Genome-wide association study dissects the genetic architecture of progesterone content in Persian walnut leaves (Juglans regia L.)

Anthony Bernard, Elisabeth Dirlewanger, Marine Delmas, Delphine Racofier, Marie-Laure Greil, Anne-Lyse Lainé, Chantal Porte, Dominique Gennetay, Matthieu Keller, Philippe Chemineau

BMC Genomics

RESEARCH

Open Access

Genome-wide association study dissects the genetic architecture of progesterone content in Persian walnut leaves (*Juglans regia* L.)

Anthony Bernard[1*], Elisabeth Dirlewanger[1], Marine Delmas[2], Delphine Racofier[3], Marie-Laure Greil[3], Anne-Lyse Lainé[4], Chantal Porte[4], Dominique Gennetay[4], Matthieu Keller[4] and Philippe Chemineau[4]

Abstract

Progesterone (P4) is an endogenous sex steroid hormone involved in the ovulatory cycle and pregnancy of animal species. In sheep and goats, P4 analogues are used to induce synchronized ovulations and oestrus behavior of the females. In humans, P4 from chemical synthesis is used to treat peri-menopausal disorders. However, such molecules are released into aquatic environment and can be a source of pollution, are prohibited in organic farms and go against the trend of "naturality" in animal production as well as in human health. A natural alternative may consist in the extraction and use of P4 in plants. Mammalian hormones were discovered in an increasing number of plant species, including walnut leaves that contain high levels of P4. We compared the content of P4 in leaves of 170 accessions of *Juglans regia* from the walnut germplasm collection of INRAE *Prunus-Juglans* Biological Resources Center previously genotyped using the Axiom™ *J. regia* 700 K SNP array. We conducted a genome-wide association study (GWAS) using multi-locus models. When collected in October, P4 content goes from 34,1 to 287,5 mg/kg dry weight of leaves. The two laciniate accessions have the largest P4 content. We identified seven significant marker-trait associations on chromosomes 1, 2, 3, 6, 7, 15 and 16, and a candidate gene involved in the metabolism of sterols, precursors of plant steroid hormones. Our results raise the huge variability of P4 content within *J. regia* and propose a candidate gene which may have a role in the control of this variability, opening the way to a potential use of walnut P4 by the pharmaceutical industry towards more natural source of chemical compounds.

Keywords Walnut, *Juglans regia* L., GWAS, Progesterone

Background

In animal production, hormonal control of the reproductive cycle in goats and sheep is crucial for breeders to synchronize or induce oestrus behavior, enabling artificial insemination and genetic progress [1]. Traditionally, this control involves about 10 days of progesterone (P4) or synthetic progestogens administered via implant, oral methods, or, most commonly, intravaginal impregnated sponges. However, these treatments are increasingly rejected by farmers due to environmental concerns [2]. Furthermore, they are prohibited

*Correspondence:
Anthony Bernard
anthony.bernard@inrae.fr
[1] INRAE, Univ. Bordeaux, UMR BFP, 33882 Villenave d'Ornon, France
[2] INRAE, Unité Expérimentale Arboricole, Domaine de La Tour de Rance, 47320 Bourran, France
[3] INRAE, Unité Expérimentale Arboricole, Domaine Des Jarres, 33210 Toulenne, France
[4] INRAE, CNRS, Université de Tours, Physiologie de La Reproduction Et Des Comportements, 37380 Nouzilly, France

in organic farming, which limits the potential for genetic improvement in these systems [1].

In humans, P4 is widely used during the menstrual cycle to alleviate painful menstruation and manage menopausal symptoms such as hot flushes and night sweats, while reducing breast cancer risk during menopausal hormone therapy [3, 4]. To date, all P4 products are derived from chemical synthesis. While P4 synthetized from diosgenin, a phytosteroid found in plants such as yam (genus *Dioscorea*), has been mentioned as a natural alternative [5], there is growing interest in identifying more sustainable and plant-based sources of P4 to address public opinion and environmental concerns. Thus, we explored the potential of using plants as natural sources of P4 for veterinary and human applications.

Progesterone has been identified in a wide range of plants [6, 7]. For instance, in rice (*Oryza sativa*) shoots, approximately 1.5 µg of P4 per kg of fresh tissue was quantified using GC–MS [8]. Lower concentrations were reported in apple (*Malus domestica*) flesh (0.15 µg/kg), onion (*Allium cepa*) bulbs (0.068 µg/kg), and mung bean (*Vigna radiata*) etiolated seedlings (0.021 µg/kg). The same authors also revealed significant variation in P4 content among plant organs. Ubiquitous at low levels, progesterone is thought to play a role in growth processes and responses to biotic and abiotic stress [9]. Notably, walnut trees (*Juglans regia* L.) stand out due to their exceptionally high P4 levels. Specifically, approximately 1 µg of P4 per gram of dry walnut leaves was observed, a concentration that is orders of magnitude higher than those found in other plant species, and comparable to the effective dose used in goats and sheep [10]. Additionally, our previous research demonstrated that walnut leaves exhibit peak P4 content in October in the Northern Hemisphere, just before senescence [11].

In walnut, genome-wide association studies (GWAS) have successfully identified the genetic bases of various quantitative traits, including flowering date [12] and fruit weight [13]. Given the high P4 content in walnut leaves, we aimed to investigate the genetic determinism of P4 content in walnut. To this end, we conducted a GWAS using multi-locus models on a diverse walnut germplasm collection maintained at the INRAE *Prunus-Juglans* Biological Resources Center, which comprises 170 accessions [14]. This study represents the first attempt to elucidate the genetic architecture of P4 content in walnut and highlights its potential as a sustainable, plant-based source of progesterone for both veterinary and human applications.

Methods

Plant material and progesterone quantification

The GWAS panel is made of 170 unique *J. regia* accessions (Table S1) belonging to the walnut germplasm collection of INRAE *Prunus-Juglans* Biological Resources Center [14] (https://doi.org/https://doi.org/10.17180/WN42-3J20), and previously used to study phenological traits [12] and fruit quality traits [13]. In addition, we included 7 accessions belonging to wild relative *Juglans* species (*J. mollis, J. nigra, and J. major)* for progesterone quantification (Table S1). All the accessions are located in the Fruit Experimental Unit of the INRAE of Bordeaux, at Toulenne and Bourran located 50 km and 120 km, respectively, south-west from Bordeaux, France (https://doi.org/https://doi.org/10.17180/9ST1-4J21).

