Unravelling DNA methylation dynamics during developmental stages in *Quercus ilex* subsp. *ballota* [Desf.] Samp.

Mónica Labella-Ortega[1*], Carmen Martín[2], Luis Valledor[3], Stefano Castiglione[4], María-Ángeles Castillejo[1], Jesús V. Jorrín-Novo[1] and María-Dolores Rey[1*]

Abstract

Background DNA methylation is a critical factor influencing plant growth, adaptability, and phenotypic plasticity. While extensively studied in model and crop species, it remains relatively unexplored in holm oak and other non-domesticated forest trees. This study conducts a comprehensive in-silico mining of DNA methyltransferase and demethylase genes within the holm oak genome to enhance our understanding of this essential process in these understudied species. The expression levels of these genes in adult and seedling leaves, as well as embryos, were analysed using quantitative real-time PCR (qRT-PCR). Global DNA methylation patterns were assessed through methylation-sensitive amplified polymorphism (MSAP) techniques. Furthermore, specific methylated genomic sequences were identified via MSAP sequencing (MSAP-Seq).

Result A total of 13 DNA methyltransferase and three demethylase genes were revealed in the holm oak genome. Expression levels of these genes varied significantly between organs and developmental stages. MSAP analyses revealed a predominance of epigenetic over genetic variation among organs and developmental stages, with significantly higher global DNA methylation levels observed in adult leaves. Embryos exhibited frequent demethylation events, while *de novo* methylation was prevalent in seedling leaves. Approximately 35% of the genomic sequences identified by MSAP-Seq were methylated, predominantly affecting nuclear genes and intergenic regions, as opposed to repetitive sequences and chloroplast genes. Methylation was found to be more pronounced in the exonic regions of nuclear genes compared to their promoter and intronic regions. The methylated genes were predominantly associated with crucial biological processes such as photosynthesis, ATP synthesis-coupled electron transport, and defence response.

Conclusion This study opens a new research direction in analysing variability in holm oak by evaluating the epigenetic events and mechanisms based on DNA methylation. It sheds light on the enzymatic machinery governing DNA (de)methylation, and the changes in the expression levels of methylases and demethylases in different organs along the developmental stages. The expression level was correlated with the DNA methylation pattern observed,

*Correspondence:
Mónica Labella-Ortega
b62laorm@uco.es
María-Dolores Rey
b52resam@uco.es
Full list of author information is available at the end of the article

showing the prevalence of *de novo* methylation and demethylation events in seedlings and embryos, respectively. Several methylated genes involved in the regulation of transposable element silencing, lipid biosynthesis, growth and development, and response to biotic and abiotic stresses are highlighted. MSAP-seq integrated with whole genome bisulphite sequencing and advanced sequencing technologies, such as PacBio or Nanopore, will bring light on epigenetic mechanisms regulating the expression of specific genes and its correlation with the phenotypic variability and the differences in the response to environmental cues, especially those related to climate change.

Keywords Holm oak, Methylated genes, DNA methyltransferase, DNA demethylase, MSAP-Seq

Introduction

Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) is a key species in the Mediterranean Basin and plays a vital role in the Dehesa agrosilvopastoral ecosystem. This system is crucial for biodiversity, holding significant environmental, economic, and social value [1]. In recent years, there has been a significant surge in interest regarding holm oak, particularly in molecular biology studies [2]. Given its remarkable adaptability to arid and semi-arid environments, its importance is expected to rise as the impacts of climate change intensify [3]. *Quercus ilex* exhibits considerable variability at the phenotypic, morphometric, physiological, and molecular levels, characteristics typical of non-domesticated, long-lived, allogamous, and anemophilous species [4, 5]. The extent to which this variability is influenced by genomic, epigenomic, and environmental factors remains largely unexplored. Molecular research on holm oak is inherently challenging due to its recalcitrance to standard molecular genetics protocols and in vitro propagation techniques. The effectiveness of these methods varies considerably across different genotypes, complicating efforts to elucidate the molecular mechanisms underlying the species' traits and adaptations [6–8]. Moreover, the cross-pollinated nature of holm oak presents a further challenge in conducting genetic studies in natural environments. The identity of the pollen donor often remains unknown, making it difficult to interpret experimental data from a genetic standpoint.

Since 2004, our research group has extensively characterized the transcriptome, proteome, and metabolome of various tissues from this forest tree species [2]. The recent release of its first draft of the genome marks a crucial milestone [9], opening new possibilities for functional genomics and epigenetics studies. While the genomes of several *Quercus* species have been released in the past decade, enabling functional genomics studies on development and stress response, the role of epigenetic marks in regulating gene expression and their integration into the developmental program remains unexplored in holm oak. To date, only a limited number of papers studying DNA methylation in *Quercus* have been published, including *Q. suber* L [10]., *Q. robur* L [11]. and *Q. lobata* Née [12]. Notably, just one publication has investigated *Q. ilex*, with Rico et al. [13] examining global methylation patterns in mature trees in response to drought.

DNA methylation is the most stable and common epigenetic mark in plants [14], maintaining genome stability and determining the expression of genes involved in plant growth, development, and environmental stresses [15]. Growth and development in plant species is orchestrated by the regulation of gene expression, which is refined by epigenetic mechanisms ensuring the accurate execution of the developmental program [16]. DNA methylation in plants occurs primarily in three distinct sequence contexts: symmetric CG and CHG, as well as asymmetric CHH (where H represents A, T or C). Among these contexts, methylation is most prevalent in the CG context in plants [17]. The regulation of DNA methylation in all three contexts (CG, CHG, and CHH) is primarily controlled by the concerted action of DNA methyltransferases and demethylases [18]. Like most plants, those in the family Fagaceae possess four classes of DNA methyltransferases, namely: Methyltransferase 1 (MET1), Chromomethylases (CMTs), DNA Methyltransferase Homologue 2 (DNMT2) and Domains Rearranged Methyltransferases (DRMs) [19]. MET1, CMT3 and CMT2 are responsible for maintaining CG, CHG and CHH methylation changes, respectively, and DRM2 for *de novo* DNA methylation. DNA demethylases include Repressor of Silencing 1 (ROS1), Demeter (DME) and Demeter-like 2 and 3 (DML2 and 3).

The Methylation Sensitive Amplified Polymorphism (MSAP) technique, widely used in non-model species [20], allows the determination of methylation patterns and the degree of methylation at cytosine residues in CCGG sites based on differences in methylation sensitivity between *HpaII* and *MspI* enzymes [21]. An advanced approach based on MSAP analysis followed by sequencing of PCR products previously digested by *HpaII* and *MspI* (MSAP-seq) was recently reviewed by Agius et al. [22]. The approach allows to identify regions of the genome that are differentially methylated at CCGG sites and to associate them to specific genetic pathways and biological processes. The study of DNA methylation in non-model species, such as holm oak, is gaining significant attention due to the remarkable ability of these species to adapt to extreme environments. Epigenetic marks are included in breeding programs as molecular markers

Labella-Ortega *et al. BMC Plant Biology*   (2024) 24:823

Page 3 of 17

for the selection of individuals with phenotypic traits of interest, such as increased growth, yield and tolerance to both biotic and abiotic factors [20]. However, the high levels of genetic diversity within these species pose a substantial challenge. In addition, intraspecific epigenetic variation may further complicate the identification of consistent methylation patterns and their association with specific phenotypic traits. Understanding the role of epigenetics in shaping tree phenotypes could reveal key factors involved in regulating growth traits and improve our understanding of the evolutionary potential of this forest species.

The primary objective of this study was to assess the dynamics of global DNA methylation levels across three developmental stages – leaves from adult trees and seedlings, as well as embryos – while concurrently comparing the transmission of methylation patterns among these stages. The specific objectives were as follows: (*a*) to identify and determine the gene expression of DNA methyltransferases and demethylases; (*b*) to analyse and compare the global DNA methylation patterns in relation to three different developmental stages; and (*c*) to identify methylated genomic sequences associated to the developmental stages. For that, the global DNA and gene-specific methylation patterns have been analysed by using the MSAP and MSAP-seq techniques. Data have been correlated with the expression level of DNA methyltransferase and demethylases, as determined by RT-q-PCR.

Results

Identification of holm oak DNA methyltransferases and demethylases

A total of 13 DNA methyltransferase and 3 demethylase genes were identified in the holm oak assembled genome. Based on available data from *A. thaliana*, they included 8 maintenance methyltransferases, 5 *de novo* methyltransferases and 3 demethylases (Table 1). The identified maintenance methyltransferases included one each for QiMET1 (gene-16946), QiCMT1 (gene-6353) and QiDNMT2 (gene-33301), in addition to three (gene-37003, gene-37004 and gene-37005) for QiCMT2 and two (gene-7580 and gene-19767) for QiCMT3. Those corresponding to *de novo* methyltransferases included one gene for QiDRM3 (gene-4322), and two each for QiDRM2 (gene-35436 and gene-19781) and QiDRM2-like (gene-6758 and gene-28358). DNA methyltransferase genes were distributed in chromosomes 1, 2, 3, 5, 6, 10 and 12, and demethylase genes in chromosomes 3, 4 and 7 (Table 1). The QiDRM2-like gene (gene-6758) was identified in chromosome 0. This was a result of less than 1% of the sequences (37 contigs, size 6.69 Mbp) allowing assignation in the first draft of the holm oak genome and having to be merged and placed in chromosome 0

Table 1 The information regarding the holm oak DNA methyltransferase and DNA demethylase genes

