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

Title: Insertion of a *mMoshan* transposable element in *PpLMI1*, 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 Lambert[1]

Carole Confolent[1, 2]

Laure Heurtevin[1]

Naïma Dlalah[1]

Véronique Signoret[1]

Bénédicte Quilot-Turion[1]

Thierry Pascal[1]

Affiliations

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

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

Authors email adresses

patrick.lambert.ugafl@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)

Abstract

Most commercial peach [*Prunus persica* (L.) Batsch] cultivars have leaves with extrafloral nectaries (EFNs). Breeders have selected this character over time, as they observed that the eglandular phenotype resulted in high susceptibility to peach powdery mildew, a major disease of peach trees. EFNs are controlled by a Mendelian locus (*E*), mapped on chromosome 7. However, the genetic factor underlying *E* was unknown. In order to address this point, we developed a mapping population of 833 individuals derived from the selfing of 'Malo Konare', a Bulgarian peach cultivar, heterozygous for the trait. This progeny was used to investigate the *E*-locus region, along with additional resources including peach genomic resequencing data, and 271 individuals from various origins used for validation. High-resolution mapping delimited a 40.6 kbp interval including the *E*-locus and four genes. Moreover, three double-recombinants allowed identifying *Prupe.7G121100*, a *LMI1-like* homeodomain leucine zipper (HD-Zip) transcription factor, as a likely candidate for the trait. By comparing peach genomic resequencing data from individuals with contrasted phenotypes, a MITE-like transposable element of the hAT superfamily (*mMoshan*) was identified in the third exon of *Prupe.7G121100*. It was associated with the absence and globose phenotype of EFNs. The insertion of the transposon was positively correlated with enhanced expression of *Prupe.7G121100*. Furthermore, a PCR marker designed from the sequence-variants, allowed to properly assign the phenotypes of all the individuals studied. These findings provide valuable information on the genetic control of a trait poorly known so far although selected for a long time in peach**.**

Introduction

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

domatia, indirectly decreasing pathogen load. Therefore, predatory mites keep leaves free of microscopic herbivores while fungivorous mites clean the leaves of detrimental fungi[7]. EFNs may thus increase the potential of EFN-bearing peach cultivars to be protected from damaging organisms by naturally-occurring biological control agents[5,6]. From an extensive study of the main varieties of the peach, Gregory[1] observed that, for the great majority, gland shapes were well defined and that, in many cases, their shape could serve to separate groups of varieties. Indeed, gland shape was generally homogenous on typical shoots, although some cultivars could exhibit mixed glands. This author identified four main types of leaves, those with reniform (kidney-shape) glands, those with globose glands, glandless leaves and leaves having indistinctive glands. He reported that glands varied in number over the leaves of a same tree and were smaller on the leaf-margin than on the petiole. Gregory[1] also observed that reniform glands were associated with single crenate leaf-margins, whereas leaf-margins were doubly and deeply serrated in eglandular individuals. In the past, fruit breeding programs had inadvertently produced peach cultivars with glandless leaves, yet without determining the effects on either natural enemies or herbivorous pests[2,8]. Mathews et al.[5] however, comparing glandular and eglandular peach trees derived from the selfing of the cultivar 'Lovell', observed that those trees with EFNs harbored significantly fewer herbivores than trees without EFNs. The latter also experienced lower growth and fruit production. Earlier, empirical observations showed that the absence of EFNs in peach cultivars resulted in high susceptibility to peach powdery mildew (PPM), one of the major diseases of the peach[9,10]. Additionally, in wild grape, Weber et al.[7] demonstrated that adding foliar sugar to plant leaves increased the number of mutualistic mites inhabiting leaf domatia, and this was negatively correlated with the extent of the establishment of grape powdery mildew, a fungal disease similar to PPM. PPM is caused by *Podosphaera pannosa* var. *persicae*[11], a member of the Ascomycete fungi, which can be responsible for serious damages in peach orchards. Indeed, the disease may induce necrosis and malformation resulting in unmarketable fruits, premature drop and shoot stunting[12]. For this reason, eglandular peach seedlings were systematically discarded during the selection process of most of the breeding programs. The Mendelian inheritance of the leaf-gland phenotype was first described by Connors[13]. The trait has an incomplete dominance, the absence being recessive and the globose shape of the nectaries representing the heterozygous phenotype. Further studies allowed to map the trait on a

single locus (*E*) on chromosome 7 of the peach[14,15], but without identifying any factor responsible for the trait and its variations. Furthermore, this same region was found associated with a minor Quantitative Trait Loci (QTL) for resistance to PPM in *P. ferganensis*[14,16]. These various studies contributed to provide evidence that EFNs might play a role in lowering PPM incidence and could be of most interest for limiting the populations of some classes of detrimental herbivores and fungi in the trees. Therefore, further investigations deserved to be conducted to identify the factor underlying the *E* locus. For this reason, in-depth study of the *E* locus was carried out, in the frame of our breeding program for resistance to pests and diseases in peach. The main objectives of the current work were to develop a high-resolution map of the *E* locus, then investigate the underlying genomic region in order to identify the factor involved in the variation of the leaf-gland phenotype as well as its possible link to susceptibility to PPM, apart from the indirect defense to fungi provided by EFNs. Then, accessorily, develop PCR marker(s) to facilitate early selection of glandular seedlings. With this aim, a large mapping population of 833 individuals, referred to as 5392², was developed from the selfing of 'Malo Konare' (clone S5392), a canning peach cultivar with globose leaf-glands, from Bulgarian origin[17]. This cultivar was selected as it was heterozygous for the trait and part of our breeding program for resistance. Peach genomic resequencing data from contrasted cultivars were used for in-depth investigation of the *E*-locus region. Additional resources including offspring derived from another cross, as well as a collection of contrasted cultivars from various origins were used to support our findings. The outcomes of this study will provide valuable information on a trait little studied in peach so far and more widely in *Prunus* species. Furthermore, they would benefit our breeding program aimed at developing multi-resistant elite peach cultivars.

