal size of the amplicons 75 bp (min 62 bp, max 85 bp),
429 T_m 65°C (min 55°C, max 72°C), primer size 25 bp (min 20 bp, max 32 bp), max self-complementarity
430 7, max 3’ self-complementarity 3, left primer end 61 bp. Primers triplets were compared with Peach
431 v2.0.a1⁴⁴, using the Basic Local Alignment Search Tool (BLAST®) at the Genome Database for
432 Rosaceae (GDR: <https://www.rosaceae.org/blast/>). Those aligning to single positions were selected for
433 genotyping the starting mapping population (Table S4). In a second step, additional SNP markers
434 focused on the interval encompassing the *E* locus were developed in order to identify recombinant
435 individuals. This was done using Next-Generation Sequencing (NGS) data derived from ‘Malo Konare’.

436 BAM files were aligned onto Peach v2.0.a1 and visualized with IGV⁴². The region containing *E* was
437 examined for SNP/indel discovery and reads including heterozygous SNP/indels compatible with the
438 KASPTM method were retrieved. Primer-triplets were then developed as above (Table S4). Mix
439 preparation and PCR reactions were performed using the KASPTM genotyping chemistry and conditions.

440 **Genetic map of linkage group 7**

441 In a first stage, linkage group 7 (G7) of ‘Malo Konare’ was constructed using the mapping dataset
442 derived from the SNP-set selected from Micheletti et al.¹⁸. Genotypic data were coded as F₂-progeny
443 type according to the JoinMap coding system. The leaf-gland trait was similarly coded as a co-dominant
444 Mendelian trait. Linkage analyses were performed using JoinMap 4.1⁴⁵. The recombination fraction
445 value was set at 0.4 and grouping was performed using the independence logarithm of odds (LOD)
446 calculation function and a minimum LOD score threshold of 3. The Kosambi mapping function⁴⁶ was
447 used to translate recombination frequencies into genetic distances. Linkage group 7 was established
448 using regression mapping procedure with three rounds per sample. In a second stage, SNP markers
449 developed for the high-resolution mapping in the interval including the *E*-locus region were added to
450 the genotypic data file and mapped similarly.

451 **High-resolution mapping of the *E* locus**

452 The extended population was genotyped using the SNP markers flanking the *E* locus in the genetic map.
453 Recombinant individuals in the interval were identified and genotyped with newly developed markers.
454 Individuals identified as recombinants in the new interval were genotyped again with a new marker-set.
455 This process was repeated iteratively until no further recombinant was observed.

456 ***In silico* analysis of the region encompassing the *E* locus**

457 The genomic region delimited by the SNP-pairs, which allowed identifying the most informative
458 recombinants, was analyzed for variants. This was done by aligning the sorted.bam files of ‘Malo
459 Konare’, ‘S10215’ and ‘S10216’ onto Peach v2.0.a1 (Ppersica_298_v2.0.fa version), under IGV⁴², and
460 by comparing them. Differences observed were then compared with sorted.bam files of ‘Zephyr’,
461 ‘Summergrand’, ‘Pamirskij 5’ and ‘Rubira’ in order to check consistency of differences regarding leaf-

462 gland phenotype, in different genetic backgrounds. The genomic region defined above was examined
463 for the presence of predicted genes, using JBrowse on the Genome Database for Rosaceae
464 (<https://www.rosaceae.org/jbrowse/>). Positions of the observed differences were compared with those
465 of the genes and their sub-features, then, genomic sequences of the candidate genes (CGs), associated
466 transcripts and predicted protein sequences, homologies and gene functions were downloaded
467 (<https://www.rosaceae.org/node/4017147>). NGS reads corresponding to the position of the selected CG-
468 variants were retrieved for ‘S10215’ and ‘S10216’ using IGV⁴², imported into CLC Main Workbench
469 version 12 (QIAGEN, Aarhus, Denmark), assembled *de novo* and compared using MUSCLE⁴⁷. In
470 addition, as a 100-bp gap remained in Peach genome v2.0.a1 in the region immediately upstream of the
471 most likely CG, 25 primers were developed (Table S3) and used for Sanger sequencing of the gap region
472 and the target CG (Genewiz, South Plainfield, NJ, USA). Assembled sequences were then compared
473 and differences between ‘S10215’ and ‘S10216’ identified. Sequences of the selected CG and the gap
474 region, were finally analyzed for comparison and possible changes in the coding sequences, as well as
475 changes in the resulting protein, using FGENESH gene-prediction program⁴⁸ with different dicot plant
476 species as model (<http://www.softberry.com>).