		Enzyme name	Enzyme abbreviation	A. thaliana gene ID	Holm oak gene ID	Location	Chromosome
DNA methyltransferase	Maintenance	METHYLTRANSFERASE 1	MET1	AT4G13610	gene-16,946	34,417,843–34,427,156	10
		CHROMOMETHYLASE 1	CMT1	AT1G80740	gene-6353	54,932,765–54,943,949	6
		CHROMOMETHYLASE 2	CMT2	AT4G19020	gene-37,003	365,798–370,564	12
					gene-37,004	371,305–379,774	12
					gene-37,005	389,005–391,789	12
		CHROMOMETHYLASE 3	CMT3	AT1G69770	gene-7580	12,597,090–12,604,980	1
					gene-19,767	46,649,165–46,651,164	5
		DNA METHYLTRANSFERASE HOMOLOG 2	DNMT2	AT5G25480	gene-33,301	26,192,553–26,198,531	2
	De novo	DOMAINS REARRANGED METHYLTRANSFERASE 2	DRM2	AT5G14620	gene-35,436	75,000,996–75,012,312	2
					gene-19,781	47,188,727–47,190,135	5
		DOMAINS REARRANGED METHYLTRANSFERASE-like	DRM2-like	AT5G14620	gene-6758	5,302,736–5,310,516	0
					gene-28,358	59,797,624–59,800,185	3
		DOMAINS REARRANGED METHYLTRANSFERASE 3	DRM3	AT3G17310	gene-4322	13,506,645–13,516,860	6
DNA demethylases		DEMETER	DME	AT5G04560	gene-23,797	40,964,476–40,976,704	4
		DEMETER-LIKE	DML-like	AT3G10010/ AT4G34060	gene-25,932	4,109,061–4,127,563	3
		REPRESSOR OF SILENCING 1	ROS1	AT2G36490	gene-31,205	44,171,595–44,183,308	7

as a result [9]. Each enzyme sequence identified in the holm oak genome was functionally annotated and categorized based on its involvement in biological processes, molecular functions, or cellular components, as detailed in Additional File 1. Furthermore, each enzyme was classified based on its involvement in DNA methylation, methyltransferase activity, and nuclear localization. Most sequences exhibited matched identity greater than 90% between *Q. ilex* and *Q. lobata*, *Q. robur*, and *Q. suber* and *Q. variabilis*; by exception of gene-19,767 (QiCMT3) that had approximately 50% matched identity between *Q. ilex* and the other *Quercus* spp. (Additional File 2). All methyltransferases, belonging to the C-5 cytosine methyltransferase family (IPR001525), displayed a variety of functional domains: the Bromo Adjacent Homology domain (PF01426, BAH), the SAM-dependent methyltransferase DRM-type domain profile (PS51680, SAM_MT_DRM), the Ubiquitin-associated domain (PS50030, UBA), the C-5 cytosine-specific DNA methylase (PF00145, DNA_methylase), the Cytosine specific DNA methyltransferase replication foci domain (PF12047, DNMT1-RFD), and the Chromo (CHRomatin Organisation MOdifier) domain (PF0385, Chromo)

(Fig. 1A). MET1 was characterized by two BAH domains, two DNMT1-RFD domains, and one DNA_methylase domain. CMTs contained one BAH domain, one Chromo domain, and one DNA_methylase domain. QiCMT2.2, QiCMT2.3, and QsCMT2.2 lacked the BAH domain, QiCMT2.1 did not have both the DNA_methylase and Chromo domains, QiCMT1 did not show the DNA_methylase domain, and QiCMT3.2 was devoid of any domains. DNMT2 exclusively possessed the DNA_methylase domain. DRM2 and DRM2-like contained the UBA, SAM_MT_DRM, and DNA_methylase domains, with QiDRM2 lacking the latter one and in QiDRM2-like.2 and QsDRM2-like, the UBA domain. DRM3 was solely characterized by the SAM_MT_DRM domain. To explore the relationship of DNA methyltransferase and demethylase genes, a phylogenetic tree was constructed using the protein sequences from *A. thaliana*, *Q. robur*, *Q. lobata*, *Q. variabilis*, *Q. suber* and *Q. ilex* (Fig. 1). In this analysis, the maintenance DNA methyltransferases (MET and CMT in one clade, and DNMT in the other) were grouped separately from *the de novo* methyltransferases. Within each enzymatic category, *A. thaliana* tended to diverge prior to the rest of the *Quercus* species,
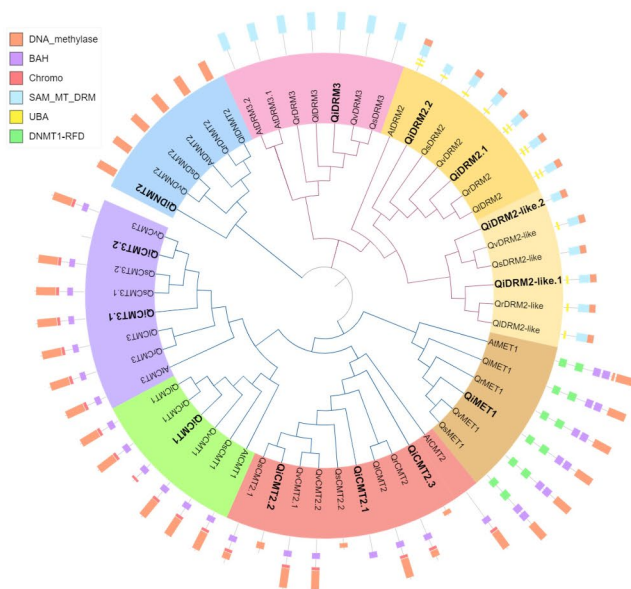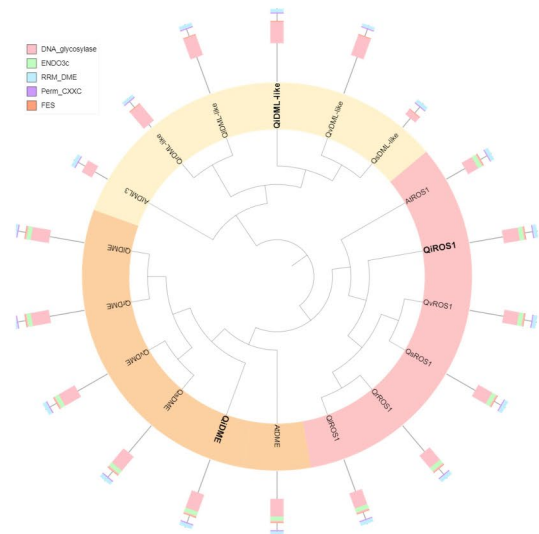

Fig. 1 Phylogeny of DNA methyltransferases and demethylases with their domain organization. The phylogenetic linage of (**A**) DNA methyltransferases was performed with Methyltransferase 1 (MET1), Chromomethylases 1 to 3 (CMT1–3), DNA (Cytosine-5-)-Methyltransferase 2 (DNMT2), Domains Rearranged Methyltransferases 2 and 3 (DRM2/3) and representing the conserved domains: DNA_methylase: the C-5 cytosine-specific DNA methylase (orange), BAH: Bromo Adjacent Homology domain (violet), Chromo: Chromo (CHromatin Organisation MOdifier) (red), SAM_MT_DRM: SAM-dependent methyltransferase DRM-type domain profile (blue), UBA: Ubiquitin-associated domain (yellow), and DNMT1-RFD: Cytosine specific DNA methyltransferase replication foci domain (green). The phylogenetic linage of (**B**) DNA demethylases was performed with DEMETER (DME), repressor of silencing 1 (ROS1) and Demeter-like (DML-like) and representing the conserved domains DNA_glycosylase: domain associated with DNA glycosylase activity (pink), ENDO3c: endonuclease III domain (green), RRM_DME: predicted RRM-fold domain (blue), Perm-CXXC: permuted zf-CXXC domain (violet), and FES: iron-sulfur cluster loop motif (orange). Different species *Arabidopsis thaliana* (At), *Quercus ilex* (Qi), *Quercus lobata* (Ql), *Quercus robur* (Qr), *Quercus suber* (Qs), and *Quercus variabilis* (Qv) were included

except for the DNMT enzyme (Fig. 1A). Demethylases were part of the DNA glycosylase family (IPR044811) and contained five key domains: a predicted RRM-fold domain (PF15628, RRM_DME), a permuted zf-CXXC domain (PF15629, Perm_CXXC), an iron-sulfur cluster loop motif (SM00525, FES), a domain associated with DNA glycosylase activity (IPR011257, DNA glycosylase), and the Endonuclease III domain (cd00056, ENDO3c), being the last one present in all demethylases except in AtDML3 and DML-like of the *Quercus* species (Fig. 1B). The DNA glycosylase domain was smaller in AtDML proteins and QsDML-like. The FES domain was absent in AtDML3. Phylogenetic analysis has revealed a distinct separation among the demethylases (DML, DME, and ROS), yet they are all grouped together.

Gene expression of DNA methyltransferases and demethylases

The expression of seven methyltransferase and two demethylase genes, out of the total 13 identified, was analysed at the three developmental stages (Fig. 2). For QiCMT2, QiCMT3, and QiDRM2, gene-37,003, gene-7580, and gene-35,436, respectively, were selected. Three

clear patterns were observed. The first corresponded to maintenance methyltransferases (QiMET1, QiCMT1, QiCMT2, and QiDNMT2) that were significantly upregulated in adults and embryos, with no differences between them, and with transcripts present at very low expression levels or not detected in leaves from seedlings. Within this first group, QiCMT3 showed a higher expression level in adult than in embryos. The second group included *de novo* methyltransferases (QiDRM2 and QiDRM3) that were significantly upregulated 6-fold and 14-fold in embryos, respectively, and 2-fold in seedlings. The third group of enzymes analysed were DNA demethylases, including QiDME and QiROS1. QiDME did not display significant differences in abundance across the various developmental stages, indicating a potentially stable role throughout the plant's growth and development.