Results

Phenotypic evaluation

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

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

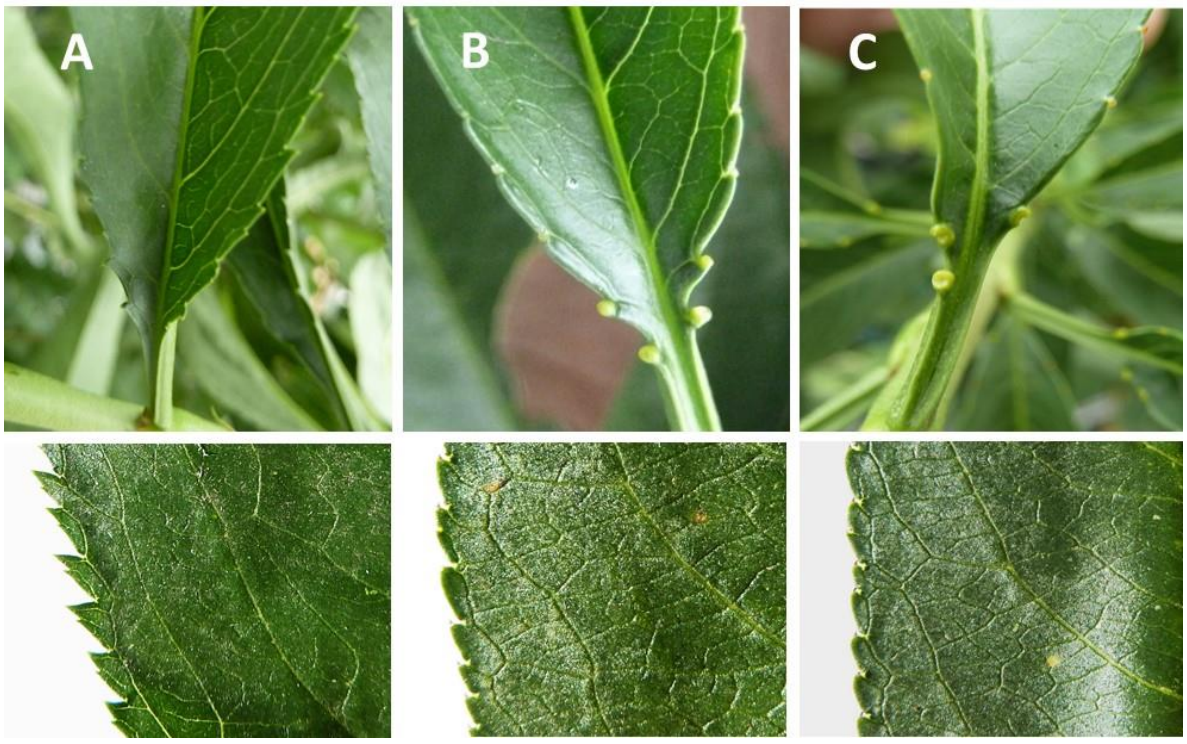

Fig. 1 Photographs of the three types of leaves observed in the population 5392². Top photos show the three different phenotypes observed for the EFNs. Bottom photos show the leaf-margins associated with each of the above phenotypes. (A) Eglandular 'S10215', (B) Globose 'Malo Konare' (C), Reniform 'S10216'.

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

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

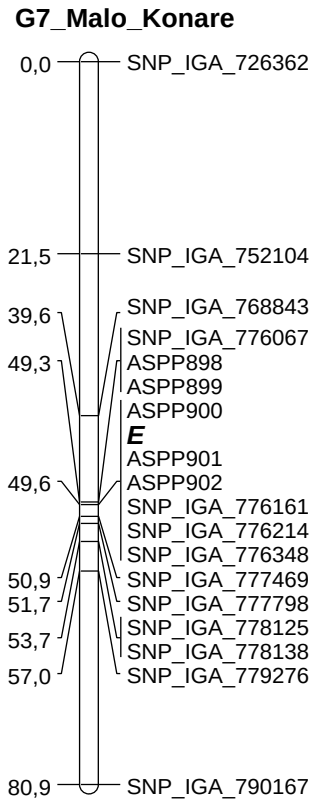

Fig. 2 Genetic map of linkage group 7 of 'Malo Konare' developed from the initial mapping population of 212 individuals. The EFN locus (***E***) is in bold and in italics. Genetic distances are in centiMorgan (cM).

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

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

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

Marker on G7	Position on the peach genome sequence v2.0	Predicted gene including the SNP/indel	S5392	S10215	S10216	5392)²_54	(5392)²_64	(5392)²_124	(5392)²_130	(5392)²_150	(5392)²_423	(5392)²_522	(5392)²_534	(5392)²_548	(5392)²_573	(5392)²_797	(5392)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
E	-	-	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.