For each *J. regia* accession, 20 leaves (approximately 100 leaflets) were collected either at ground level or from a platform 3–5 m above ground between October 6th and October 7th 2021 from adult trees managed for fruit production. Leaves were air-dried at < 30 °C using a semi-industrial dryer used for walnut fruit in the Fruit Experimental Station. Drying generally took < 48 h. Dry leaves were then stored at room temperature in darkness until needed. As described in detail in [11], each walnut leaf sample used for the P4 assay comprised about 10 leaflets. Briefly the different steps were: gentle grinding of the leaves; extracting a 25 mg sample with 2 mL of ethanol; vortexing then agitating at room temperature overnight; centrifuging; evaporating; redissolving with Tris-BSA buffer and storing at 4 °C until required for enzyme immunoassay (EIA). Under these conditions, calculated from two successive overnight extractions of the same samples, the maximum extraction yield was 90%. The final values were corrected for this bias.

Concentrations of progesterone were measured by EIA for blood plasma as described by [15], and described in detail in [11] for walnut leaves, using 25 µL of the final Tris-BSA solution above. The method included reference samples of known values at regular intervals during the assay, which allowed us to estimate the CV and were used as quality controls. The sensitivity was 0.25 mg/kg, and the interassay CV was < 5.8%. The specificity of this P4 antibody was tested against 27 P4-related molecules. Three compounds (5α-dihydroprogesterone, isopregnanonole, and 5β-dihydroprogesterone) cross-reacted at > 2% with our antibody (22, 15 and 14%, respectively; see [11].

Outlier phenotypes and extreme phenotypes were determined using R package "rstatix" [16] as follows: phenotypes above Q3 + (1.5xIQR) or below Q1 − (1.5xIQR) were considered as outliers, whereas phenotypes above Q3 + (3xIQR) or below Q1 − (3xIQR) were considered as extreme points, where Q1, Q3 and IQR are the first

quantile, the third quantile and the interquartile range (Q3 – Q1), respectively.

In 2023, new sampling was performed the October 2nd in order to validate the consistency of the P4 content in six contrasted accessions (3 with low and 3 with high P4 contents in 2021) and to test for a "year" effect in the Fruit Experimental Unit of the INRAE of Bordeaux, at Toulenne. These six genotypes were sampled twice depending on the position of the leaves in the tree (top and bottom) to test for a "position" effect in the Fruit Experimental Unit of the INRAE of Bordeaux, at Toulenne. They were also sampled at two locations (Toulenne and Bourran, World Geodetic System 1984 coordinates 44.575575, −0.283197 and 44.333304, 0.412251, respectively) to test for a "site" effect as both locations have clones of the selected accessions from different years of plantation (Toulenne between 1978 and 2002, Bourran between 2010 and 2020). R package "rstatix" [16] was used to perform comparison of means.

In 2023 as well, two new laciniate walnut accessions were collected: one from the walnut germplasm collection of INRAE *Prunus-Juglans* Biological Resources Center not included in the core collection of 170 accessions, called 'Lacinié', and one from the "Jardin Botanique de l'Université de Strasbourg" (Botanical Garden of University of Strasbourg), called 'Lacinié Strasbourg' (Table S1).

Genome-wide association analysis, LD blocks and search of annotations

Data management and visualization were performed using R package "tidyverse" [17]. P4 quantification was performed two times (technical replicate) for each accession and the mean was used as phenotypic data. SNP genotyping, quality control of these SNPs, population structure and kinship analyses were already performed in previous studies [12, 13]. The accessions were genotyped using the Axiom™ *J. regia* 700 K SNP array containing 609,658 SNPs uniformly distributed over the 16 J. *regia* chromosomes [18]. Filtering metrics were performed by ThermoFisher considering following thresholds: dish quality control greater than or equal to 0.82, and quality control call rate of 97%. For the quality control steps performed with "PLINK1.9″ software" [19], SNPs were filtered using stringent thresholds: SNP call rate (> 90%), minor allele frequency (MAF > 5%), and redundancy in the genome (SNP probes aligning in duplicated regions). Finally, 364,275 robust SNPs (59.8% of the total number of SNPs) were retained for the following genome-wide analysis.

GWAS was carried out using the R package "GAPIT version 3" [20, 21]. Three multi-locus models were tested: the Multi-Locus Mixed Model (MLMM) [22], the Fixed and random model Circulating Probability Unification method (FarmCPU) [23], and the BLINK model [24], using automatically implemented configurations in GAPIT. Familiar relatedness was accounted for using a kinship matrix estimated with the VanRaden algorithm implemented in GAPIT. In order to correct for population structure, the best number of principal components to include in our models was selected using the "model. selection" function implemented in GAPIT according to the Bayesian Information Criterion (BIC). Significant marker-trait associations were determined using 5% Bonferroni correction.

MLMM, FarmCPU, and BLINK address relatedness with distinct strengths and limitations. MLMM controls relatedness by incorporating a mixed model framework, reducing false positives, but its computational intensity can be limiting. FarmCPU improves computational efficiency by separating marker testing from variance component estimation steps using only a subset of significant markers, but it may struggle with correlated marker effects, leading to false negatives. BLINK enhances speed and power by simplifying random effects and exploiting linkage disequilibrium, but it may underperform when relatedness is strong. Together, these models offer complementary approaches for addressing the challenge of related germplasm populations.

Each genome position of the identified marker-trait associations was investigated to explore the extension of the surrounding linkage disequilibrium (LD) blocks using the "solid spine of LD" method implemented in HaploView v4.2 software [25]. We searched for candidate genes within the defined LD blocks using the walnut nuclear gene annotation and mapped into the new chromosome-scale reference 'Chandler' genome v2.0 [26] (https://treegenesdb.org).

Correlation of P4 content with other traits measured in the walnut core collection

The walnut core collection was previously phenotyped in 2018 for 8 traits related to phenology such as budbreak date, male flowering date and female flowering date [12], and for 25 traits related to fruit quality and nutritional composition such as kernel weight, nut length, shell thickness and polyunsaturated fatty acids content [13]. We investigated the correlation level of these traits with P4 content quantified in 2021. We phenotyped as well the global foliage diseases in June 2023 and August 2023 to investigate whether P4 could be related to any form of tolerance to foliage diseases for the 170 accessions, based on total brown spots. We used the following scale: 0: no brown spots on foliage, 1: < 5% of the foliage affected by brown spots, 2: between 5 and 25% of the foliage affected, 3: between 25 and 75% of the foliage affected, 4: > 75% of

leaves with the presence of brown spots, so diseases generalized throughout the tree. Pearson correlations were investigated using the R packages "corrplot" [27] and "PerformanceAnalytics" [28].