Global DNA methylation patterns

A total of 387 MSAP fragments were scored of which 279 were identified as methylation-susceptible loci (MSL) and 108 as non-methylated loci (NML). Shannon's diversity index was higher in MSL (0.5793) than it was in
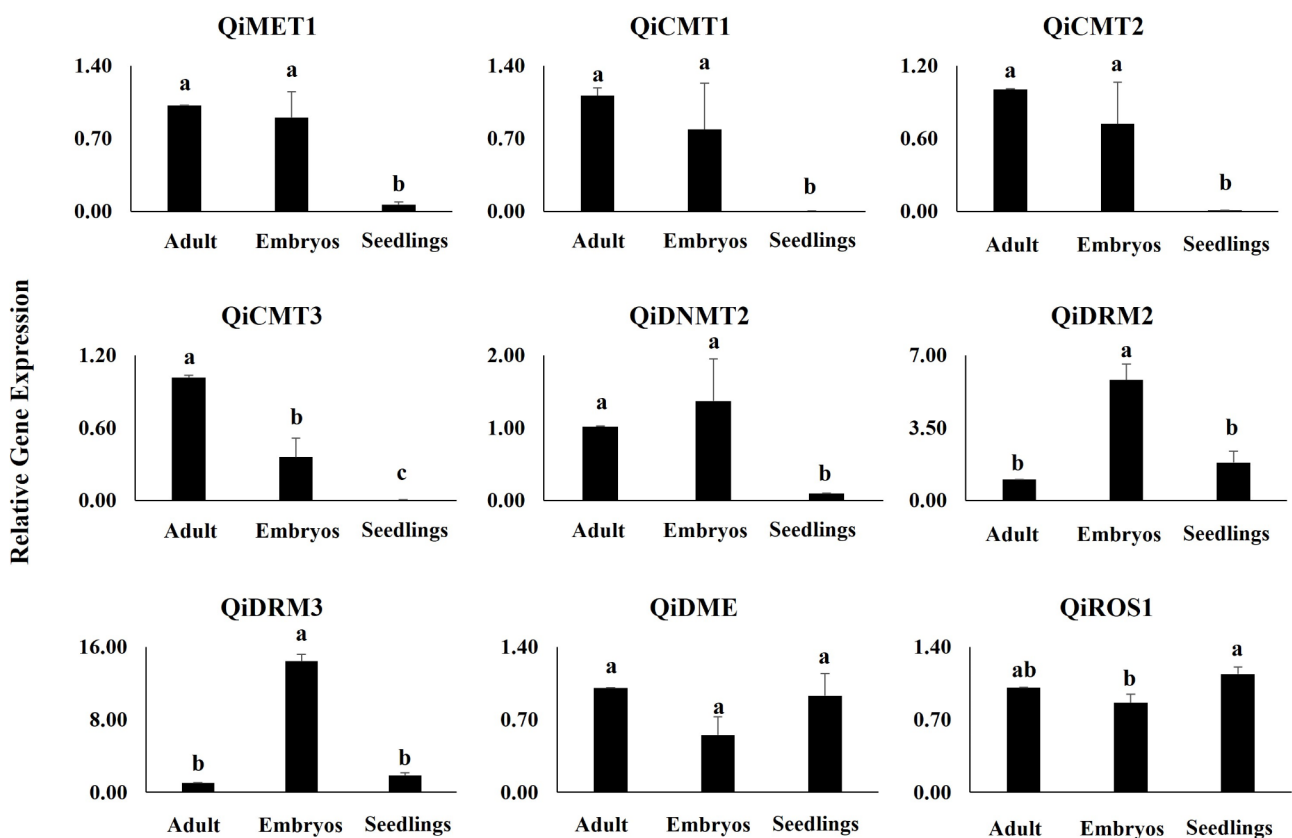
Fig. 2 Expression profiles for maintenance DNA methyltransferase (QiMET1, QiCMT1, QiCMT2; QiCMT3 and QiDNMTT2), *de novo* DNA methyltransferase (QiDRM2 and QiDRM3), and demethylase (QiDME and QiROS1) -encoding genes in adult, embryos, and seedlings of holm oak. The gene expression level is given in relative gene expression. Data are mean of three biological replicates ± standard error (SE). Different letters indicate significant differences between developmental stages according to ANOVA and LSD test

NML (0.3825), and differences between overall genetic and epigenetic diversity were significant (Wilcoxon Rank-sum test<0.0001). Analyses revealed differences in DNA cytosine methylation patterns across developmental stages (Fig. 3A-B), with estimated PhiPT=0.4020 (*p*<0.0001), with 59.8% of total epigenetic variability residing within developmental stages and 40.2% between stages. Genetic tests revealed 74.6% of variability to be within developmental stages and 25.4% between them (estimated PhiPT=0.2541, *p*<0.0001). A total of 171 MSAP fragments were scored in MSL at the three developmental stages. Unmethylated (type I) fragments were more abundant than methylated (type II−IV) fragments; also, the former were slightly more frequent in seedlings (Fig. 3C; Additional File 3). Total methylation amounted

to 38.6% in adults, 39.8% in embryos and 34.5% in seedlings. Fully methylated fragments (type III+IV) were more abundant than hemimethylated fragments (type II), especially in embryos (35.1% and 4.7%, respectively), followed by seedlings (32.7% and 1.8%, respectively) and adults (31.0% and 7.6%, respectively). Stabilization events often exceeded methylation events, with 31.4% in MSAP fragments. Methylated fragments exhibited the greatest numbers of methylation−demethylation events at all developmental stages, followed by demethylation events in embryos (8.2%) and *de novo* methylation events in adults and seedlings (4.6% and 5.2%, respectively; Fig. 3D).
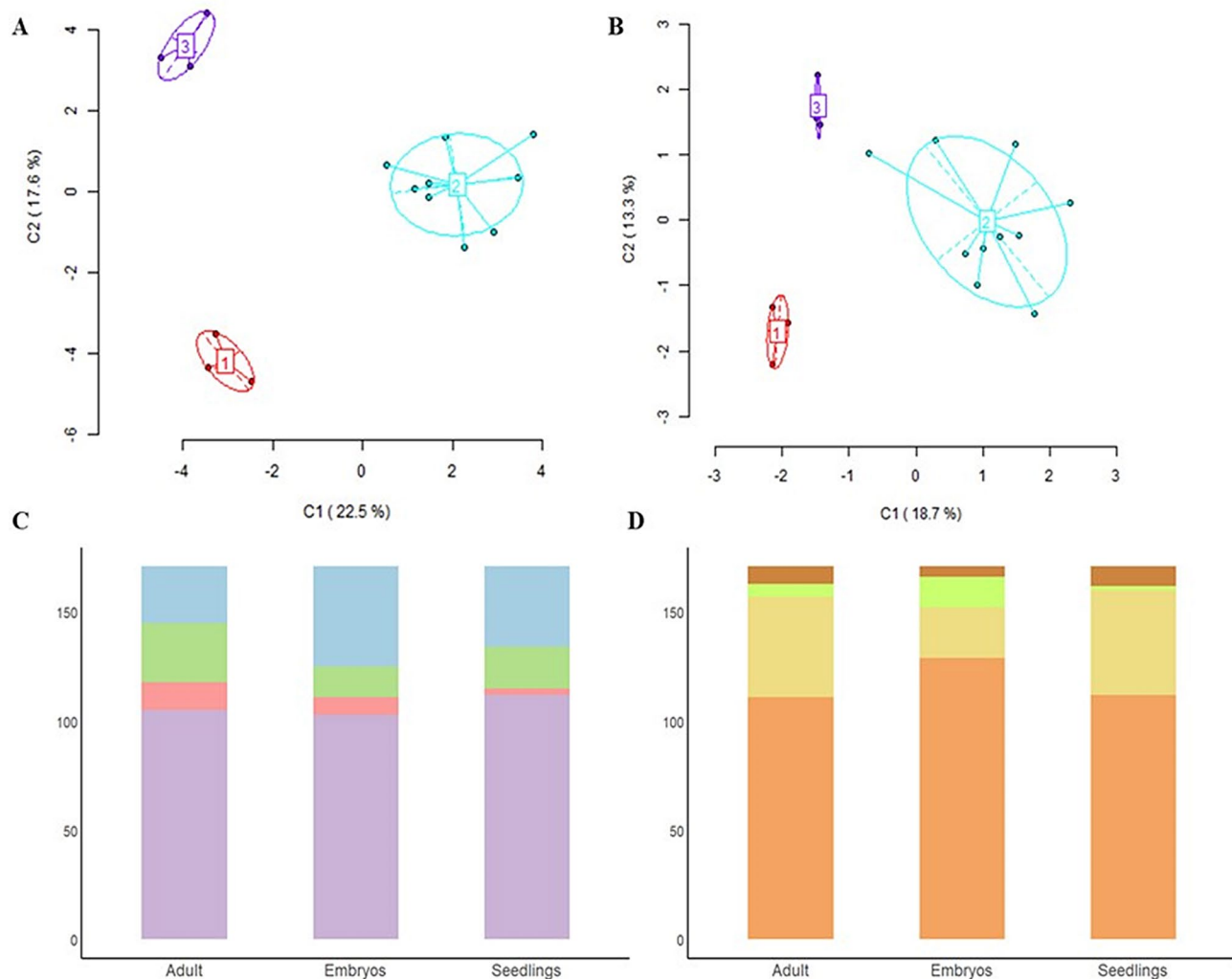

Fig. 3 Global DNA patterns for adult, embryos and seedlings of holm oak as established with MSAP. (**A**) PCoA plot of the fragments classified as MSL (methylation-susceptible loci). (**B**) PCoA plot of the fragments classified as non-methylated loci (NML). The values 1, 2 and 3 indicate adults, seedlings, and embryos, respectively. (**C**) Types I (purple), II (pink), III (green) and IV (blue) correspond to unmethylated, hemimethylated, hypermethylated and fully methylated fragments, respectively. (**D**) Stability (methylation-susceptible loci, MSL, and non-methylated loci, NML) (orange) and epigenetic changes (de-methylation, *de novo* methylation, and methylation−demethylation events, in green, brown, and yellow, respectively) detected at each developmental stage