In silico analysis

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

Table 2 Predicted genes observed in the 70.5-kbp genomic region comprised between 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 (*Glycine max*) /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 (*Arabidopsis thaliana*) /Q9LF92	Phloem protein 2-A15/ AT3G53000.1
Prupe.7G121100	Pp07:14436305..14437630	Putative homeobox-leucine zipper protein ATHB-51 (*Arabidopsis thaliana*) /Q9LZR0	Homeobox 51/ AT5G03790.1
Prupe.7G121200	Pp07:14438623..14440853	60S ribosomal protein L24 (*Prunus avium*)/ Q9FUL4	Ribosomal protein L24e family protein/ AT3G53020.1
Prupe.7G121300	Pp07:14441874..14445148	Protein kinase dsk1 (*Schizosaccharomyces pombe*)/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 (*Daucus carota*)/P20075	Embryonic cell protein 63/ AT2G36640.1
Prupe.7G121600	Pp07:14471118..14474276	Probable glycosyltransferase At5g03795 (*Arabidopsis thaliana*) /Q9FFN2	Exostosin family protein/ AT5G03795.1
Prupe.7G121700	Pp07:14480000..14482717	Pentatricopeptide repeat-containing protein At5g03800 (*Arabidopsis thaliana*) /Q9FFN1	Pentatricopeptide repeat (PPR) superfamily protein/ AT5G03800.1
Prupe.7G121800	Pp07:14483790..14494599	E3 ubiquitin-protein ligase UPL7 (*Arabidopsis thaliana*) /Q9SCQ2	Ubiquitin-protein ligase 7/ AT3G53090.2

Reads from the eglandular ‘S10215’, the reniform ‘S10216’, ‘Summergrand’ and ‘Pamirskij 5’, as well as the globose ‘Zephyr’ and ‘Malo Konare’ were aligned onto Peach v2.0.a1 derived from the reniform peach *cv*. Lovell (Plov2-2N) and compared. A total of two hundred and seventy-seven variants between ‘S10215’ and ‘S10216’and heterozygous in ‘Malo Konare’, were identified among which six SNPs and an indel in the ASPP899-ASPP901 region (Table S2). However, no relationship was observed between any of the variants and the trait, except for the indel, which clearly differentiated eglandular, reniform and globose accessions. For the other 276 variants, the eglandular ‘S10215’ had the same haplotype as the reniform ‘Summergrand’ and ‘Lovell’ (Plov2-2N), as well as the globose ‘Zephyr’. In contrast, the reniform ‘S10216’ was highly similar to ‘Pamirskij 5’ (reniform), except for an 11-kbp region upstream of the indel, for which ‘Pamirskij 5’ had the same haplotype as the above four other accessions (Table S2). The indel was located in *Prupe.7G121100*, a gene annotated as putative homeobox-leucine zipper protein ATHB-51 (Table 2). According to Gene Ontology, *Prupe.7G121100*, has a DNA-binding transcription factor activity and is involved in bract formation and leaf morphogenesis. Based on these findings, 25 primers (Table S3) were developed from consensus regions between ‘S10215’ and ‘S10216’ in order to sequence the interval encompassing *Prupe.7G121100*, as well as the 100-bp gap which remained in the peach genome sequence reference[44] (Peach v2.0.a1) immediately upstream of the CG (position Pp07:14436205..14436304). The sequencing of the gap region resulted in sequences 9-fold longer than expected (905 bp and 903 bp for S10215 and S10216 respectively), therefore impacting coordinates downstream (Fig. S1). Sequence comparison allowed identifying a 590-bp insertion in the last coding DNA sequence (CDS) of *Prupe.7G121100*, in the eglandular ‘S10215’, as well as two additional polymorphisms due to differences in the number of CT repeats in two SSRs present in the gap region (Fig. S1). The 590-bp insertion was located between positions Pp07:14437331 and Pp07:14437332, disrupting the initial reading frame (Fig. S1). BLASTN search against NCBI database allowed finding a high similar hit (98% of identity) with an insertion fragment of 588