Results

P4 quantification and interannual content in outlier phenotypes

A large variability was observed in P4 content in 2021 within the INRAE *J. regia* germplasm collection. The values vary from 34.1 µg/g for the Chinese accession 'Lu Guang' to 287.5 µg/g for the accession 'Lacinié Suisse' (Table S1). Among the accessions, the two unique laciniate accessions have the largest P4 content, with 238.6 µg/g for 'Laciniata' and 287.5 µg/g for 'Lacinié Suisse'. These two accessions have laciniate leaflets (also called cutleaf walnut tree) and are mainly used for ornamental purpose (Figure S1). The two laciniate walnut trees are outlier phenotypes but not extreme outliers, so that they can bring useful biological information in our GWAS. Compared to *J. regia*, the individuals of wild relative *Juglans* species have tremendously less P4 (Table 1). In average, *J. mollis*, *J. nigra* and *J. major* have 0.6, 0.7 and 0.4 µg/g of P4 in their leaves, respectively, against 102.2 µg/g for the *J. regia* core collection.

As we selected contrasted accessions to validate the consistency of the P4 content between the years, both assumptions, normal distribution and homogeneity of variance, are not met ($p=6.63e\text{-}03$ for Shapiro-Wilk test and $p=8.49e\text{-}03$ for Levene test). Consequently, a non-parametric Wilcoxon rank sum test (with independent samples) was used and no significant difference was found between the P4 concentrations in 2021 and 2023 of the 6 contrasted accessions (Fig. 1).

No significant difference in P4 concentration between the top and the bottom of the trees of the 6 genotypes at Toulenne was detected using a Wilcoxon rank sum test. However, we found a tendency to obtain higher P4 values at the bottom of the tree, for the 3 genotypes having more P4 (Figure S2). When comparing the sites, no significant difference neither was found in P4 concentration between clones at Toulenne and clones at Bourran. However, we found a tendency to obtain higher P4 values at Toulenne, for the 3 genotypes having more P4 (Figure S3).

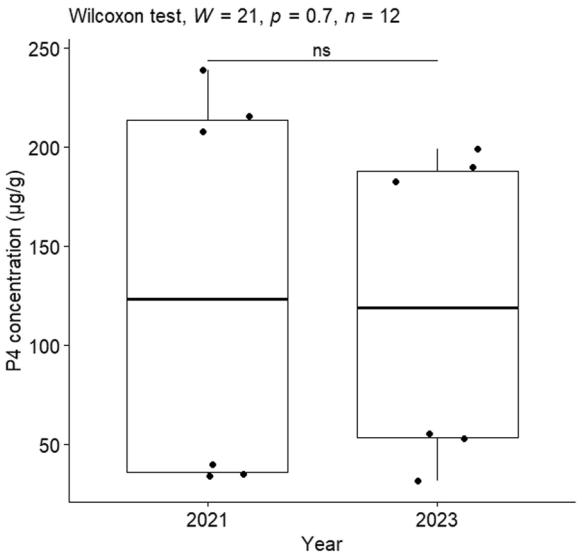

Fig. 1 Boxplots of the "year" effect (2021 vs. 2023) on the P4 concentration using six contrasted phenotypes (Wilcoxon rank sum test)

Regarding laciniate walnut accessions, the new samplings conducted in 2023 give contrasted results. For the two accessions already sampled in 2021 and used for GWAS, the P4 content is stable between years. We obtained 238,6 µg/g in 2021 and 189,6 µg/g (mean of "top" and "bottom" sampling) in 2023 for 'Laciniata', and 287.5 µg/g in 2021 and 201.4 µg/g in 2023 for 'Lacinié Suisse'. Both accessions, outliers in 2021, still gave high levels of P4 concentration in 2023. However, for the two new laciniate walnut accessions collected only in 2023, the P4 content is low: 42.6 µg/g for 'Lacinié' and 66.3 µg/g for 'Lacinié Strasbourg'.

Table 1 Progesterone content in *J. regia* and wild relatives

***Juglans* species**	**INRAE ID**	**Progesterone adjusted concentration (µg/g)**
J. mollis	MO 1	0,6
J. nigra	NG 209	1,1
J. nigra	NG 232–4	0,3
J. major	MJ 2–2	0,2
J. major	MJ 1	0,6
J. major	MJ 5–1	0,1
J. major	MJ 5–2	0,5
J. regia	Mean core collection	102,2

Cofounding factors for GWAS and marker-trait associations

Both population structure and kinship are common cofounding factors in GWAS [27]. In our *J. regia* germplasm collection, the accessions are structured in two main groups, accordingly to their origin: Eastern Europe and Asia or Western Europe and America [12]. Based on an analysis of variance, a significant variation of P4 content between these groups of structure was detected (Figure S4). However, in our analysis, we defined that the best number of principal components to include in the GWAS models for accounting for structure was zero using the BIC (Table S2), as previously found for phenological

traits [12] and fruit quality traits [13]. This means that the structure of our germplasm is mostly accounted for using the kinship matrix, supporting the best number of Principal Components (PCs) to include of zero. To remove any doubt, we considered both "PCA = 0" and "PCA = 2" in our GAPIT analysis.

For our trait, the models used are adapted since the quantile–quantile plots have a line shape with a tail (Figure S5). By using 1% Bonferroni threshold (solid green line), we found seven associations with the P4 content, depending on the GWAS model used and the number of PCs included. When considering zero PCs as structure covariable, the Manhattan plots indicate marker-trait associations on chromosomes 2 and 16 detected using both FarmCPU and BLINK models. The MLMM approach identified the region on chromosome 2, even though not significant (Fig. 2). We also identified an association on chromosome 1 using FarmCPU model only. When considering two PCs as structure covariable, the association on chromosome 2 is found as well using FarmCPU, but other associations are detected. The other associations are on chromosomes 7 and 15 using both FarmCPU and BLINK models, and on chromosomes 3 and 6 using FarmCPU model only (Fig. 2).

The association on chromosome 2 is detected three times (FarmCPU/PCA0, BLINK/PCA0 and FarmCPU/PCA2). The associations on chromosomes 7 and 15 are detected twice with PCA = 2 (FarmCPU and BLINK). The association on chromosome 16 is detected twice with PCA = 0 (FarmCPU and BLINK) (Table 2, Fig. 2). The SNP AX-170913249 on chromosome 2, position 6,219,027 bp, is the most significant when considering PCA0, and explains between 7.72 and 12.75% of the P4 content variance according to the GWAS model. The SNP AX-170971757 on chromosome 16, position 13,373,812 bp, is less significant but explains more variance, between 12.18 and 19.03% according to the GWAS model. For the three associations detected using FarmCPU only (on chromosomes 1, 3 and 6), the variance explained is between 3.14 and 6.66%. By comparing associations detected with both models (FarmCPU and BLINK) the variance explained is higher with the BLINK model.