Identification of methylated genomic regions

Sequencing of pre-selected PCR products provided an average ~2 million paired-end reads per sample and restriction enzyme (Table 2). That number reduced to ~900 000 after filtering for the presence of the HM-adapter, followed by the CGG sequence at 3′ or 5′ and 50 bp of minimal read length. The degree of mapping of the reads to the holm oak reference genome ranged from 74.40% (*HpaII*-digested seedlings) to 91.40% (*MspI*-digested adults) (Table 2). Scoring the genomic sequences observed in all three biological replicates for each developmental stage allowed the identification of a total of 281−415 sites in nuclear genes, 17−23 in chloroplast genes and 17−33 in repetitive regions. Finally, the dataset for the genomic sequences observed at the three developmental stages with both enzymes included 235 nuclear genes sites (106 exons, 41 introns and 88 promoters), 57 intergenic regions (258 CCGG sites), 11 chloroplast genes (27 CCGG sites) and 43 repetitive regions (53 CCGG sites) (Additional File 4). Out of the 573 analysed sequences, methylation was detected in various genomic regions, including 93 nuclear gene sites (consisting of 47 exons, 29 promoters, and 17 introns), 114 intergenic regions, 9 chloroplast genes, and 30 repetitive regions. These methylated sites were classified as type II, III, or IV and were found to be present in at least one developmental stage (Additional File 4).

A total of 17 nuclear gene sites showed two methylated sites in the same genomic region. That was the case with *Metalloendoproteinase 2-MMP*, which had two sites in exon 1, and *Putative uncharacterized protein ART2*, with two sites in the promoter (Additional File 4). Other genes exhibited a single methylated site in two different genomic sequences (e.g., *NADH-ubiquinone oxidoreductase chain 2*, with sites in exon 1 and the promoter; *E3 ubiquitin protein ligase RIE1*, with sites in exon 6 and the promoter; and *Protein Ycf2*, with sites in the promoter and intron 1). Genomic regions with more than three methylated sites in the same gene were also identified including *Photosystem I P700 chlorophyll a apoprotein A2* (exon 1, and introns 1 and 3), *Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta*, chloroplastic (three sites in the promoter), *Cytochrome f* (four sites in exon 2, one exon 1 and one in the promoter) and *uncharacterized protein ycf68* (three sites in the promoter, one in exon 4 and two in intron 1) (Additional File 4).

The degree of methylation was similar in adults (36.3%) and seedlings (37.7%), but somewhat lower in embryos (30.0%) (Table 3). More fully methylated genes were observed in adults (28.9%) than in seedlings (20.8%) and embryos (15.7%); in contrast, more hemi-methylated genes were scored in seedlings (16.9%) that in embryos (14.3%) and adult (7.3%). As regard genomic sequences, more marked methylation was observed in intergenic regions (15.4%) and nuclear genes (13.5%) than in repetitive sequences (4.5%) and chloroplast genes (1.3%) (Table 3). Also, type IV methylation was the most frequent in all genomic sequences, followed by type II methylation (Table 3).

Intergenic regions exhibited an increased occurrence of *de novo* methylation and demethylation events (Fig. 4). The number of events was dependent on the genomic sequence and developmental stage. In exons, adults and seedlings exhibited more *de novo* methylation events than demethylation events. In introns, *de novo* methylation events were observed mainly in embryos and demethylation events in adult plants. In promoters, adults and seedlings exhibited more *de novo* methylation and demethylation events, respectively. By contrast, in chloroplast genes, adults and embryos exhibited a similar number of *de novo* methylation events whereas seedlings only had demethylation events. In repetitive regions, *de novo* methylation events were more frequent in embryos than they were in adults and seedlings — which were similar in this respect. In intergenic events, embryos and adults had more *de novo* methylation and demethylation events, respectively. The number of methylation and demethylation events was much higher in intergenic events than in all other genomic sequences. In general, the genomic sequences exhibited substantial stability (68.5%) —by exception, methylated sequences had a slightly lower stability (8.8%).

The GO classification in Additional File 4 provides functional categorization of the genomic sequences associated with biological processes, cellular components, and molecular functions. When focusing on methylated

Table 2 Mean values of the MSAP-Seq statistics for data processing in adult, embryos and seedlings of holm oak

	Adult		Embryos		Seedlings	
	HpaII	**MspI**	**HpaII**	**MspI**	**HpaII**	**MspI**
Initial number of reads [mln]	1.82	2.15	1.88	2.12	2.11	2.33
Number of reads after filtering [mln]*	0.99	0.71	0.89	1.02	0.77	0.1
Percentage of mapped reads [%]	91.01	91.42	82.13	90.58	74.44	82.33
Total no. of common nuclear genes	313	308	415	369	349	281
Total no. of common chloroplast genes	23	22	19	20	22	17
Total no. of common repetitive regions	18	18	33	22	31	17

* The presence of the either *HpaII*-related or *MspI*-related adapters and a minimum length of 50 bp

Table 3 Total number of non-methylated (type I), hemimethylated (type II), hypermethylated (type III) and fully methylated (type IV) patterns in nuclear (exons, introns and promoters) and chloroplast genes, and repetitive regions identified in adult, embryos and seedlings, of holm oak

	Adult	Embryos	Seedlings
Nuclear genes (exons)			
I (unmethylated)	66	66	62
II (hemimethylated)	5	6	35
III (hypermethylated)	7	4	1
IV (fully methylated)	28	30	8
Nuclear genes (introns)			
I (unmethylated)	26	31	28
II (hemimethylated)	5	5	5
III (hypermethylated)	4	0	1
IV (fully methylated)	6	5	7
Nuclear gene (promoters)			
I (unmethylated)	68	64	62
II (hemimethylated)	8	7	12
III (hypermethylated)	3	5	1
IV (fully methylated)	9	12	13
Chloroplast genes			
I (unmethylated)	22	19	18
II (hemimethylated)	1	5	2
III (hypermethylated)	1	0	3
IV (fully methylated)	3	3	4
Repetitive regions			
I (unmethylated)	27	29	25
II (hemimethylated)	2	13	16
III (hypermethylated)	1	4	2
IV (fully methylated)	23	7	10
Intergenic Regions			
I (unmethylated)	156	192	162
II (hemimethylated)	21	46	27
III (hypermethylated)	20	8	8
IV (fully methylated)	61	12	61
Type I	365	401	357
Type II	42	82	97
Type III	36	21	16
Type IV	130	69	103
Total amplified genes	573	573	573
Fully methylated genes (%)[1]	28.97	15.71	20.77
Hemi-methylated genes (%)[2]	7.33	14.31	16.93
Total methylated genes (%)[3]	36.30	30.02	37.70

The percentage of polymorphic MSAP, fully methylated and hemimethylated fragments as well as total methylation were estimated using the formulas described by Abid et al. (2017): [1] Fully methylated fragments (%) = [(III+IV)/(I+II+III+IV)] × 100; [2] Hemimethylated fragments (%) = [(II)/(I+II+III+IV)] × 100; [3] Total methylation (%) = [(II+III+IV)/(I+II+III+IV)] × 100

nuclear genes present at one or more developmental stages, a notable finding was the prevalence of genes involved in photosynthesis, ATP synthesis, coupled electron transport, and defence response processes (Fig. 5A). Many genes involved in molecular functions were associated to NADH dehydrogenase (ubiquinone) activity, ATP binding or metal ion binding (Fig. 5B). On the other hand, genes in the cellular component category belonged to the chloroplast, mitochondrion, integral membrane component or nucleus categories mainly (Fig. 5C).

The genomic sequences exhibiting type II methylation events (i.e., hemimethylation) and/or type III events (hypermethylation) at two or more developmental stages allowed a set of 10 nuclear genes (6 exons and 4 promoters) to be selected as potentially methylated genes in holm oak (Table 4). Especially prominent among them were *Photosystem I P700 chlorophyll a apoprotein A1*, *Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta*, *Metalloendoproteinase 2-MMP* and *Dynamin-related protein 3 A*, involved in photosynthesis, fatty acid biosynthesis, defence response and peroxisome fission processes, respectively.

Discussion

Recent studies on several plant species have shown that DNA methylation strongly influence plant growth, development, adaptability, and phenotypic plasticity [23, 24]. It is therefore a crucial factor to consider in plant breeding programmes not only gene variants, but also DNA methylation and other epigenetics mechanisms determining how gene expression is regulated. This is the case of holm oak and other orphan, non-domesticated, species for which breeding is based on the selection of elite genotypes for ulterior clonal propagation [2]. Furthermore, research on epigenetic marks could shed light on the recalcitrance of the species to in vitro propagation [6, 7]. Moreover, the study of DNA methylation in holm oak is crucial to understand how these organisms adapt to their environment, contributing to the development of strategies for conservation and management. Recognizing the limited research on DNA methylation in holm oak, this study aims to further our understanding of its underlying mechanisms and significance. To accomplish this goal, the researchers examined the differences in DNA-cytosine methylation patterns across various developmental stages, including embryos, leaves from seedlings, and leaves from adult trees.