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

Analysis with FGENESH

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

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

Expression analysis of *Prupe.7G121100*

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

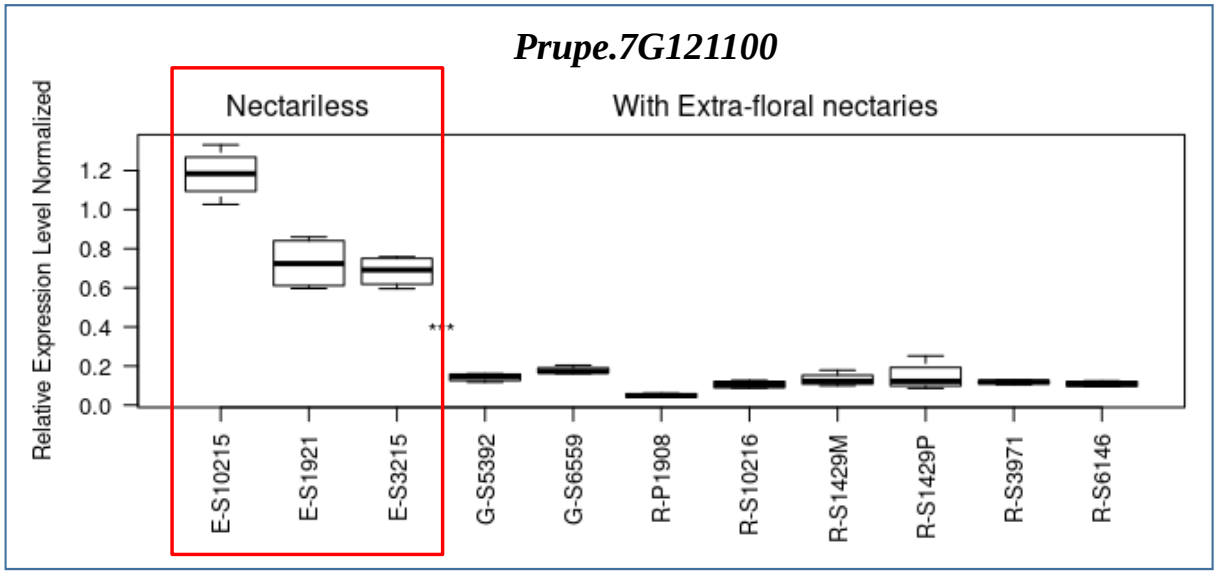

Fig. 4 Relative expression of *Prupe.7G121100* in ten accessions contrasting for EFNs. Expressions were normalized with the constitutive genes *PpTEF2* and *PpRPL13*. Eglandular, globose and reniform individuals are denoted by the letter E, G or R, before the accession number, respectively. The three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

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

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

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

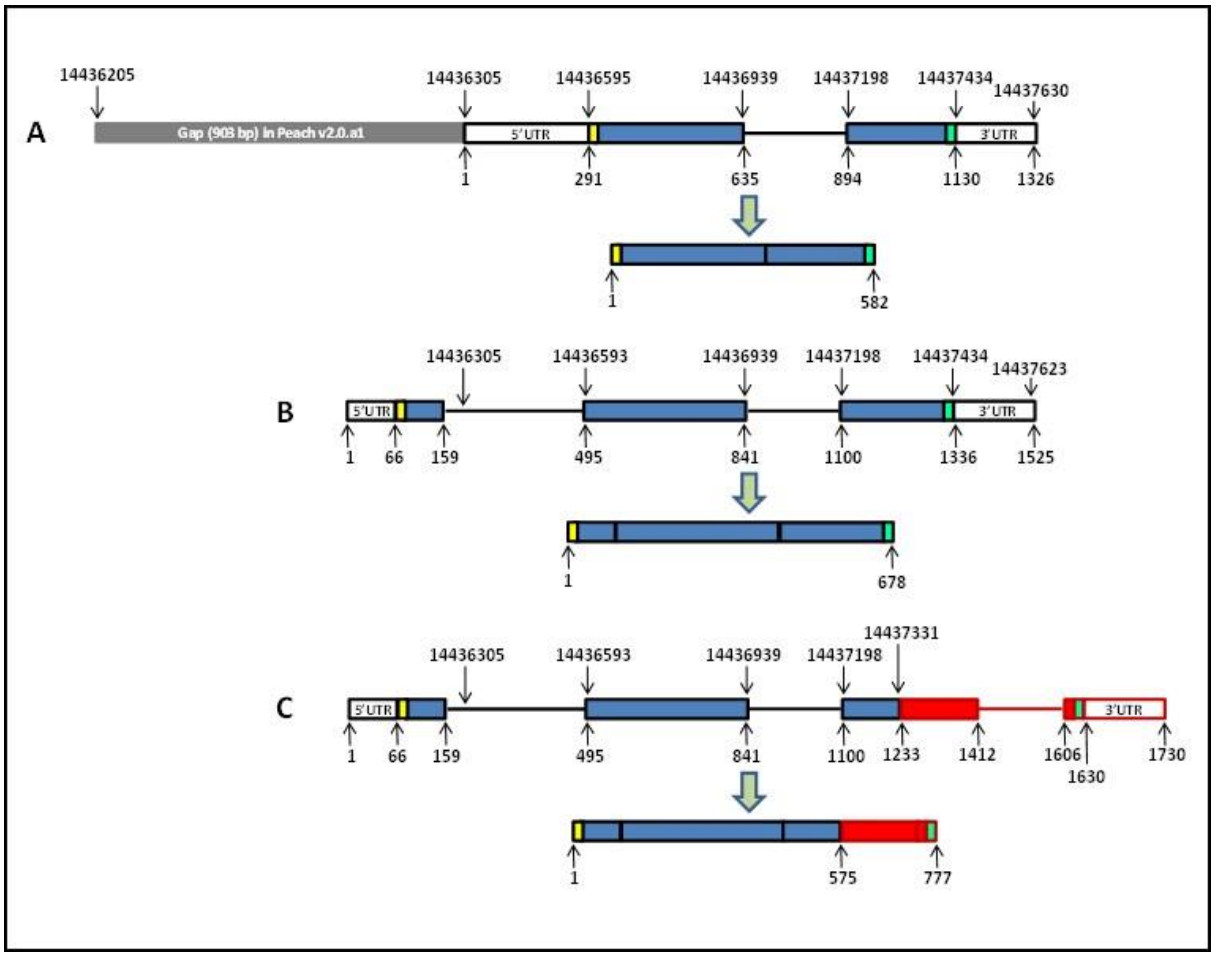

Fig. 3 Diagrams of *PpLMI1*. Spliced transcripts are displayed below their respective primary transcripts (**A**) Truncated *PpLMI1* (Prupe 7.G121100) as annotated in Peach v2.0.a1. CDSs are shown as blue rectangles, 5' and 3'-UTR as white rectangles, introns as black lines. The start and stop codons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