The effects of alleles are high, between 17.03 to 28.76 µg/g in absolute value, depending on the association and the GWAS models (Table 2). The boxplots of alleles for each association are given in Fig. 3. They all have an additive effect, except the SNP AX-171113163 on chromosome 3 having an overdominance effect. For most associations, the two laciniate accessions have the same allele (chromosomes 1, 3, 6, 7, and 15), as we can observe the two dots outside the boxplots corresponding to 'Laciniata' (238.6 µg/g) and 'Lacinié Suisse' (287.5 µg/g) (Fig. 3). However, for the two associations on chromosomes 2 and 16, 'Lacinié Suisse' is homozygous and 'Laciniata' is heterozygous at these loci.

To go further, we ran the GWAS discarding the two laciniate accessions to check whether they bias the analyses. Results show consistency, with the association on chromosome 2 at 6,219,027 bp that remains the major one, identified for both FarmCPU and BLINK models, with PCA = 0. When discarding them, some differences are highlighted for minor loci, such as the one on chromosomes 15: the peak SNP on chromosome 15 was found at 15,524,979 bp in both models and PCA = 2 when the two laciniate accessions are included, whereas we found it at 15,506,389 bp in both models and PCA = 0 when discarding them.

LD blocks and candidate genes

Using the "solid spine of LD" method, we found differences in LD level around the marker-trait associations. For instance, the SNPs of the two associations on chromosomes 1 and 2 belong to a LD block of less than 1 kb and 3 kb, respectively, whereas the LD block corresponding to the SNP of the association on chromosome 15 is more than 100 kb (Table 3). Unfortunately, within the shorts LD blocks corresponding to the two associations on chromosomes 1 and 2, no gene were found. Even if the association on chromosome 2 seems to be highly reliable, we cannot propose any candidate gene associated. To remove any doubt, we searched for genes present on either side of the SNP in a 200 kb window (SNP AX-170913249 at 6,219,027 bp on chromosome 2, so a window between 6,119,027 and 6,319,027 bp). Within this 200 kb window, we identified nine genes but none appear to be related to hormone metabolism (Table S3). However, we identified a total of 15 genes included in the LD blocks of the seven associations (Table 3). Among these 15 genes, four genes overlap the physical position of the associated SNP and deserve attention. The SNP of the association on chromosome 6, even though it's not the most significant, is within the gene *Jr06_01700*, and its functional annotation according to SwissProt corresponds to a "*sterol 3-beta-glucosyltransferase UGT80B1*" identified in *Arabidopsis thaliana*.

To better understand the metabolism of steroids in *J. regia*, we looked for genes related to "steroid", used as keyword in the functional annotation of the genome. In total, 23 genes have the keyword "steroid" in their functional annotation (Table S4) but none of them are close (less than 1 Mb) to a GWAS peak.

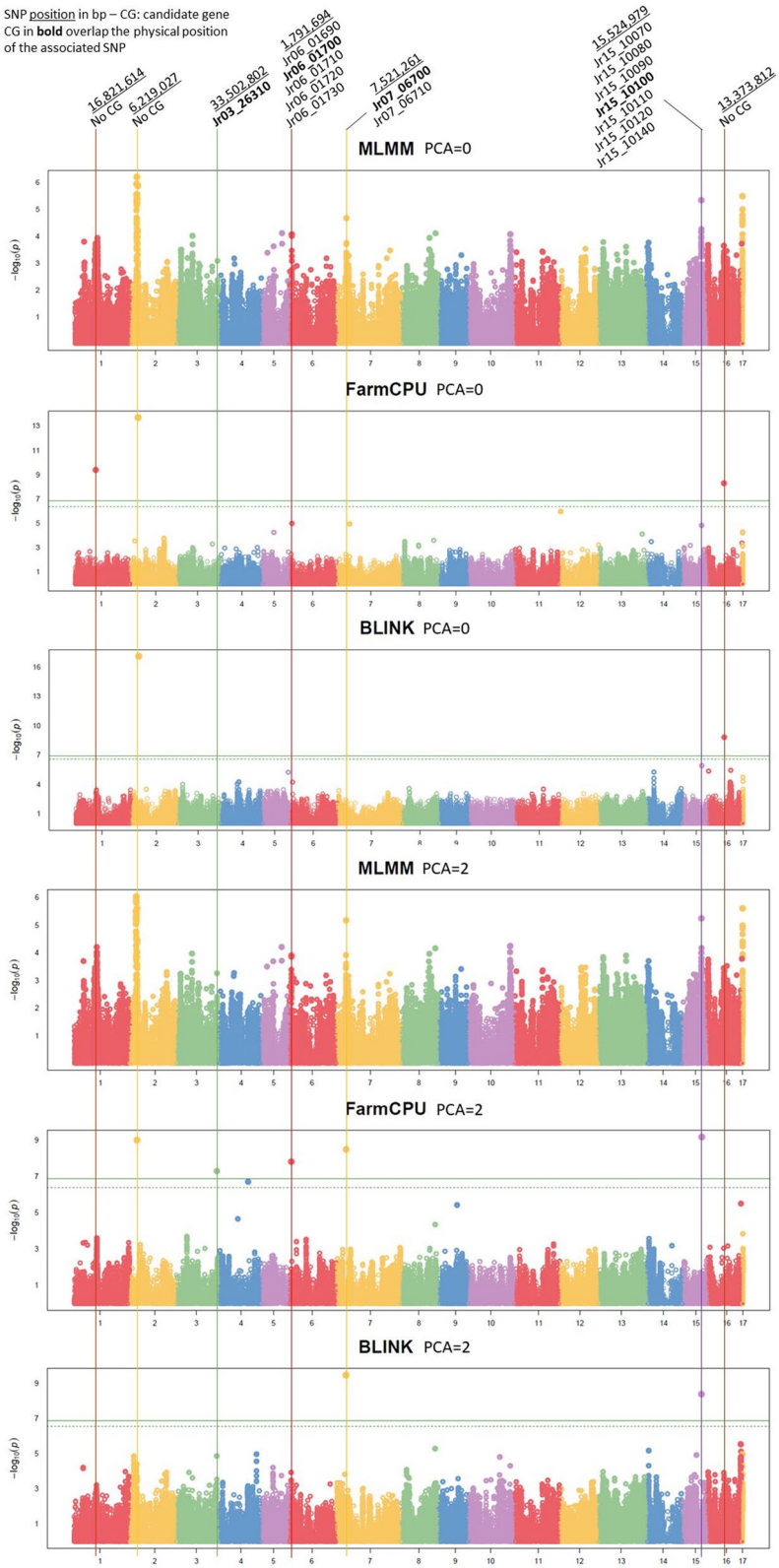

Fig. 2 Manhattan plots of the GWAS analysis conducted on the P4 concentration. From top to bottom: MLMM/PCA0, FarmCPU/PCA0, BLINK/PCA0, MLMM/PCA2, FarmCPU/PCA2, BLINK/PCA2. Horizontal solid green line and horizontal dotted green line correspond to 1% and 5% Bonferroni threshold, respectively, automatically implemented in GAPIT