We took advantage of the recent publication of the first draft of the holm oak genome to identify the complete list of DNA methyltransferase- and demethylase-encoding genes [9]. This allowed us to identify 13 DNA methyltransferase genes and three demethylase genes involving eight maintenance methyltransferases, five *de novo* methyltransferases and three demethylases (Table 1). All identified enzymes were consistent with previous reports on other species of the family Fagaceae [19, 25], — by exception, DME3 (DME-like), which is involved in active DNA demethylation in *A. thaliana*, is absent from the family Fagaceae [19]. This gene is required for endosperm gene imprinting and seed viability in *A. thaliana* [26]. The gene expression of several DNA methyltransferases and
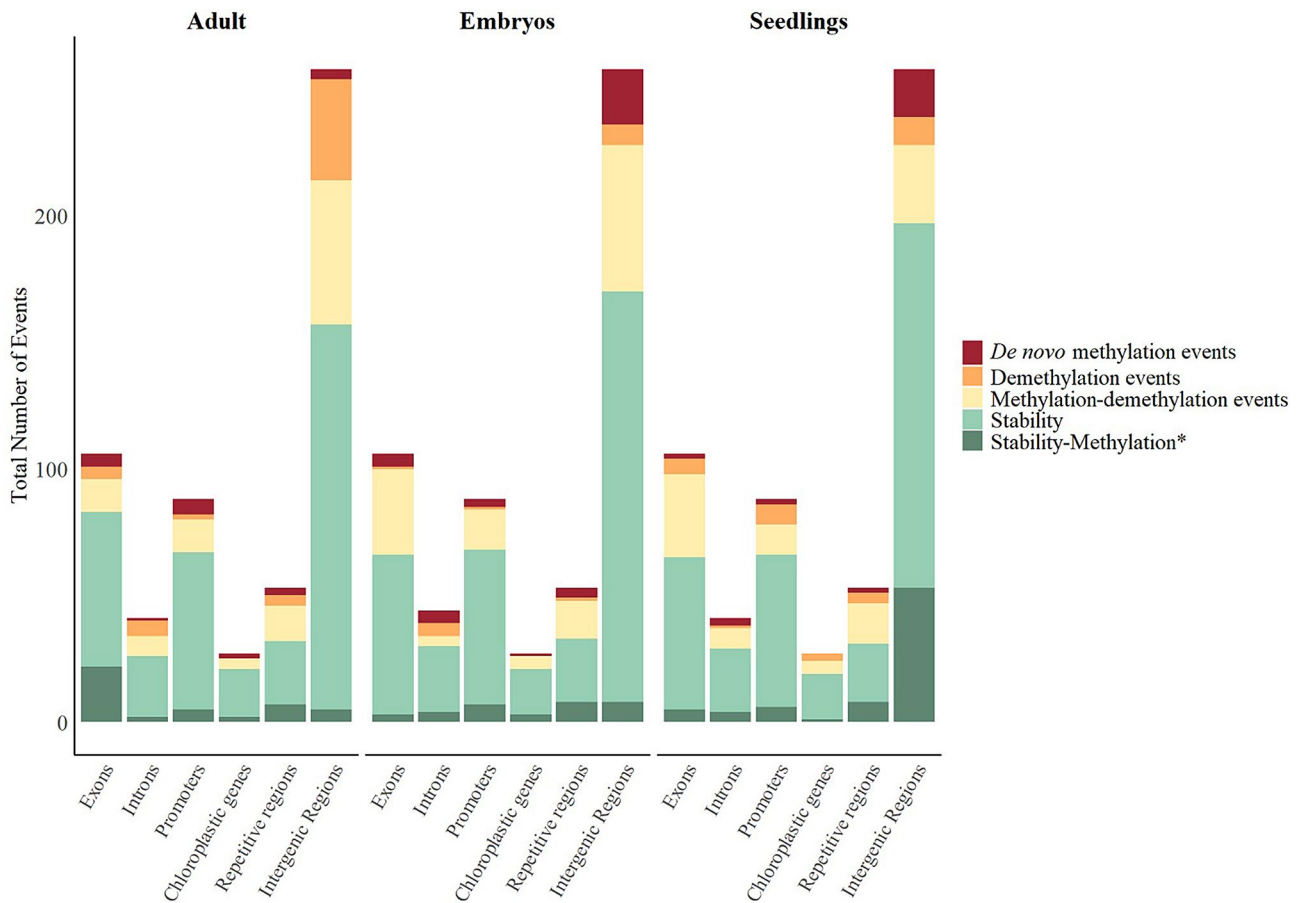
Fig. 4 Total number of stabilization and methylation events (demethylation, *de novo* methylation and methylation – demethylation) observed in adult plants, embryos, and seedlings of holm oak at the genomic sequence level

The asterisk indicates the number of methylated bands exhibiting stability

demethylases in all the developmental stages was determined. Enzyme-encoding genes exhibited an organ-dependent response. Although significant differences were observed in gene expression across the analysed genes, the majority exhibited low expression levels, with the remarkable exception of QiDRM2 and QiDRM3 [27]. Both genes were significantly upregulated 6-fold and 14-fold in embryos, respectively. Throughout plant embryonic development, significant alterations in gene regulation and DNA methylation dynamics occur to establish specific patterns of gene expression and determine the cellular identity of developing tissues [28]. *QiDRM2* and *QiDRM3* play pivotal roles in *de novo* DNA methylation, overseeing the establishment of new methylation patterns during embryogenesis [27, 28]. Thus, an increase in both enzymes in embryos could indicate increased DNA methylation activity at these early developmental stages, potentially influencing the regulation of gene expression and cell determination in embryonic tissues.

MSAP analysis was employed to study the global DNA methylation patterns in CCGG sequences across

different developmental stages, including an adult tree and its offspring at embryo and seedling levels. The findings revealed that epigenetic variability surpassed genetic variability, with Shannon's diversity index being 0.5793 for epigenetic variation and 0.3825 for genetic variation. This observation implies that epigenetic variability could play a significant role in the high molecular diversity reported for holm oak. DNA methylation is a key mechanism behind phenotypic plasticity, allowing trees to modify their growth and traits in response to environmental changes and identifying the epigenetic patterns that allow trees adapt and survive under changing conditions, which is crucial in the context of climate change. We identified various developmental stage-related DNA methylation profiles and found an increased proportion of total methylation in adult plants relative to seedlings. This result is consistent with those for other forest species such as *Pinus radiata* and *Eucalyptus urophylla* x *E. grandis* [29, 30]. Also, they suggest progressive developmental stage-related DNA methylation in holm oak and highlight its role in preserving genomic structure and stability as previously discovered in various eukaryotic
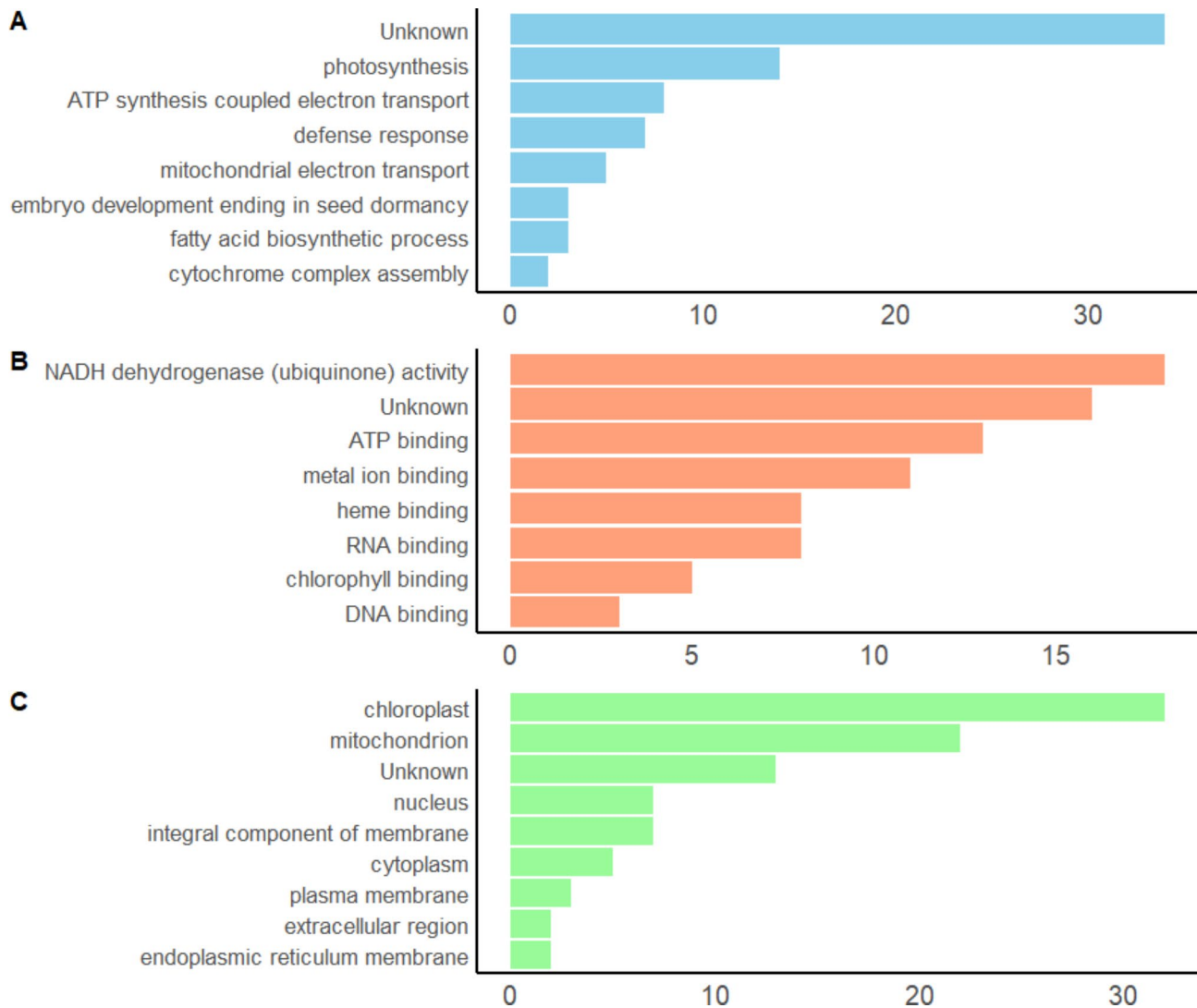
Fig. 5 Histogram of the first eight methylated nuclear genes represented in the three main GO categories: biological processes (**A**), molecular functions (**B**) and cellular components (**C**). All other genes assigned to each GO category are shown in Additional File 4