477 **Gene expression analysis**

478 Eight cultivars with contrasted phenotypes and both wild species (*Prunus davidiana* P1908 and *Prunus*
479 *kansuensis* S1429) were selected for expression analysis. Foliar samples were collected from the part of
480 the leaves including leaf-glands, or from the region including the base of the leaf-blade and the upper
481 part of the petiole for the eglandular individuals. Regarding *Prunus kansuensis*, two samples were
482 collected in order to make comparisons: one from the margin of the leaf-blade where reniform glands
483 were visible, the other from the base, close to the petiole. Samples were immediately frozen in liquid
484 nitrogen. Total RNA was isolated using the Macherey-Nagel[®] NucleoSpin[®] RNA Plant kit (Thermo
485 Fischer Scientific, Waltham, MA USA) following manufacturer’s instructions. RNA concentration and
486 quality were assessed using a NanoDrop[™] ND-1000 spectrophotometer (Thermo Fisher Scientific) and
487 the Agilent 2100 Bioanalyzer System (Agilent, Santa Clara, CA USA). For reverse transcription
488 analysis, primer pairs composed of primers on either side of the second intron of *Prupe.7G121100* (Fig.

489 S3) were designed using Primer3⁴³ and GenScript® Tool (Table S5). One microgram of total RNA per
490 sample was then subjected to cDNA synthesis using the AffinityScript RT kit (Agilent) according to the
491 manufacturer's instructions. A SYBR green real-time PCR assay was thereby carried out in a final
492 volume of 15 µl of a reaction mixture containing 7 µl of 2x Brilliant III SYBR® Green qPCR Master
493 mix (Agilent), 0.5 µM of each primer and 100 ng of cDNA template. Reaction mixtures without cDNA
494 were used as negative controls. Amplification reactions were run in a 96 well plate on a Stratagene
495 Mx3005P (Agilent) under the following conditions: 95°C for 30 s, followed by 40 cycles of denaturation
496 at 95°C for 10 s, annealing at 60°C for 30 s and extension at 72°C for 15 s. Reactions were performed
497 using four biological and three technical replicates for each sample. Amplification values were then
498 normalized using two genes as constitutive controls, as recommended by Bustin et al.⁴⁸: *PpTEF2*
499 (translation elongation factor 2) and *PpRPL13* (60S ribosomal protein L13), both having previously
500 been tested and selected for their stability. Two-way analysis of variance (ANOVA) was used to assess
501 the independent effect of the presence of EFNs and that of the insertion on the expression of
502 *Prupe.7G121100*. Tests were performed using a script of 'RqPCRAnalysis' R-package⁴⁹ customized to
503 generate box-plots with R studio⁵⁰. Significance threshold was set to $p < 0.01$.

504 **3' RACE PCR**

505 Transcripts of reniform and eglandular accessions were amplified using the Invitrogen 3' Rapid
506 Amplification of cDNA Ends (3' RACE) system (Thermo Fisher Scientific). The 3' RACE procedure
507 was carried out as recommended by the supplier. The first strand cDNA was synthesized using 1 µg of
508 total RNA of each of the individuals and the adapter primer (AP) targeting the poly(A) region of the
509 mRNA. The synthesis reaction was followed by the amplification of the target cDNA in a final volume
510 of 50 µl containing 2 µl (1/10) of the above reaction, 1x reaction buffer, 0.2 mM each dNTP, 1.5 mM
511 MgCl₂, 0.2 µM each of the following primers, the antisense abridged universal amplification primer
512 (AUAP) provided in the kit and a custom sense primer developed in the second exon of the gene (Table
513 S6), and 2.5 U of GoTaq® Hot Start Polymerase (Promega). Amplification reactions were run on an
514 Eppendorf Mastercycler egradient (Eppendorf AG, Hamburg, Germany) under the following
515 conditions: 94°C (2min) followed by 35 cycles at 94°C (45 sec), 57°C (45 sec), 72°C (1.5 min) and a

516 final extension at 72°C (5 min). PCR products were visualized in a 1.5 % agarose gel stained with
517 ethidium bromide. Nested amplification reactions were then carried out as above using 1/5 of the second
518 reaction, the antisense AUAP primer and additional custom sense primers developed downstream of the
519 first sense primer. In addition, as numerous stretches of poly(A) were included in the sequence of the
520 transposon, which interfered in the hybridization of the adapter primer (AP) to the poly(A) region of the
521 mRNA, antisense primers were developed based on each of the predictions derived from the analysis
522 with FGENESH (Table S6). PCRs were carried out as above except for the annealing temperature which
523 was lowered to 55°C. Resulting amplicons were then sent for sequencing (Genewiz). Finally, upstream
524 regions were amplified and sequenced to obtain the complete transcripts.