Genotyping with the ASPP900 marker

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

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

Discussion

The aim of our study was to identify the genomic factor responsible for the presence/absence of EFNs in peach, a Mendelian trait previously mapped on chromosome 7 (ref. 14, 15), but little studied so far. The fine-mapping approach allowed delimiting the trait to an interval of 10.7 kbp between positions Pp07:14,428,469 and Pp07:14,439,211. Micheletti et al.[18], using the ISPC 9K SNP peach array[19] and a collection comprising 750 reniform and 190 globose accessions, identified a single SNP, SNP_IGA_776161 (position Pp07:14,469,094) as associated to the leaf-gland type. This association was not fully congruent with our observations as 'S10215' (eglandular), 'Zephyr' (globose), 'Summergrand', 'Rubira' and the peach genome reference derived from 'PLov2-2N' (reniform) had the same allele combination (C/C) for this SNP, whereas 'S10216' and 'Pamirskij 5', both reniform were T/T. Nevertheless, taking into account the limited number of SNPs on the array corresponding to the trait region, as well as possible misclassification of some individuals of the collection, the results of the two studies were convergent. Coupling the results of the fine-mapping approach with the comparison of the genomic sequences of accessions contrasting for the trait, then allowed to clearly identifying a single candidate gene, *Prupe.7G121100*, among the three genes included in the above interval, and more broadly, among the 12 genes comprised in the longer 70.5-kbp genomic region encompassing the latter. Indeed, *Prupe.7G121100* has two variants: the regular one, associated with the presence of EFNs and homozygous in reniform individuals, and a second one including a 590-bp insertion homozygous in eglandular individuals. This insertion was identified as a MITE-like transposable element of the hAT superfamily, termed as *mMoshan*[20]. Miniature inverted-repeat transposable elements (MITEs) are non-autonomous class II transposable elements. They are considered a major driving force for generating allelic diversity in plant genomes[21]. MITES account for 3.89% of the peach genome[22] with 0.16% for the 491 *Moshan* elements identified. *Moshan* elements are unique to Rosaceae and the *mMoshan* class is predominant with 432 elements[20]. Interestingly, two of these *mMoshan* elements generated no obvious target site duplication, as the element inserted in *Prupe.7G121100*, suggesting that these three elements

were atypical. Wang et al. [20] identified 29 *mMoshan* which were inserted in genes, among which 14 in exons. The 29 genes were distributed over all the chromosomes but none in chromosome 7. The *mMoshan* in *Prupe.7G121100* was not detected probably because Peach genome v2.0.a1 was derived from the reniform double haploid 'Lovell' Plov2-2N and therefore does not include the inserted element. This author observed that genes including *mMoshan* elements showed relatively lower expression levels compared with genes lacking these elements and this was consistent with previous studies on MITES[23]. However, this was not the case in our study since *Prupe.7G121100* demonstrated enhanced expression in eglandular individuals compared to that in globose and reniform ones, which were quite similar. *mMoshan* elements contain several *cis*-regulatory elements such as MYB and WRKY binding sites in the first third of the sequence, which could be involved in upregulation of the transcription[20]. However, when we take into account the minor differences in gene expression observed between reniform and globose individuals, and the similarity of the phenotypes of their leaf-margins, as well as the incomplete dominance of the trait, this seems not correlated. It would be therefore interesting to investigate possible functional differences between the two alleles This point needs a dedicated approach. *Prupe.7G121100* was annotated as putative homeobox-leucine zipper protein ATHB-51, a member of the class I (HD-Zip I) superfamily of transcription factors. (HD-Zip) proteins are unique to plants. They include the peculiar combination of a DNA-binding homeodomain (HD) and an adjacent Leucine zipper (Zip) motif, which mediates protein-dimer formation[24]. Saddic et al. [25] identified ATHB-51 as a meristem identity regulator and named it LATE MERISTEM IDENTITY (*LMI1*) based on its regulation functions; accordingly, we will further refer to *Prupe.7G121100* as *PpLMI1*. These authors showed that *LMI1* was a direct target of LEAFY (*LFY*), a central meristem identity regulator in *Arabidopsis thaliana* as well as a direct upstream activator of a second meristem identity regulator, the MADS-box transcription factor CAULIFLOWER (*CAL*). *LMI1* acts together with *LFY* to induce *CAL* expression, the interaction between these three genes corresponding to a feed-forward loop transcriptional network motif[26]. *LMI1* thus belongs to the complex of genes including others transcription factors, such as APETALA1 (*AP1*), involved in the meristem identity switch leading to flower formation[25,27]. Interestingly, the *mMoshan* transposable element identified on chromosome 5, with the highest