Table 2 Summary of association mapping results related to P4 content

GWAS model/ Number of PCA	Chr [a]	SNP	Physical position (bp)	Significance [b]	Minor Allele Frequency	R^2 [c]	Alleles/Effect [d]
FarmCPU/PCA0	1	AX-170606550	16,821,614	4.17E-10	0.23	3.14	T/C (20.70)
	2	***AX-170913249***	***6,219,027***	2.08E-14	0.47	7.72	T/C (21.42)
	16	*AX-170971757*	*13,373,812*	5.14E-09	0.30	12.18	G/A (−17.31)
BLINK/PCA0	***2***	***AX-170913249***	***6,219,027***	7.85E-18	0.47	12.75	T/C (28.76)
	16	*AX-170971757*	*13,373,812*	1.58E-09	0.30	19.03	G/A (−20.92)
FarmCPU/PCA2	***2***	***AX-170913249***	***6,219,027***	1.01E-09	0.47	5.45	T/C (18.49)
	3	AX-171113163	33,502,802	5.15E-08	0.13	6.66	T/C (−20.61)
	6	AX-170677879	1,791,694	1.57E-08	0.38	3.19	C/A (−17.03)
	7	*AX-171133694*	*7,521,261*	3.29E-09	0.39	5.89	G/A (−18.03)
	15	*AX-171019542*	*15,524,979*	6.82E-10	0.47	5.22	G/A (−16.63)
BLINK/PCA2	*7*	*AX-171133694*	*7,521,261*	3.38E-10	0.39	17.42	G/A (−26.30)
	15	*AX-171019542*	*15,524,979*	4.10E-09	0.47	11.75	G/A (−20.97)

[a] Chr, abbreviation for Chromosome

[b] The significance value indicated is the p-value, which is significant if lower than the False Discovery Rate

[c] R^2 is the percentage explained variance

[d] The allelic effect is the difference in mean of P4 content between genotypes with one or other allele

The sign is with respect to the allele that is first mentioned

Associations in *italic* are detected using both FarmCPU and BLINK models for PCA = 0 or PCA = 2

Associations in *italic and bold* are detected for PCA = 0 and PCA = 2

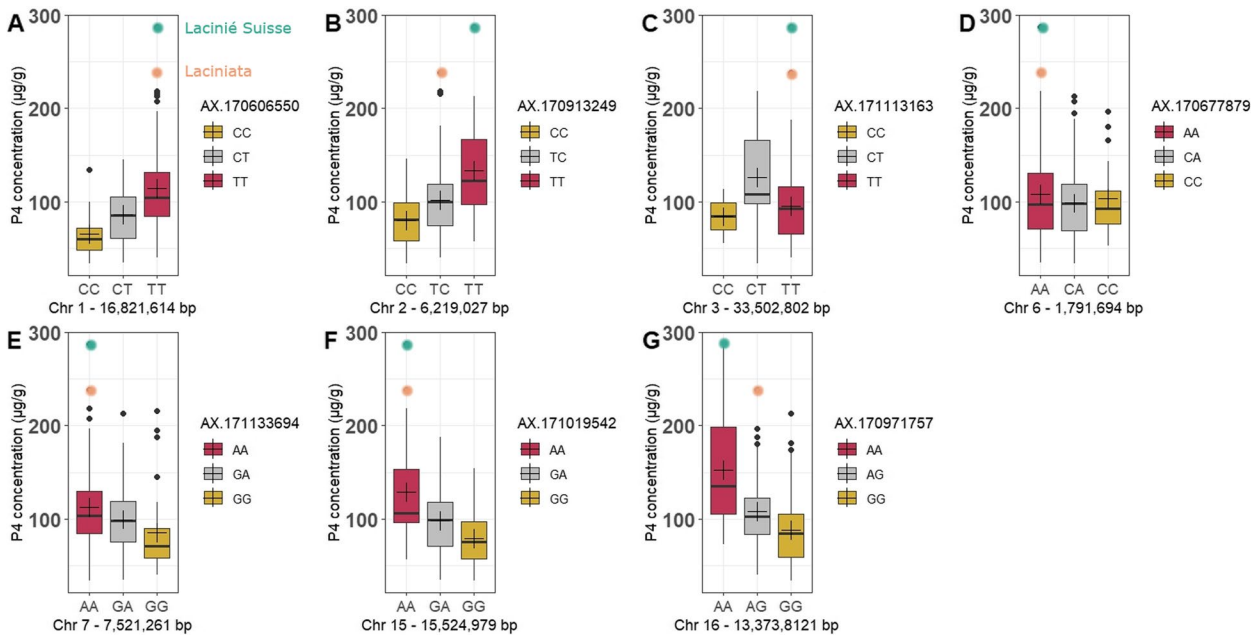

Fig. 3 Boxplots of the allelic effects of the seven SNPs associated with the P4 concentration

Correlation of P4 content with other traits measured in the walnut core collection

Using the phenotyping data obtained in 2018 for several traits related to phenology [12] and fruit quality traits [13], we investigated the correlation level of these traits with P4 content from 2021. Results indicate no correlation between P4 and any of the 33 traits (Figure S6). Moreover, the P4 content from 2021 is not correlated with global foliage diseases based on brown spots observed in June and August 2023 (Figure S7).