525 **Development of the diagnostic marker ASPP900**

526 One primer-triplet based on the PCR-based KASPT[™] method (<https://www.biosearchtech.com/>) was
527 developed in order to differentiate each of the three phenotypes encountered (Table S4). It was
528 composed of one forward primer specific of the glandular phenotype (20 nucleotides in CDS 3 of
529 *Prupe.7G121100* starting 15 nucleotides before the insertion position), one forward primer specific of
530 the eglandular phenotype (18 nucleotides astride the 9 last nucleotides of the transposon and the first 9
531 nucleotides of CDS3 after the insertion), and one common reverse primer (20 nucleotides in CDS3,
532 starting 75 nucleotides downstream of the insertion point). Positions of the primers on the sequence are
533 shown in Fig. S1.

534 **Acknowledgements**

535 We thank Amélie Emanuel currently at the UMR-BPMP (Montpellier) and Céline Roques for their
536 involvement in the initial genotyping and phenotyping work of the mapping population. We are grateful
537 to Henri Duval from GAFL, for having managed the PeachReseq project with MGX. We thank
538 Christophe Tuero from GAFL, for his help in the phenotyping and Marine Delmas from the *Prunus-*
539 *Juglans* Genetic Resource Centre, Domaine des Jarres, 33210 Toulence (France) for providing peach
540 samples. We also thank Caroline Le Baron for her help with 3' RACE PCR and we are grateful to the
541 technical staff of the experimental domains of 'Saint Maurice' and 'Les Pins de l'Amarine' (INRAE-

542 UGAFL) for their technical contribution to tree management of the peach collection and the population
543 5392². Finally we thank Jean-Luc Gallois (UGAFL) for his help and suggestions as well as Jean-Luc
544 Poëssel (UGAFL) for proofreading the manuscript.

545 **Conflict of interests**

546 The authors declare no potential conflict of interests of any kind.

547 **Data Availability**

548 The datasets supporting the current study are available from the corresponding author upon request, in
549 strict accordance with the policy of the Journal.

550 **Supplementary information** accompanies the manuscript on the *Horticulture Research* website
551 <http://www.nature.com/hortres>

552 **References**

553
554 1 Gregory, C.T. The taxonomic value and structure of the peach leaf glands. *N.Y. Cornell Agric. Exp.*
555 *Sta. Bull.* **365**, 183-222 (1915).

556 2 Okie, W.R. Handbook of peach and nectarine varieties. USDA-ARS, Byron, GA (1998).

557 3 Koptur, S. Extrafloral nectary-mediated interactions between insects and plants. *Insect-Plant*
558 *Interactions*. Vol. IV (ed. E.A. Bernays), pp. 81-129. CRC Press, Boca Raton, FL, USA (1992).

559 4 Marazzi, B., Bronstein, J.L. & Koptur, S. The diversity, ecology and evolution of extrafloral nectaries:
560 current perspectives and future challenges. *Ann. Bot.* **111**, 1243-1250 (2013).

561 5 Mathews, C.R., Bottrell, D.G. & Brown, M.W. Extrafloral nectaries alter arthropod community
562 structure and mediate peach (*Prunus persica*) plant defense. *Ecological Applications* **19**, 722-
563 730 (2009).

564 6 Jones, I.M., Koptur, S. & von Wettberg, E.J. The use of extrafloral nectar in pest management:
565 overcoming context dependence. *J. of Appl.Ecol.* **54**, 489-499 (2017).