Table 4 Methylated genes identified in adult, embryos, and seedlings of holm oak. Type II, III and IV indicate hemimethylated, hypermethylated and fully methylated patterns, respectively

ID	Sequence	Annotation	Adult	Embryos	Seedlings
gene-14,558	exon2	Glycine-rich RNA-binding protein GRP1A	III	IV	II
gene-14,913	exon1	Glycine-rich protein 2	III	IV	II
gene-20,869	exon2	Cytochrome f	IV	III	II
gene-28,646	exon1	Metalloendoproteinase 2-MMP	III	IV	II
gene-33,063	exon2	Dynamin-related protein 3 A	II	IV	II
gene-37,848	exon1	E3 ubiquitin-protein ligase At1g12760	III	II	II
gene-14,188	promoter	Putative uncharacterized protein ART2	III	IV	II
gene-20,865	promoter	Photosystem I P700 chlorophyll a apoprotein A1	II	III	IV
gene-20,868	promoter	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	II	III	IV
gene-29,659	promoter	Uncharacterized protein ycf68	II	II	II

organisms [31]. Consistent with the results of Valledor et al. [32], demethylation was more marked in embryos, which are seemingly needed for transcriptional activation during germination. A decrease in plant methylation prior to germination has been widely discovered [33] in species such as *Q. suber*, where embryonic germination decreased global DNA methylation [10].

While significant progress has been made in understanding the overall epigenetic landscape of holm oak, the specific methylated genomic sequences within this species remain unidentified. This knowledge gap hinders our comprehension of the precise roles that DNA methylation plays in regulating gene expression, cellular processes, and the species' adaptability. Recently, the MSAP protocol was enhanced by deep PCR selective amplicon sequencing with Next-Generation Sequencing (NGS). A combination of MSAP and NGS has been used on crops such as *Triticum aestivum* and *Hordeum vulgare* [34–36], and also on forest species such as *E. grandis* and *Populus alba* [37, 38]. Although we used the recently released first draft of the holm oak genome, the mean proportion of mapped reads, 85.3%, was comparable to that previously reported by Chwialkowska et al. [35, 36]. Based on the genome size of holm oak (842.2 Mbp) [9], our MSAP-Seq analysis screened approximately 34.7% of all methylated genomic sequences. This value was obtained by excluding the second PCR step according to Guarino et al. [37] and using both *EcoRI*-A and *HpaII/MspI*-G primers during pre-selective amplification according to Rico et al. [13]. We compiled a dataset consisting of 573 genomic sequences, approximately 43.1% of which were methylated. Nuclear genes exhibited several methylated sites per gene in different samples, as previously described in Bewick and Schmitz [39]. Some (*Cytochrome f, NADH-ubiquinone oxidoreductase chain 4, Photosystem I P700 chlorophyll a apoprotein A2* and uncharacterized protein ycf68) had more than four methylated sites. As expected, more unmethylated genes (56.9%) than methylated genes were observed at the three developmental stages. The study revealed that intergenic regions and nuclear genes exhibited higher methylation levels compared to repetitive sequences and chloroplast genes. Similarly, exons displayed more pronounced methylation patterns than promoters. Gene-containing regions were thus enriched. *HpaII* and *MspI* only allow one to identify CCGG sites within the CG or CHG context. Therefore, the MSAP technique overestimates DNA methylation in these sequences. For this reason, we found an increased number of nuclear genes relative to repetitive sequences [36–38]. The most methylated regions in repetitive sequences were hAT, Gypsy and CACTA, which were heavily methylated as a defensive mechanism against TEs transposition [40, 41].

Although the occurrence of DNA methylation in chloroplast genomes is highly controversial, some studies have reported cytosine methylation to occur in the chloroplast [42, 43]. In this study, approximately 1.3% of all chloroplast sequences were in fact methylated. The degree of methylation varied across developmental stages, with seedlings exhibiting the highest level (50.0%), followed by embryos (42.1%) and adult plant (22.7%). Ohta et al. [44] also reported minimal methylation (0.5%) in amyloplast and chloroplast genomes in pea leaves and cotyledons.

GO classification allowed us to identify the target biological processes for the epigenetic machinery in adults, embryos and seedlings. The study revealed an increased proportion of methylated genes associated with photosynthesis, ATP synthesis coupled electron transport, and defence response categories. Unsurprisingly, DNA methylation has an impact on genes involved in these metabolic activities since adaptation to the highly diverse and dynamic conditions in natural habitats requires precise regulation [45]. Recently, Sela et al. [46] found embryonic photosynthesis to deeply influence future growth and development in *A. thaliana*. DNA methylation can regulate genome functioning and induce plant resistance and adaption to abiotic stresses as a result. The expression of defence-related genes is crucial for plant immunity against pathogens and contributes to their ability to adapt to changing conditions [47]. The relationship between DNA methylation and gene expression depends on several factors including genomic region, transposon activity and organ [48], so it is highly complex.

A total of 10 nuclear genes to be methylated in at least two developmental stages (Table 4). Chen et al. [49] previously found cytochrome f and Photosystem I P700 chlorophyll a apoprotein A1 (psaA) to be hypermethylated in photoperiod-thermo-sensitive sterile male rice. Also, Saito et al. [50] found supplementation with micronutrients such as vitamins and minerals to increase DNA methylation in the acetyl-CoA carboxylase alpha promoter. This enzyme has been proposed as an epigenetic regulator controlling lipid biosynthesis. Also, JUMONJI protein with E3 ubiquitin-protein ligase (JMJ24) has been deemed a regulator of transposable element silencing [51]. The ORTH family, which is considered an *E3 ubiquitin-protein ligase*, is also involved in the regulation of epigenetic mechanisms at heterochromatic and euchromatic loci in plants [52]. The putative *uncharacterized protein ycf68* exhibits differential methylation patterns under cold stress conditions in rice, as demonstrated by Pan et al. [53]. While the selected genes have not previously been associated with methylation, their potential involvement in plant growth and development makes them intriguing candidates for further investigation (dynamin-related protein 1 A and 2) [54, 55] regulating

various aspects of RNA metabolism (glycine-rich RNA binding proteins) [56], and in plant defence against microbial pathogens (metalloendoproteinase) [57].

Conclusion

It appears that this study is indeed the first comprehensive DNA methylation research focused on identifying DNA methyltransferases and demethylases, their expression levels, global methylation patterns, and methylated genomic sequences in three developmental stages (adult, embryo, and seedling) of holm oak. While some previous studies have investigated specific aspects of DNA methylation in this species, such as somatic embryogenesis and tissue-specific gene silencing, none have provided a thorough examination of the epigenetic landscape across multiple developmental stages. Our analysis exposed methylation mechanisms similar to those previously found in the family Fagaceae. Thirteen DNA methyltransferase genes and three demethylase genes were identified within the holm oak genome, exhibiting organ and developmental stage-dependent expression patterns. In this study, it was discovered that epigenetic variability was higher than genetic variability. This finding suggests that the observed molecular diversity in holm oak may, at least in part, be attributed to epigenetic changes. At global methylation pattern level, adult organs exhibited higher global DNA methylation levels than embryos and seedlings, with embryos having more demethylation events and seedlings more *de novo* methylation events. At genomic sequence level, consistent methylation levels across developmental stages that were especially high in intergenic regions and nuclear genes was observed. Also, several methylated genes linked to critical processes have been identified, including those of photosynthesis, ATP synthesis coupled electron transport and defence response. A set of 10 methylated genes involved in regulating transposable element silencing, lipid biosynthesis, growth and development, and response to biotic and abiotic stress are described. Thus, this study has expanded our current understanding of epigenetic regulation in holm oak development by providing novel insights into the roles of DNA methylation and the enzymes involved in this process.

In conclusion, although this work provides a preliminary analysis of global DNA methylation in holm oak, it also highlights the inherent limitations of current techniques. In particular, the MSAP method produces binary results (presence/absence of bands) indicating methylated or unmethylated sites without providing a quantitative view of the degree of methylation. Furthermore, MSAP does not allow quantification of average methylation levels in specific contexts such as CG, CHG and CHH. To overcome these limitations and gain a more comprehensive understanding of methylation patterns, we are currently developing whole genome bisulphite sequencing (WGBS). This approach will provide a more detailed and quantitative view of DNA methylation. Furthermore, the integration of WGBS with advanced sequencing technologies, such as PacBio or Nanopore sequencing, will enhance our ability to identify methylated genes and elucidate their functional implications, thus deepening our understanding of the epigenetic mechanisms underlying the adaptability and evolutionary potential of holm oak.

Materials and methods

Search for DNA methyltransferases and demethylases in holm oak

A total of 8 DNA methyltransferase and 4 DNA demethylase protein sequences of *Arabidopsis thaliana* were downloaded from TAIR (https://www.arabidopsis.org/, accessed 18 September 2022), and used as query to identify homologous sequences in the holm oak genome with the software BLAST+ [58]. Subsequently, homologous sequences in holm oak were BLASTed with the online-available tool NCBI Protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed 19 September 2022) to identify orthologues in *Q. lobata*, *Q. robur* and *Q. suber*, and BLAST+to obtain those from *Q. variabilis* [59]. Conserved domains were examined by searching Inter-Pro v. 3.1 (http://www.ebi.ac.uk/interpro/, accessed 20 September 2022) [60]. To identify the chromosomal locations in holm oak, the genome annotation file was utilized as a reference [9]. DNA methyltransferase and demethylase protein sequences from *A. thaliana*, *Q. ilex*, *Q. lobata*, *Q. robur*, *Q. suber* and *Q. variabilis* were aligned by using MUSCLE algorithm, and Maximum Likelihood method coupled with the Jones-Taylor-Thornton substitution model were used to construct a phylogenetic tree, which was further validated using a bootstrap method with 1000 replicates through the MEGA 11 software [61]. The trees were redesigned and visualized using the Interactive Tree of Life Tool (iTO) [62].