Table 3 LD blocks of marker-trait associations and candidate genes

Physical position in bp	LD block interval [a]	Gene ID	Gene interval [b]	Functional annotation (Swiss-Prot, top BLASTX hit)
Chr1—16,821,614	16,820,773—16,821,614	No gene	-	-
Chr2—6,219,027	6,216,101—6,219,027	No gene	-	-
Chr3—33,502,802	33,497,846—33,502,802	Jr03_26310	**33,501,980—33,503,100**	Dehydration-responsive element-binding protein 2A;Oryza sativa
Chr6—1,791,694	1,786,880—1,823,176	Jr06_01690	1,786,424—1,787,349	Citrate-binding protein;Hevea
		Jr06_01700	**1,788,194—1,806,047**	Sterol 3-beta-glucosyltransferase UGT80B1;Arabidopsis
		Jr06_01710	1,808,219—1,810,023	UPF0496 protein 3;Oryza sativa
		Jr06_01720	1,823,106—1,827,332	G-box-binding factor 4;Arabidopsis
		Jr06_01730	1,823,219—1,826,939	G-box-binding factor 4;Arabidopsis
Chr7—7,521,261	7,506,157—7,549,756	Jr07_06700	**7,511,949—7,525,431**	U-box domain-containing protein 33;Arabidopsis
		Jr07_06710	7,548,667—7,554,856	U-box domain-containing protein 33;Arabidopsis
Chr15—15,524,979	15,449,989—15,584,983	Jr15_10070	15,445,471—15,450,221	E3 ubiquitin-protein ligase MBR1;Arabidopsis
		Jr15_10080	15,479,372—15,486,594	Not annotated
		Jr15_10090	15,488,719—15,493,543	Nuclear poly(A) polymerase 1;Arabidopsis
		Jr15_10100	**15,506,085—15,527,810**	Probable protein phosphatase 2C 8;Arabidopsis
		Jr15_10110	15,527,700—15,527,810	T-complex protein 1 subunit delta;Arabidopsis
		Jr15_10120	15,533,040—15,546,025	DExH-box ATP-dependent RNA helicase DExH3;Arabidopsis
		Jr15_10140	15,584,523—15,586,811	Probable S-adenosylmethionine-dependent methyltransferase At5g38780;Arabidopsis
Chr16—13,373,812	13,364,187—13,405,303	No gene	-	-

[a] LD blocks are defined using the "solid spine of LD" method implemented in HaploView software

[b] The candidate genes in bold overlap the physical position of the associated SNP

Discussion

High levels of P4 are specific to *J. regia*, stable between years, not related to laciniate foliage, and not correlated to other traits

The large variability observed in P4 content in 2021 within the INRAE walnut germplasm collection is specific to *J. regia* (from 34.1 to 287.5 µg/g), compared to some accessions from wild relative species *J. mollis*, *J. nigra* and *J. major* (from 0.1 to 1.1 µg/g). Thought to exist only in animals, progesterone was first detected in apple (*Malus domestica*) seeds at levels around 0.5 µg/g [29], so at similar levels found in our wild relative individuals of *Juglans* species. Progesterone was also detected in black pine and loblolly pine (*Pinus nigra* and *Pinus taeda*) at various levels according to the organ, from 0.08 µg/g in pollen to 15.5 µg/g in dry mature wood [30, 31]. The P4 content in our *J. regia* germplasm is higher, even compared to the work of [10] where 1 µg/g was found in the same species. Such differences can be explained by extraction and quantification methods used. Since these early detections of P4 in plants, many methods were used, especially thin-layer chromatography, radioimmunoassays (RIA) and gas chromatography followed by single or tandem mass spectrometry (GC–MS/MS) [32]. The EIA method used here, which is close to the RIA method, is sensitive for detection of P4 content in walnut leaves and has a good specificity [11]. It gave concentrations in the same range as GC–MS/MS (tens of mg/kg dry leaves) even though concentrations in the same samples tended to be higher using EIA rather than GC–MS/MS [11]. EIA allowed to perform a large number of samples at low cost which would have been impossible with GC–MS/MS. Overall, the P4 concentrations measured in the present experiment were much higher than the concentrations previously observed in plants in the literature [32]. This places the walnut and their leaves in a very specific situation in plants.

When quantifying the P4 content a second time in 2023 in six accessions having contrasted phenotypes in 2021, we did not find significant difference. This suggests that the P4 content is stable between years for an accession and that our one-year phenotyping must be sufficient to conduct a powerful GWAS. We did not find as well significant difference in P4 content when testing two locations and two positions of the shoots. However, even though it is not significant, sampling at Toulenne and at the bottom of the tree might be preferable to obtain higher levels of P4 for genotypes having more P4.

Regarding laciniate walnut accessions, the two outliers in 2021 still gave high levels of P4 concentration in 2023. However, for the two new laciniate walnut accessions collected only in 2023 ('Lacinié' and 'Lacinié Strasbourg'),

the tremendous lower contents suggest that high levels of P4 are not specific is the peculiar laciniate foliage in walnut. The absence of correlation between P4 and any other traits suggests that progesterone is not involved in phenological events (such as budbreak and flowering dates) nor fruit characteristics (nut volume, nut shape, ease of cracking and nutritional composition) in walnut. Moreover, global foliage diseases based on brown spots observed in June and August 2023 are mainly related to walnut blight and walnut anthracnose diseases. The absence of correlation between the P4 content from 2021 and global foliage diseases, even though phenotyping was conducted in different years, indicates that diseases are not influenced by the P4 content.

A major locus on chromosome 2 and a candidate gene on chromosome 6 identified for the variation of P4 content in *J. regia*

A total of seven associations with the P4 content were identified, depending on the GWAS model used and the number of PCs included. Associations in agreement between GWAS models and included covariates give weight to their reliability. Thus, the association on chromosome 2 detected three times (SNP AX-170913249, position 6,219,027 bp) can be considered as a major locus, as it remains the major one when discarding the two laciniate outlier accessions. However, no gene was found within the corresponding LD block, probably because LD is too strong in this genomic area in our panel. When looking for a larger region (200 kb window), the nine genes identified are not related to hormone metabolism. This suggests that the effect of this locus on P4 content may consist in potential gene regulatory mechanisms or long-range, distant interactions.

The association on chromosome 6 is within a gene encoding for a "*sterol 3-beta-glucosyltransferase UGT80B1*" identified in *Arabidopsis thaliana*. This enzyme catalyzes the attachment of an UDP-glucose to a sterol, forming a sterol glucoside [33]. It has been showed in *Digitalis lanata* that sitosterol could be converted in progesterone [34]. More precisely, pregnenolone is formed from side-chain-cleavage of both sitosterol and cholesterol, and is converted to isoprogesterone before being converted to progesterone [35], so sterols can lead to formation of progesterone [36]. While sterol 3-beta-glucosyltransferase does not directly participate in the biosynthesis of progesterone, its role in converting free sterols to sterol glucosides may be crucial for maintaining the balance of sterol pools, essential precursors for the biosynthesis of steroid hormones.

Using the functional annotation of the genome, 23 genes that include "steroid" as keyword in their name were identified in 12 about the 16 chromosomes, but all are more distant than 1 Mb of any GWAS peak. Three of them are genes encoding for a "*membrane steroid-binding protein 1 (or 2)*", on chromosomes 1, 3 and 10, confirming that steroid hormones receptors are present and may have a crucial role also in walnut tree hormone signalization. The gene on chromosome 1, *Jr01_12240*, starts at 9,315,325 bp whereas the SNP associated is at 16,821,614 bp. The gene on chromosome 3, *Jr03_06480*, starts at 4,908,396 bp whereas the SNP associated is at 33,502,802 bp. [37] first identified "*membrane steroid-binding protein 1*" encoding genes in *A. thaliana* (*AtMSBP1*) as steroid hormones receptors, including progesterone, and homologous genes were then studied in *Oryza sativa* by [8]. Interestingly, the 3 genes above coding for steroid receptors owns to the large family of steroid membrane receptors [38] and not of steroid nuclear receptors [39] which does not seem present in the walnut genome as in any other plants [40]. However, these later ones are considered as those involved in many effects of steroids in vertebrates. Our SNP-P4 associations identified may intervene in a broader regulating metabolic pathway associated to steroid hormones.