- 566 7 Weber, M.G., Porturas, L.D. & Taylor, S.A. Foliar nectar enhances plant–mite mutualisms: the effect
567 of leaf sugar on the control of powdery mildew by domatia-inhabiting mites. *Ann. Bot.* **118**(3),
568 459-466 (2016).
- 569 8 Scorza, R. & Sherman, W.B. Peaches. pp 325-440 In: Janick J, Moore JN (eds.) Fruit breeding,
570 Volume 1: Tree and tropical fruits. John Wiley & Sons, NY (1996).
- 571 9 Watking, W. & Brown A.G. Genetic response to selection in cultivated plants: Gene frequencies in
572 varieties of *Prunus persica*. *Proc. Soc. Lond. B. Biol. Sci.* **145**, 337-347 (1956).
- 573 10 Saunier, R. Contribution to the study of relationships between certain characteristics of simple genetic
574 determination in the peach tree and susceptibility of peach cultivars to oidium, *Sphaerotheca*
575 *pannosa* (Wallr.). *Lev. Ann. Amelior. Plant* **441**, 235-243 (1973).
- 576 11 Weinhold, A.R. The orchard development of peach powdery mildew. *Phytopathology* **51**, 478-481
577 (1961).
- 578 12 Pascal, T., Pfeiffer, F. & Kervella, J. Powdery mildew in the peach cultivar Pamirkij 5 is genetically
579 linked to the *Gr* gene for leaf color. *Hort. Sci.* **45**(1), 150-152 (2010).
- 580 13 Connors, C.H. Inheritance of foliar glands of the peach. *Proc. Am. Soc. Hortic. Sci.* **18**, 21-27 (1921).
- 581 14 Dettori M.T., Quarta, R. & Verde, I. A peach linkage map integrating RFLPs, SSRs, RAPDs, and
582 morphological markers. *Genome* **44**, 783–90 (2001).
- 583 15 Dirlewanger, E. et al. Comparative mapping and marker-assisted selection in Rosaceae fruit crops.
584 *Proc. Natl. Acad. Sci. USA* **101**, 9891-9896 (2004).
- 585 16 Verde, I., Quarta, R., Cedrola, C. & Dettori, M.T. QTL analysis of agronomic traits in a BC1 peach
586 population. *Acta Hortic.* **592**, 291-297 (2002).
- 587 17 Dabov, S. Malo Konare - a new canning peach variety. *Rastenievj Nauki* (1985).

- 588 18 Micheletti, D. et al. Whole-genome analysis of diversity and SNP-major gene association in peach
589 germplasm. *PloS one* **10**(9), e0136803. <https://doi.org/10.1371/journal.pone.0136803>
590 (2015)
- 591 19 Verde, I. et al. Development and evaluation of a 9 K Array for peach by internationally coordinated
592 SNP detection and validation in breeding germplasm. *PloS one* **7**(4), e35668. [https://doi:](https://doi.org/10.1371/journal.pone.0035668)
593 [10.1371/journal.pone.0035668](https://doi.org/10.1371/journal.pone.0035668) (2012).
- 594 20 Wang, L. et al. Evolutionary origin of Rosaceae-specific active non-autonomous hAT elements and
595 their contribution to gene regulation and genomic structural variation. *Plant Mol. Biol.* **91** (1-
596 2), 179-191 (2016).
- 597 21 Bennetzen, J.L. & Wang, H. The contributions of transposable elements to the structure, function,
598 and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530 (2014).
- 599 22 Chen, J., Hu Q., Zhang, Y., Lu C & Kuang, H. P-Mite : a database for plant miniature inverted-repeat
600 transposable elements. *Nucleic Acids Res.* **42**, Database issue (2013).
- 601 23 Lu, C., Chen J., Zhang, Y., Hu, Q., Su W & Kuang, H. Miniature inverted-repeat transposable
602 elements (MITEs) have been accumulated through amplification bursts and play important
603 roles in gene expression and species diversity in *Oryza sativa*. *Mol. Biol. Evol.* **29**, 1005-1017
604 (2012).
- 605 24 Henriksson, E. et al. Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns
606 and phylogenetic relationships. *Plant Physiology* **139**, 509–518 (2005).
- 607 25 Saddic, L.A. et al. The LEAFY target LMI1 is a meristem identity regulator and acts together with
608 LEAFY to regulate expression of *CAULIFLOWER*. *Development* **133**, 1673-1682 (2006).
- 609 26 Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl.*
610 *Acad. Sci USA* **100**(21), 11980–11985 (2003).