percentage of identity with that inserted in *PpLMI1*, was in the promoter region of an homolog of AGL8 (agamous-like 8) transcription factor, a MADS-box negatively regulated by APETALA1, suggesting a possible involvement of APETALA1 in the regulation of *PpLMI1*. However, *LMI1* has also additional *LFY*-independent roles in leaf morphogenesis and bract formation[25]. For instance, *LMI1* regulates leaf growth in *Arabidopsis thaliana*[28] as well as organ proportions such as stipules size, via an endoreduplication-dependent trade-off, that limits tissue size and cell proliferation, through the activation of the mitosis blocker *WEE1*[29]. Moreover, modifications of *GhLMI1-D1b*, one of its homologues[30] were found to be responsible for the major leaf shapes in Upland cotton (*Gossypium.hirsutum*). This designates *LMI1-LIKE* genes (along with the *KNOXI* genes) as evolutionary hotspots that have been recruited in angiosperms to modify leaf shape[31]. In this way, Chang et al. [32] demonstrated that *LMI1-like* and *KNOX1* genes coordinately control leaf development in dicotyledons and that different expression patterns of these two genes correspond to the formation of different leaf marginal structures. The same way, loss of function of *CrLMI1*, a likely ortholog of *LMI1* was reported to decrease leaf serration in *Capsella rubella*[33]. This is in agreement with the results of our study, in which increased leaf-margin serration was found strictly associated with the absence of EFNs, concurrently with the higher expression of *PpLMI1*. This relationship between leaf serration and absence of EFNs was already reported in previous studies[13]. These findings thus contribute to confirm the possible involvement of *PpLMI1* in leaf-margin structures in peach and accordingly in the phenotype of EFNs. Likewise, in cucumber (*Cucumis sativus*), *mict*, a class I HD-Zip factor which sequence had 52% of identity with *LMI1*, regulates multicellular trichome development[34]. Regarding the EFNs, however, molecular genetic understanding of their formation is still underdeveloped. No study to date is available in tree species and only a few ones have been published in annual plants. For instance, Hu et al.[35] identified *GaNEC1*, a gene encoding a PB1 domain-containing protein, as positive regulator of nectary formation in cotton, which silencing led to a smaller size of foliar nectary phenotype. However, EFNs were located in the leaf midribs and their conformation was different than peach EFNs. Phenotypic diversity usually results from diversity in the genetic organization, regulation and/or expression of underlying developmental programs[4]. In the case of EFNs, such underlying programs have been poorly investigated. The gene *CRABS CLAW* (*CRC*), a YABBY transcription factor[36,37], appears to be an early-

functioning regulator of the development of both floral and extrafloral nectaries in core eudicots[38,39]. But while the location of floral nectaries may be determined by *CRC* along with several upstream MADS box floral homeotic genes and other unknown regulatory genes[38], the development of EFNs may involve the recruitment of different transcriptional control networks than those needed in floral nectaries[39]. This means that the program involved in EFN development may be closely associated with that of the EFN-bearing organ[4], the leaf, in the case of peach trees. As a result, the functional characteristics of *LMI1*, its involvement in leaf morphogenesis as well as in the meristem identity switch leading to flower formation, suggest that this transcription factor might have a pivotal role in the regulation of different characteristics of the leaf. This makes *PpLMI1* a most likely candidate for the presence/absence of EFNs in peach. Therefore, a plausible hypothesis is that functional modification of *PpLMI1* associated with the insertion of the HD-Zip I element might trigger endoreduplication. This would result in changes in the cell-wall composition of the lamina as well as in leaf margins, through a developmental program involving target-genes of *PpLMI1*, leading notably to serrated leaves and the absence of EFN. In addition, changes in cell-wall composition of the leaf-blade surface could thus make easier the development of fungi, such as *Podosphaera pannosa*, on the leaves. Modification of the cell walls at regions targeted by pathogen attack is a common response to infection and the inability to do so, or the presence of weakened cell walls, might explain, at least in part, the susceptibility to pathogens[40]. As a result, these changes, along with the absence of the positive effects associated with the presence of domitia-inhabiting mutualist mites, and fungivore mites attracted by EFN nectar, might be responsible for enhanced susceptibility to PPM in eglandular individuals, as compared to those with EFNs. Further studies need however to be undertaken in order to assess our hypothesis.

Conclusion

In this study, we were interested in identifying the genetic factor responsible for the presence/absence of EFNs in peach. In our knowledge, this is the first time that a molecular genetic approach has been undertaken to clarify the genetic basis of this Mendelian trait, in peach and, more broadly in *Rosaceae* perennial crops. Based on our results, *PpMLI1* appears the most likely candidate gene for this character. A comprehensive study of the genomic region including *PpMLI1* study did not bring to light another

alternative candidate. In addition, its characteristics, regulation functions as a meristem identity regulator as well as its role in leaf morphogenesis, make it highly plausible its involvement in the control of the presence/absence of EFNs as well as its association, at some extent, with the variation of the susceptibility to PPM, in link with cell-wall changes. However, this has to be further validated functionally. Virus induced gene silencing (VIGS) method could be considered as a relevant approach, as genetic transformation in peach is currently an obstacle. In addition, a broader study including the expression of *PpMLI1* in the meristem as well as that of genes interacting with *PpMLI1* or target genes such as *WEE1*, may be undertaken to elucidate the molecular interactions underlying this interesting trait. This study is thus a first step. Nevertheless, in the short term and from a breeder perspective, ASPP900 marker already allows differentiating the different phenotypes at the seedling level, and could then be used in peach breeding programs.