Progesterone is involved in shoot and root growth, seed germination, reproductive development, and biotic/abiotic stress regulation [9], but further studies on walnut tree are needed to better understand its role in this species. It is also interesting to mention that in the walnut tree, only the leaves contain substantial amounts of progesterone [11]. This probably means that its role would be exerted in leaves rather than in roots, seed or reproductive organs such as flowers and fruits. Moreover, even though we did not detect here any link with global foliage diseases, we cannot exclude that progesterone acts not at the walnut tree itself, but rather interacts with pests or diseases transmitted by aphids. Five walnut cultivars were already reported to largely differ regarding the reproductive efficiency of the "dusk-veined aphid" (*Panaphis juglandis*), which is a major pest in walnut production [41].

For practical applications, these results could have significant implications for walnut breeding programs worldwide. Identifying SNPs associated with P4 content in walnut leaves enables breeders to adopt marker-assisted selection strategies, accelerating the development of new cultivars that combine high walnut yield with elevated P4 levels in the foliage. Such dual-purpose cultivars could address multiple agricultural and industrial needs, providing a sustainable source of P4 for pharmaceutical or nutraceutical applications while maintaining excellent nut production. Furthermore, this approach could optimize land use efficiency and expand the economic value of walnut orchards by leveraging leaves as a secondary product. In France, for instance, walnut is an agricultural sector in danger because of

climate change limiting walnut production. These findings could help producers to improve the value of their orchards in this difficult environmental context.

Conclusions

High concentration of P4 was found in fresh leaves of several accessions of *J. regia* trees of the INRAE walnut germplasm collection. By conducting GWAS using multi-locus models, we highlighted genomic regions linked to the variation of this concentration among the 170 accessions. Significant associations were identified on seven about the 16 chromosomes in walnut. The association on chromosome 2 seems particularly reliable, as it was detected using three model/PC configurations, but we failed in identifying any candidate gene in this region. Candidate genes were identified on chromosomes 3, 6, 7 and 15. The gene encoding for a sterol 3-beta-glucosyltransferase on chromosome 6 appears to be an interesting candidate. The presence of steroid membrane receptors genes and the absence of steroid nuclear receptor genes in the walnut genome could be crucial to decipher the role of P4 in plant physiology. These results can speed up the use of plant natural source of P4 for the veterinary and human health. As prospects, transient transformation of the identified candidate gene, quantification of precursors of P4 and of P4 membrane receptors in walnut leaves could be interesting to better understand the functional role of this gene within the metabolism of steroids in walnut.

Abbreviations

BIC	Bayesian information criterion
BLINK	Bayesian-information and linkage-disequilibrium iteratively nested keyway
EIA	Enzyme immunoassay
FarmCPU	Fixed and random model circulating probability unification
GAPIT	Genome association and prediction integrated tool
GC-MS	Gas chromatography-mass spectrometry
GWAS	Genome-wide association study
INRAE	Institut national de recherche pour l'agriculture, l'alimentation et l'environnement
IQR	Interquartile range
LD	Linkage disequilibrium
MAF	Minor allele frequency
MLMM	Multi-locus mixed-model
P4	Progesterone
PC	Principal component
RIA	Radio immunoassay
SNP	Single nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11341-2.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank the INRAE Prunus-Juglans Biological Resources Center for the maintenance of the collection, and more information is available at: Prunus-Juglans BRC, INRAE, member of BRC4Plants, INRAE, 2024. Biological Resource Centres for plants of AgroBRC-RARe. https://doi.org/10.17180/WN42-3J20. We thank the Fruit Tree Experimental Units of the INRAE in Toulenne and Bourran for helping us to collect the leaves, and more information is available at: UEA. (2024). Arboricultural Experimental Facility. INRAE. https://doi.org/10.17180/9ST1-4J21.

Authors' contributions

PC, AB and MK initiated the project. AB performed statistical and GWAS analyses, and wrote the manuscript. A-LL, CP and DG performed the P4 quantification. MD and M-LG provided fresh leaves of the INRAE walnut germplasm collection. DR conducted global foliage diseases observations. ED provided its experience to interpret the GWAS results. PC and AB conceived and coordinated the research. All co-authors revised the manuscript.

Funding

This work was supported by the Centre Val de Loire regional authority under CAPRINOIX Contract Number 85-C4646, by the programme cadre CNE "Maîtrise de la reproduction des petits ruminants" and by INRAE-PHASE specific funding. Phytosterol assays were supported by the INRAE grant Vegetruies of the METABIO programme.

Data availability

The phenotypic raw dataset generated and used in this work is available in Table S1. The SNP genotyping raw dataset in "hapmap" format is freely and openly accessed on the "Portail Data INRAE" official institute repository, via the identifier "INRAE's Walnut Genotyping Resources" and the following Digital Object Identifier (DOI): https://doi.org/10.15454/XPKII8. The dataset is called "3_GWAS_SNP_hapmap.txt". The additional file called "5_List_of_ID_SNP.tab" allows to link the array identifier name of the accessions used in this study with the identifier name of the INRAE walnut germplasm collection.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 November 2024 Accepted: 10 February 2025
Published online: 15 February 2025

References

1. Pellicer-Rubio MT, Ferchaud S, Fréret S, Tournadre H, Fatet A, Boulot S, Pavie J, Leboeuf B, Bocquier F. Les méthodes de maîtrise de la reproduction disponibles chez les mammifères d'élevage et leur intérêt en agriculture biologique. Prod Anim. 2009;22(3):255–70.
2. Gonzalez-Bulnes A, Menchaca A, Martin GB, Martinez-Ros P. Seventy years of progestagen treatments for management of the sheep oestrous cycle: where we are and where we should go. Reprod Fertil Dev. 2020;35(2):441–52.
3. Prior JC. Progesterone for treatment of symptomatic menopausal women. Climateric. 2018;21(4):358–65.
4. Eden J. The endometrial and breast safety of menopausal hormone therapy containing micronised progesterone: A short review. Aust N Z J Obstet Gynaecol. 2017;57(1):12–5.
5. Sato K, Seto K. The effect of *Dioscorea esculenta* powder on prostaglandin E2 and cytochrome c oxidase subunit 2 levels, menstrual pain, and premenstrual syndrome in young women: A randomized double-blind