- 611 27 Xu, M. et al. Arabidopsis BLADE-ON-PETIOLE1 and 2 promote floral meristem fate and
612 determinacy in a previously undefined pathway targeting APETALA1 and AGAMOUS-
613 LIKE. *The Plant Journal* **63**, 974–989 (2010).
- 614 28 Vlad, D. et al. Leaf shape evolution through duplication, regulatory diversification, and loss of a
615 homeobox gene. *Science* **343**, 780–787 (2014).
- 616 29 Vuolo, F. et al. LMI1 homeodomain protein regulates organ proportions by spatial modulation of
617 endoreduplication. *Genes Dev.* **33**, 377 (2018).
- 618 30 Andres, R.J. et al. Modifications to a LATE MERISTEM IDENTITY1 gene are responsible for the
619 major leaf shapes of Upland cotton (*Gossypium hirsutum* L.). *Proc. Natl. Acad. Sci USA*
620 **114**(1), E57-E66 (2016).
- 621 31 Maugarny-Calès, A. & Laufs, P. Getting leaves into shape: a molecular, cellular, environmental and
622 evolutionary view. *Development* **145**, dev161646 (2018).
- 623 32 Chang, L.J., Mei, G.F., Hu, Y., Deng, J.Q. & Zhang, T.Z. LMI1-like and KNOX1 genes coordinately
624 regulate plant leaf development in dicotyledons. *Plant Mol. Biol.* **99**, 449-460 (2019).
- 625 33 Sicard, A. et al. Repeated evolutionary changes of leaf morphology caused by mutations of a
626 homeobox gene. *Curr. Biol.* **24**(16), 1880-1886 (2014).
- 627 34 Zhao, J.L. et al. Transcriptome analysis in *Cucumis sativus* identifies genes involved in multicellular
628 trichome development. *Genomics* **105**, 296–303 (2015).
- 629 35 Hu, W. et al. Genetic and evolution analysis of extrafloral nectary in cotton. *Plant Biotech. Jour.* **18**
630 (10), 2081-2095 (2020).
- 631 36 Bowman, J.L. & Smyth, D.R. CRABS CLAW, a gene that regulates carpel and nectary development
632 in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains.
633 *Development* **126**, 2387–2396 (1999).

- 634 37 Gross, T., Broholm, S. & Becker, A. CRABS CLAW Acts as a Bifunctional Transcription Factor in
635 Flower Development. *Front Plant Sci.* **9**, 835 (2018).
- 636 38 Lee, J.Y. et al. Activation of *CRABS CLAW* in the nectaries and carpels of Arabidopsis. *Plant Cell*
637 **17**, 25–36 (2005a).
- 638 39 Lee, J.Y. et al. Recruitment of *CRABS CLAW* to promote nectary development within the eudicot
639 clade. *Development* **132**, 5021–5032 (2005b).
- 640 40 Bhosale, R., Maere, S. & De Veylder, L. Endoreplication as a potential driver of cell wall
641 modifications. *Curr. Opin. in Plant Biol.* **51**, 58-65 (2019).
- 642 41 Desnoues, E. et al. Dynamic QTLs for sugars and enzyme capacities provide an overview of genetic
643 control of sugar metabolism during peach fruit development. *J. Exp. Bot.* **67**(11), 3419–3431
644 (2016).
- 645 42 Thorvaldsson, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-
646 performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**(2),
647 178-192 (2013).
- 648 43 Untergasser, A. et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **40**(15), e115
649 (2012).
- 650 44 Verde, I. et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing
651 improve chromosome-scale assembly and contiguity. *BMC Genomics* **18**, 225 (2017).
- 652 45 Van Ooijen, J.W. JoinMap® 4.1, Software for the calculation of genetic linkage maps in experimental
653 populations of diploid species. Kyazma B.V., Wageningen, Netherlands (2012).
- 654 46 Kosambi, D.D. The estimation of map distance from recombination values. *Ann. Eugenics* **12**, 172-
655 175 (1944).
- 656 47 Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids*
657 *Res.* **47**(W1), 636–641 (2019).

658 48 Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D. Automatic annotation of eukaryotic genes,
659 pseudogenes and promoters. *Genome Biol.* 7, Suppl 1: P. 10.1-10.12 (2006).

660 49 Bustin, S.A. et al. The MIQE guidelines: minimum information for publication of quantitative real-
661 time PCR experiments. *Clin. Chem.* **55**, 611-622 (2009).

662 50 Hilliou, F. & Tran, T. RqPCRAnalysis: Analysis of Quantitative Real-time PCR Data. *Proceedings*
663 *of the International Conference on Bioinformatics Models, Methods and Algorithms* pages
664 202-211 (2013).

665 51 R Core Team: A language and environment for statistical computing. R Foundation for Statistical
666 Computing, Vienna, Austria. URL. <https://www.R-project.org/>. (2020).