Material and methods

Plant material

The initial mapping population included 212 individuals derived from the self-pollination of 'Malo Konare' (clone S5392). 'Malo Konare' is a canning peach cultivar developed in 1984 at the Fruit-growing Institute in Plovdiv (Bulgaria). It originated from the cross 'Stoika' × 'New Jersey Cling 97', has globose leaf-glands and shows strong resistance to powdery mildew. 'Stoika' was for its part derived from 'House Kling' and 'Ferganskyi Zheltyi' (1973), a clone of *Prunus ferganensis*. The population was further extended to 833 individuals for the fine mapping of the leaf-gland region and identifying recombinants. This population will be referred to as 5392^{2}. In addition, 271 individuals were used to validate phenotype/genotype association in different genetic backgrounds: at first, 149 accessions with contrasting leaf-gland phenotypes (Table S1), including 143 peach cultivars from various origins, two accessions of wild species close to peach, *Prunus davidiana* (Carr.) and *Prunus kansuensis* (Koehne), one accession of *Prunus ferganensis* (Kost. & Rjab.), two double haploids and a possible triploid; then a sample of 122 individuals from a complex breeding population, referred to as $BC2^{41}$. The latter was derived from two successive crosses (F_1 and back-cross) including *Prunus davidiana* clone P1908 and peach cv. 'Summergrand', followed by a final cross derived from a mixture of pollen of the back-cross

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

DNA isolation

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

Leaf-gland phenotyping

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

Next Generation Sequencing of accessions

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

Library preparation kit from Illumina (Illumina Inc. San Diego CA, USA) following recommendations provided by the supplier. 125-bp paired-end sequencing was performed utilizing the Illumina HiSeq 2500 sequencing platform and the sequence by synthesis (SBS) technique. Base calling was performed by the Real Time Analysis (RTA) software. Raw Illumina paired-end reads were subsequently trimmed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Potential contaminants were investigated using FastQ Screen software (Babraham Institute) and Bowtie2 aligner (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Resulting reads were aligned onto Peach v2.0.a1 using BWA-MEM (v0.7.12-r1039) and the Ppersica_298_v2.0.fa version. BAM files (*.sorted.bam and *.sorted.bam.bai) were generated in order to visualize sequences under the Integrative Genomics Viewer (IGV) tool[42].

Marker development and genotyping

'Malo Konare' has been genotyped earlier in the frame of the European project Fruitbreedomics[18], using the IPSC peach 9K SNP array v1[19]. Based on the available SNP dataset, a first set of heterozygous SNPs was selected to develop the genetic map of linkage group 7 of 'Malo Konare' and insure a sufficient coverage of the group. Genotyping was done using the PCR-based KASP™ (Kompetitive Allele Specific PCR) method from LGC Biosearch technologies (https://www.biosearchtech.com/). Primer-triplets (two competitive allele-specific forward primers and one common reverse primer for each marker) were developed from the 60-bp genomic sequence available on either side of the SNPs (https://www.rosaceae.org/species/rosaceae_family_genera/IRSC_SNP_array), using Primer3[43] under the following primer-picking conditions: optimal size of the amplicons 75 bp (min 62 bp, max 85 bp), Tm 65°C (min 55°C, max 72°C), primer size 25 bp (min 20 bp, max 32 bp), max self-complementarity 7, max 3' self-complementarity 3, left primer end 61 bp. Primers triplets were compared with Peach v2.0.a1[44], using the Basic Local Alignment Search Tool (BLAST®) at the Genome Database for Rosaceae (GDR: https://www.rosaceae.org/blast/). Those aligning to single positions were selected for genotyping the starting mapping population (Table S4). In a second step, additional SNP markers focused on the interval encompassing the *E* locus were developed in order to identify recombinant individuals. This was done using Next-Generation Sequencing (NGS) data derived from 'Malo Konare'.

BAM files were aligned onto Peach v2.0.a1 and visualized with IGV[42]. The region containing *E* was examined for SNP/indel discovery and reads including heterozygous SNP/indels compatible with the KASP™ method were retrieved. Primer-triplets were then developed as above (Table S4). Mix preparation and PCR reactions were performed using the KASP™ genotyping chemistry and conditions.

Genetic map of linkage group 7

In a first stage, linkage group 7 (G7) of 'Malo Konare' was constructed using the mapping dataset derived from the SNP-set selected from Micheletti et al. [18]. Genotypic data were coded as F_2-progeny type according to the JoinMap coding system. The leaf-gland trait was similarly coded as a co-dominant Mendelian trait. Linkage analyses were performed using JoinMap 4.1[45]. The recombination fraction value was set at 0.4 and grouping was performed using the independence logarithm of odds (LOD) calculation function and a minimum LOD score threshold of 3. The Kosambi mapping function[46] was used to translate recombination frequencies into genetic distances. Linkage group 7 was established using regression mapping procedure with three rounds per sample. In a second stage, SNP markers developed for the high-resolution mapping in the interval including the *E*-locus region were added to the genotypic data file and mapped similarly.