controlled trial. Nutr Health. 2022. https://doi.org/10.1177/02601060221130889.
6. Simons R, Grinwich D. Immunoreactive detection of four mammalian steroids in plants. Can J Bot. 1989;67:288–96.
7. Simerský R, Novák O, Morris DA, Pouzar V, et al. Identification and quantification of several mammalian steroid hormones in plants by UPLC-MS/MS. J Plant Growth Regul. 2009;28:125–36.
8. Iino M, Nomura T, Tamaki Y, Yamada Y, et al. Progesterone: its occurrence in plants and involvement in plant growth. Phytochemistry. 2007;68(12):1664–73.
9. Li H, Chen L, Chen H, Xue R, et al. The Role of Plant Progesterone in Regulating Growth, Development, and Biotic/Abiotic Stress Responses. Int J Mol Sci. 2022;23(18):10945.
10. Pauli GF, Friesen JB, Gödecke T, Farnsworth NR, et al. Occurrence of progesterone and related animal steroids in two higher plants. J Nat Prod. 2010;73:338–45.
11. Chemineau, P., Lainé A.L., Gennetay, D., Porte, C., Chesneau, D., Laclie C., Goudet, G., Meunier, M., Delmas, M., Greil, M.L., Lière, P., Pianos A., Bernard, A., Dirlewanger, E., Delgadillo, J.A., Keller, M. The walnut tree as a source of progesterone for reproductive control in goats. Animal 2024 (accepted).
12. Bernard, A., Marrano, A., Donkpegan, A., Brown, P. J. *et al*. Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian Walnut (*Juglans regia* L.). BMC Genomics 2020a;21:203.
13. Bernard, A., Crabier, J., Donkpegan, A., Marrano, A. *et al*. Genome-wide association study reveals candidate genes involved in fruit trait variation in Persian walnut (*Juglans regia* L.). Front. Plant. Sci. 2020b;11:607213.
14. Roux-Cuvelier M, Grisoni M, Bellec A, Bloquel E, Charron C, et al. Conservation of horticultural genetic resources in France. Chronica Horticulturae. 2021;61(2):21–36.
15. Canépa S, Lainé AL, Bluteau A, Fagu C, Flon C, Monniaux D. Validation d'une méthode immunoenzymatique pour le dosage de la progestérone dans le plasma des ovins et des bovins. Cahier des Techniques de l'INRA. 2008;64:19–30.
16. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2, 2023, https://rpkgs.datanovia.com/rstatix/.
17. Wickham H, Averick M, Bryan J, Chang W, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4(43):1686.
18. Marrano A, Martínez-García PJ, Bianco L, Sideli GM, Di Pierro EA. *et al*. A new genomic tool for walnut (Juglans regia L.): development and validation of the high-density Axiom™ J. regia 700K SNP genotyping array. Plant Biotechnol J. 2019;17:1027–1036.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR. *et al*. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2008;81. Available online at: http://pngu.mgh.harvard.edu/purcell/plink/.
20. Wang J, Zhang Z. GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. Genom Proteom Bioinform. 2021;19(4):629–40.
21. Lipka AE, Tian F, Wang Q, Peiffer J, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–9.
22. Segura V, Vilhjálmsson BJ, Platt A, Korte A, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 2012;44:825–30.
23. Liu X, Huang M, Fan B, Buckler ES, et al. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 2016;12: e1005767.
24. Huang, M., Liu, X., Zhou, Y., Summers, R. M. *et al*. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 2018;8(2):giy154.
25. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.
26. Marrano, A., Britton, M., Zaini, P. A., Zimin, A. V. *et al*. High-quality chromosome-scale assembly of the walnut (*Juglans regia* L) reference genome. GigaScience 2020;9: giaa050.
27. Wei, T., Simko, V. R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.94), 2024, https://github.com/taiyun/corrplot.
28. Peterson, B. G., Carl, P. PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis. R package (Version 2.0.6), 2020, https://github.com/braverock/performanceanalytics.
29. Gawienowski AM, Gibbs CC. Identification of cholesterol and progesterone in apple seeds. Steroids. 1968;12:545–50.
30. Šaden-Krehula M, Tajić M, Kolbah D. Sex hormones and corticosteroids in pollen of *Pinus nigra*. Phytochemistry. 1979;18:345–6.
31. Carson, J. D., Jenkins, R. L., Wilson, E. M., Howell, W. M., Moore, R. Naturally occurring progesterone in loblolly pine (*Pinus taeda* L.): A major steroid precursor of environmental androgens. Environ. Toxicol. Chem. Int. J. 2008;27:1273–1278.
32. Janeczko A. The presence and activity of progesterone in the plant kingdom. Steroids. 2012;77(3):169–73.
33. Duperon R, Duperon P. Intracellular-localization of UDP-glucose-sterol glucosyl transferase and UDP-galactose-sterol galactosyl transferase activities in the leaves of tomato (Solanum lycopersicon L, Solanaceae). Comptes Rendus de l'Académie des Sciences, Série. 1987;III(304):235–8.
34. Bennett R, Heftmann E, Winter BJ. Conversion of sitosterol to progesterone by *Digitalis lanata*. Die Nat. 1969;56:463.
35. Lindemann P. Steroidogenesis in plants–Biosynthesis and conversions of progesterone and other pregnane derivatives. Steroids. 2015;103:145–52.
36. Tarkowská D. Plants are Capable of Synthesizing Animal Steroid Hormones. Molecules. 2019;24(14):2585.
37. Yang XH, Xu ZH, Xue HW. Arabidopsis Membrane Steroid Binding Protein 1 Is Involved in Inhibition of Cell Elongation. Plant Cell. 2005;17(1):116–31.
38. Thomas P, Pang Y, Jan Kelder J. Membrane progesterone receptors on the cell membrane: A review highlighting potential export motifs in mPRα regulating its trafficking to the cell surface. Steroids. 2023;199: 109295.
39. Loosfelt H, Atger M, Misrahi M, Guiochon-Mantel A, Meriel C, Logeat F, Benarous R, Milgrom E. Cloning and sequence analysis of rabbit progesterone-receptor complementary DNA. PNAS. 1986;83(23):9045–9.
40. Baker ME. Steroid receptors and vertebrate evolution. Mol Cell Endocrinol. 2019;496: 110526.
41. Akkopru EP, Atlihan R, Okut H, Chi H. Demographic Assessment of Plant Cultivar Resistance to Insect Pests: A Case Study of the Dusky-Veined Walnut Aphid (Hemiptera: Callaphididae) on Five Walnut Cultivars. J Econ Entomol. 2015;108:379–87.

Publisher's Note