Plant material

The mature leaf samples and acorns used for this study were collected from a single *Quercus ilex* subsp. *ballota* [Desf.] Samp individual located in Aldea de Cuenca, Córdoba, Spain (38°12′08″ N, 5°19′16″ W) [9]. The voucher specimen was deposited at the Global Biodiversity Information Facility of University of Cordoba, Spain (deposition number. 36840-1) [63]. The MSAP assay was carried out using three replicates of the mature tree leaves (10 leaves per replicate), three replicates of acorn embryos at stage M9 [64] (30 embryos per replicate), and ten replicates of three-month-old seedling leaves (one seedling per replicate). The MSAP-seq and qRT-PCR analyses utilized the same three replicates for mature tree and

embryos as before, while for the seedlings, the three replicates with the highest number of MSAP fragments were selected. All plant material was immersed in liquid nitrogen and stored at −80 °C prior to DNA extraction.

DNA extraction and methylation-sensitive amplified polymorphism (MSAP) assay

Genomic DNA from adults, embryos and seedlings was extracted with CTAB [9]. All samples were treated with RNAse (10 mg/mL) (ThermoFisher scientific, Massachusetts, USA). DNA quality and concentration were assessed by using 1% agarose gel and a Qubit fluorometer (ThermoFisher scientific, Massachusetts, USA), respectively. MSAP analysis was performed by following previously reported protocols [21, 65]. Briefly, 1 µg genomic DNA was digested with two methylation-sensitive isoschizomers (*HpaII* and *MspI*; New England BioLabs, Massachusetts, USA) in combination with *EcoRI* (ThermoFisher scientific, Massachusetts, USA). The CpG methylation status of anonymous regions of the genome was determined by *HpaII* and *MspI* cut DNA sequences at the same tetra-nucleotide motif (5′-CCGG-3′) differing in sensitivity to cytosine methylation at the restriction site (Additional File 5). DNA was digested with *HpaII* and *MspI* at 37 °C for 3 h and redigested with *EcoRI* at the same temperature and time. Digested DNA was ligated with single strand oligonucleotide adaptors (Additional File 6) to prevent reconstruction of restriction sites. The ligation reaction was done with T4-DNA ligase (Invitrogen, Massachusetts, USA) according to the manufacturer's instructions. Pre-selective amplification was performed in a final volume of 20 µL, using 3 µL of ligated DNA previously diluted 4-fold with sterilized ultra-pure water (Additional File 6). Each PCR mixture contained $1 \times$ PCR buffer with $MgCl_2$, 0.25 mM dNTP, 0.17 µM primers and 0.02 U/µL DNA polymerase (Biotools S.A., Madrid, Spain). Those fragments 300–900 bp in length were verified in 2% agarose gel.

Selective amplification was accomplished by using four primer combinations (ECO adapter + ACG/ACC and HM adapter + TAG/TCC) and labelling with HEX fluorochrome according to Rico et al. [13] under the same PCR conditions as for pre-selective amplification (Additional File 6). The selective amplification conditions for PCR included 23 amplification cycles, denaturing at 94 °C for 0.5 min, annealing at 60 °C for 0.5 min, extension at 72 °C for 1 min and final extension at 72 °C for 10 min. MSAP-PCR fragments were identified by using ABI3130 XL (Applied Biosystems, California, USA) in combination with the GeneScan-400 HDROX size standard (ThermoFisher scientific, Massachusetts, USA) and analysed with GeneMapper v.4.1 (ThermoFisher scientific, Massachusetts, USA), for sizing. Following the sequencing process, the obtained fragments, ranging from 70 to 500 bp

(base pairs) in length, were analysed and organized into a binary character matrix. In this matrix, a value of 1 was assigned to represent the presence of a specific fragment, while a value of 0 indicated its absence. Four different types of methylation statuses were considered [35, 36], namely: type I (unmethylated loci) when a fragment was present in both *HpaII*/*EcoRI* and *MspI*/*EcoRI*; type II (hemimethylated loci) when present in *HpaII*/*EcoRI* only and thus indicating methylation of external cytosine in one strand; type III (hypermethylated loci) when present in *MspI*/*EcoRI* alone and thus confirming methylation of internal cytosine in at least one strand; and type IV (fully methylated loci) when no fragment was observed with either *HpaII*/*EcoRI* or *MspI*/*EcoRI*, which suggested methylation of both (internal and external) cytosines or the absence of digestion sites by effect of some mutation (Additional File 5). The appearance and disappearance of MSAP fragments was classified as a demethylation, *de novo* methylation or methylation−demethylation event [29]. The resulting binary matrix was analysed with the R package *msap* [66]. Developmental differentiation was determined via analyses of molecular variance (AMOVA) to estimate Phi-st as fixation index, using the package *pegas* with 10 000 permutations. The package *ade4* was used for principal coordinate analysis (PCoA) to expose epigenetic (MSL, methylation-susceptible loci) and genetic variability (NML, no methylated loci). Differences between genetic and epigenetic diversity were estimated in terms of Shannon's diversity index, and statistical significance in the differences between overall genetic and epigenetic diversities was assessed with the Wilcoxon rank sum test [67].

Sequencing of pre-selective amplification PCR products for MSAP-Seq

Methylation-Sensitive Amplification Polymorphism Sequencing (MSAP-Seq) allows to identify methylated genomic sequences by combination of traditional MSAP analysis and next-generation sequencing (NGS). The MSAP-Seq technique involves cleaving genomic DNA with restriction enzymes and then using NGS to directly sequence the resulting fragments of the pre-selective amplification. The sequencing procedure included repeating pre-selective amplification with a final volume of 50 µL and purification as recommended for the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). A genomic library was built for each sample by using an in-house developed protocol derived from [64]. Samples were fragmented by random digestion with the enzyme dsDNA Fragmentase (New England BioLabs, Massachusetts, USA), and universal Illumina-adapter was used to ligate the molecules at both ends. All libraries were dual-indexed for post-sequencing demultiplexing and pooled in equimolar amounts according to the quantitative

results provided by the Qubits DNA HS assay kit (ThermoFisher scientific, Massachusetts, USA). The pool was sequenced in a fraction of a NovaSeq PE150 flow cell from Illumina, Inc., aiming for a total of 15 gigabases output. The quality of each read was assessed with the software FastQC v. 0.11.9 [68]. Reads were filtered for the presence of HM-adapter by using Cutadapt v. 1.2 [69], and only those with adapter and a minimum length of 50 bp were retained for further analysis. Those remaining were aligned with the holm oak genome [9] by using bowtie2 v. 2.2.4 [70], and just reads with a MAPQ of at least 30 were retained by using Samtools v. 1.13 [71]. CCGG sites were identified by using the R package *msgbsR* [72] and annotated with *GenomicRanges* [73]. Sites were classified into nuclear genes (at exon, intron, and promoter positions), chloroplast genes, repetitive regions, or intergenic regions (without annotation in the genome used). A promoter was defined as a region located 2500 bp upstream the gene concerned and identified by using a Python script (https://github.com/EdoardoPiombo/promoter_extractor)). Intron annotations were obtained by using the *GenomicFeature* package in R [73], and intergenic regions with *complement* in BEDTools v. 2.30.00 [74]. The methylation status of each gene was identified and compared between developmental stages. The holm oak genome was used for Gene Ontology (GO) classification [9]. The appearance or disappearance of unmethylated and methylated regions at the three developmental stages examined was classified as either stabilization or methylation events [29].

Gene expression analysis

RNA was extracted from three biological replicates for each tissue following the protocol described by Echevarría-Zomeño et al. [75]. Briefly, RNA was extracted from adult and seedling leaves and embryos (50 mg of fresh tissue), previously macerated in liquid nitrogen, with the commercial kit InviTrap® Spin Plant RNA Mini (STRATEC Molecular GmbH, Berlin, Germany) according to the manufacturer's instructions. To remove any possible residual DNA, the samples were digested using DNase I from ThermoFisher according to the manufacturer's instructions. The quality of the extracted RNA was tested on 1% agarose gel, using Redsafe as an intercalating agent (Intron Biotechnology, Seongnam, Korea). The quantity and integrity of RNA was determined by electrophoresis using the Agilent 2100 Bioanalyser (Agilent Technologies, California, USA). Samples with an RNA Integration Number (RIN) value≥7 and A260:A280 ratio around 2.0, were selected. Once the quality and quantity of the RNA was verified, complementary DNA (cDNA) from 1 µg RNA was synthesised by a Reliance Select cDNA Synthesis Kit (Bio-Rad, California, United States) according to the manufacturer's instructions and used in

the qRT-PCR reaction in an iCycler system using SYBR-Green (Bio-Rad, California, United States). To obtain amplified DNA fragments ranging from 101 to 134 nucleotides, primers were designed using the Primer3-Plus web-based program (https://primer3plus.com), based on the sequences identified in Rey et al. [9]. The reaction mixtures comprised 2x Sso Advanced Universal SYBR® Green supermix (Bio-Rad, Hercules, USA), with each forward and reverse primer at a concentration of 0.2 µM, 50 ng of cDNA, and nuclease-free water adjusted to reach a total reaction volume of 10 µl. A two-step amplification protocol was employed, beginning with an initial denaturation step at 95 °C for 30 s, followed by 45 cycles of denaturation at 95 °C for 15 s and annealing/polymerization at 60 °C for 60 s. Subsequently, a melting analysis ranging from 60 °C to 97 °C at a rate of 0.1 °C/s was conducted to validate the specificity of the target amplicon. The expression levels of all analysed genes were determined in each reaction using the threshold cycle (Ct value). Relative differential expression was determined by using the $2^{-\Delta\Delta Ct}$ method [76]. Elongation factor 1 alpha and AAA-type ATPase family protein, which exhibit constitutive expression under tested experimental conditions [77], were used as internal references. Each reaction was performed in duplicate. The primers used are listed in Additional File 7.