High-resolution mapping of the *E* locus

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

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

The genomic region delimited by the SNP-pairs, which allowed identifying the most informative recombinants, was analyzed for variants. This was done by aligning the sorted.bam files of 'Malo Konare', 'S10215' and 'S10216' onto Peach v2.0.a1 (Ppersica_298_v2.0.fa version), under IGV[42], and by comparing them. Differences observed were then compared with sorted.bam files of 'Zephyr', 'Summergrand', 'Pamirskij 5' and 'Rubira' in order to check consistency of differences regarding leaf-

gland phenotype, in different genetic backgrounds. The genomic region defined above was examined for the presence of predicted genes, using JBrowse on the Genome Database for Rosaceae (https://www.rosaceae.org/jbrowse/). Positions of the observed differences were compared with those of the genes and their sub-features, then, genomic sequences of the candidate genes (CGs), associated transcripts and predicted protein sequences, homologies and gene functions were downloaded (https://www.rosaceae.org/node/4017147). NGS reads corresponding to the position of the selected CG-variants were retrieved for 'S10215' and 'S10216' using IGV[42], imported into CLC Main Workbench version 12 (QIAGEN, Aarhus, Denmark), assembled *de novo* and compared using MUSCLE[47]. In addition, as a 100-bp gap remained in Peach genome v2.0.a1 in the region immediately upstream of the most likely CG, 25 primers were developed (Table S3) and used for Sanger sequencing of the gap region and the target CG (Genewiz, South Plainfield, NJ, USA). Assembled sequences were then compared and differences between 'S10215' and 'S10216' identified. Sequences of the selected CG and the gap region, were finally analyzed for comparison and possible changes in the coding sequences, as well as changes in the resulting protein, using FGENESH gene-prediction program[48] with different dicot plant species as model (http://www.softberry.com)).

Gene expression analysis

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

S3) were designed using Primer3[43] and GenScript® Tool (Table S5). One microgram of total RNA per sample was then subjected to cDNA synthesis using the AffinityScript RT kit (Agilent) according to the manufacturer's instructions. A SYBR green real-time PCR assay was thereby carried out in a final volume of 15 µl of a reaction mixture containing 7 µl of 2x Brilliant III SYBR® Green qPCR Master mix (Agilent), 0.5 µM of each primer and 100 ng of cDNA template. Reaction mixtures without cDNA were used as negative controls. Amplification reactions were run in a 96 well plate on a Stratagene Mx3005P (Agilent) under the following conditions: 95°C for 30 s, followed by 40 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 30 s and extension at 72°C for 15 s. Reactions were performed using four biological and three technical replicates for each sample. Amplification values were then normalized using two genes as constitutive controls, as recommended by Bustin et al.[48]: *PpTEF2* (translation elongation factor 2) and *PpRPL13* (60S ribosomal protein L13), both having previously been tested and selected for their stability. Two-way analysis of variance (ANOVA) was used to assess the independent effect of the presence of EFNs and that of the insertion on the expression of *Prupe.7G121100.* Tests were performed using a script of 'RqPCRAnalysis' R-package[49] customized to generate box-plots with R studio[50]. Significance threshold was set to $p < 0.01$.

3' RACE PCR

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

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

Development of the diagnostic marker ASPP900

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

Acknowledgements

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

UGAFL) for their technical contribution to tree management of the peach collection and the population 5392[2]. Finally we thank Jean-Luc Gallois (UGAFL) for his help and suggestions as well as Jean-Luc Poëssel (UGAFL) for proofreading the manuscript.

Conflict of interests

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

Data Availability

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

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

References

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

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

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

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

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

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

7 Weber, M.G., Porturas, L.D. & Taylor, S.A. Foliar nectar enhances plant–mite mutualisms: the effect of leaf sugar on the control of powdery mildew by domatia-inhabiting mites. *Ann. Bot.* **118**(3), 459-466 (2016).

8 Scorza, R. & Sherman, W.B. Peaches. pp 325-440 In: Janick J, Moore JN (eds.) Fruit breeding, Volume 1: Tree and tropical fruits. John Wiley & Sons, NY (1996).

9 Watking, W. & Brown A.G. Genetic response to selection in cultivated plants: Gene frequencies in varieties of *Prunus persica*. *Proc. Soc. Lond. B. Biol Sci.* **145**, 337-347 (1956).

10 Saunier, R. Contribution to the study of relationships between certain characteristics of simple genetic determination in the peach tree and susceptibility of peach cultivars to oïdium, *Sphaerotheca pannosa* (Wallr.). *Lev. Ann. Amelior. Plant* **441**, 235-243 (1973).

11 Weinhold, A.R. The orchard development of peach powdery mildew. *Phytopathology* **51**, 478-481 (1961).

12 Pascal, T., Pfeiffer, F. & Kervella, J. Powdery mildew in the peach cultivar Pamirkij 5 is genetically linked to the *Gr* gene for leaf color. *Hort. Sci*. **45**(1), 150-152 (2010).

13 Connors, C.H. Inheritance of foliar glands of the peach. *Proc. Am. Soc. Hortic. Sci.* **18**, 21-27 (1921).

14 Dettori M.T., Quarta, R. & Verde, I. A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. *Genome* **44**, 783–90 (2001).

15 Dirlewanger, E. et al. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. USA* **101**, 9891-9896 (2004).

16 Verde, I., Quarta, R., Cedrola, C. & Dettori, M.T. QTL analysis of agronomic traits in a BC1 peach population. *Acta Hortic.* **592**, 291-297 (2002).

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.013680319
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:
593 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 Thorvaldsdóttir, 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).

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

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

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

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