Statistical analysis

Statistical analysis of the gene expression data was conducted using STATISTIX 10.0 software (Analytical Software, Florida, USA). Analysis of variance (ANOVA) was performed based on a completely randomized design, and mean separations were determined using the least significant difference (LSD) test with a significance level of 0.05.

Abbreviations

CMTs	Chromomethylases
DME	Demeter
DML2 and 3	Demeter-like 2 and 3
DNMT2	DNA Methyltransferase Homologue 2
DRMs	Domains Rearranged Methyltransferases
GO	Gene Ontology
MET1	Methyltransferase 1
MSAP	Methylation-sensitive amplified polymorphism
MSAP-Seq	Methylated genomic sequences were identified via MSAP sequencing
MSL	Methylation-susceptible loci
NGS	Next-Generation Sequencing
NML	Non-methylated loci
qRT-PCR	Quantitative real-time PCR
ROS1	Repressor of Silencing 1

Supplementary Information

Additional Figure 1: MSAP band patterns established by using *Hpall/EcoRI* and *Mspl/EcoRI* and the correspondence methylation status of

CCGG sites. Black squares denote methylated cytosines and +/− whether a sequence was cut/not cut, or a band was/was not present. **Additional Table 1:** List of primers used for analysis of relative expression of DNA methyltransferases and demethylases. **Additional Table 2**: List of primers used. ECO denotes digestion with *EcoRI* and HM digestion with *HpaII* or *MspI*. **Additional Table 3**: Gene ontology classification of DNA methyltransferases and demethylases in holm oak. **Additional Table 4**: Percent identity in DNA methyltransferase and demethylase protein sequences between *Q. ilex* and *Q. lobata*, *Q. robur*, *Q. suber* and *Q. variabilis*. **Additional Table 5**: Global DNA patterns and levels for holm oak adults, embryos and seedlings

Additional Table 6: List of genomic sequences identified in holm oak adults, embryos and seedlings including methylation type (I, II, III and IV denote non-methylated, hemimethylated, hypermethylated and fully methylated genes, respectively). A GO analysis (biological process, molecular function, and cellular component) was done with the nuclear genes

Author contributions
M.-D. Rey and J.V. Jorrín-Novo conceived and designed the experiments. M. Labella-Ortega, C. Martín, L. Valledor, S. Castiglione, M.-Á. Castillejo and M.-D. Rey conducted the tests. M. Labella-Ortega, L. Valledor and M.-D. Rey analysed the sequencing data. M. Labella-Ortega and M.-D. Rey wrote the original draft. All authors have read and approved the final version of the manuscript.

Data availability
The datasets generated and/or analysed during the current study are available in the NCBI SRA repository, with accession number SAMN41532855, SAMN41532856, SAMN41532857, SAMN41532858, SAMN41532859, SAMN41532860 (https://www.ncbi.nlm.nih.gov/bioproject/687489). All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Ethical consideration
The plant material was identified by Prof. Jesús V. Jorrín Novo. All the methods involving plant material were performed in accordance with national or institutional guidelines.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Agroforestry and Plant Biochemistry, Proteomics and Systems Biology, Department of Biochemistry and Molecular Biology, University of Cordoba, UCO-CeiA3, Cordoba 14014, Spain
[2]Department of Biotechnology–Plant Biology, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid (UPM), Ciudad Universitaria s/n, Madrid 28040, Spain
[3]Plant Physiology Lab, Department of Organisms and Systems Biology and University Institute of Biotechnology (IUBA), University of Oviedo, Cat. Rodrigo Uría s/n, Oviedo 33006, Spain
[4]Department of Chemistry and Biology "A. Zambelli", University of Salerno, via Giovanni Paolo II 132, Fisciano, Salerno 84084, Italy

References

1. Pulido FJ, Díaz M, Hidalgo de Trucios SJ. Size structure and regeneration of Spanish holm oak *Quercus ilex* forests and dehesas: effects of agroforestry use on their long-term sustainability. Ecol Manag. 2001;146:1–13.
2. Maldonado-Alconada AM, Castillejo MA, Rey MD, Labella-Ortega M, Tienda-Parrilla M, Hernández-Lao T, Honrubia-Gómez I, Ramírez-García J, Guerrero-Sanchez VM, López-Hidalgo C, et al. Multiomics Molecular Research into the recalcitrant and Orphan *Quercus ilex* Tree species: why, what for, and how. Int J Mol Sci. 2002;23:9980.
3. Quero JL, Villar R, Marañón T, Zamora R. Interactions of drought and shade effects on seedlings of four *Quercus* species: physiological and structural leaf responses. New Phytol. 2006;170:819–33.
4. San-Eufrasio B, Sánchez-Lucas R, López-Hidalgo C, Guerrero-Sánchez VM, Castillejo MA, Maldonado-Alconada AM, Jorrín-Novo JV, Rey MD. Responses and differences in tolerance to water shortage under climatic dryness conditions in seedlings from *Quercus* spp and andalusian *Q ilex* populations. Forests. 2020;11.
5. López-Hidalgo C, Trigueros N, Menéndez MR, Jorrin-Novo JV. Phytochemical composition and variability in *Quercus ilex* Acorn morphotypes as determined by NIRS and MS-based approaches. Food Chem. 2021;338:127803.
6. Martínez MT, San José MV, Vieitez AM, Cernadas MJ, Ballester A, Corredoira E. Propagation of mature *Quercus ilex* L (holm oak) trees by somatic embryogenesis. Plant Cell Tissue Organ Cult. 2017;131:321–33.
7. Corredoira E, Merkle SA, Martínez MT, Toribio M, Canhoto JM, Correia SI, Ballester A, Vieitez AM. Non-zygotic embryogenesis in hardwood species. Crit Rev Plant Sci. 2019;38:29–97.
8. Pavese V, Moglia A, Milani AM, Marino LA, Martinez MT, Torello Marinoni D, Botta R, Corredoira E. Advances in *Quercus ilex* L breeding: the CRISPR/Cas9 technology via ribonucleoproteins. Front Plant Sci. 2024;15.
9. Rey MD, Labella-Ortega M, Guerrero-Sánchez VM, Carleial R, Castillejo MA, Ruggieri V, Jorrín-Novo JV. A first draft genome of holm oak (*Quercus ilex* subsp *ballota*), the most representative species of the Mediterranean forest and the Spanish agrosylvopastoral ecosystem *dehesa*. Front Mol Biosci. 2023;10.
10. Pérez M, Viejo M, LaCuesta M, Toorop P, Cañal MJ. Epigenetic and hormonal profile during maturation of *Quercus suber* L somatic embryos. J Plant Physiol. 2015;173:51–61.
11. Michalak M, Plitta-Michalak MP, Naskręt-Barciszewska M, Barciszewski J, Bujarska-Borkowska B, Chmielarz P. Global 5-methylcytosine alterations in DNA during ageing of *Quercus robur* seeds. Ann Bot. 2015;116:369–76.
12. Sork VL, Cokus SJ, Fitz-Gibbon ST, Zimin AV, Puiu D, Garcia JA, Gugger PF, Henriquez CL, Zhen Y, Lohmueller KE et al. High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. Nat Commun. 2022;13.
13. Rico L, Ogaya R, Barbeta A, Peñuelas J. Changes in DNA methylation fingerprint of *Quercus ilex* trees in response to experimental field drought simulating projected climate change. Plant Biol. 2014;16:419–27.
14. Lucibelli F, Valoroso MC, Aceto S, Plant DNA, Methylation. An epigenetic mark in development, environmental interactions, and evolution. Int J Mol Sci. 2022;23.
15. Kumar S. Epigenomics of plant responses to environmental stress. Epigenomes. 2018;2.
16. Yamaguchi N. Editorial: epigenetics in Plant Development. Front Plant Sci. 2022;13.
17. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11:204–20.
18. Chan SWL, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nat Rev Genet. 2005;6:351–60.
19. Alves S, Braga Â, Parreira D, Alhinho AT, Silva H, Ramos MJN, Costa MMR, Morais-Cecílio L. Genome-wide identification, phylogeny, and gene duplication of the epigenetic regulators in Fagaceae. Physiol Plant. 2022;174:e13788.
20. Thiebaut F, Hemerly AS, Ferreira PCG. A role for epigenetic regulation in the adaptation and stress responses of non-model plants. Front Plant Sci. 2019;10:246.
21. Reyna-López GE, Simpson J, Ruiz-Herrera J. Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Mol Gen Genet. 1997;253:703–10.

22. Agius DR, Kapazoglou A, Avramidou E, Baranek M, Carneros E, Caro E, Catiglione S, Cicatelli A, Radanovic A, Ebejer JP, Gackowski D, Guarino F, Gylyás A, Hidvégi N, Hoenicka H, Inácio V, Johannes F, Karalija E, Lieberman-Lazarovich M, Martinelli F, Maury S, Mladenov V, Morais-Cecílio L, Pecnika A, Testillano PS, Todorov D, Valledor L. Vassileva, V. exploring the crop epigenome: a comparison of DNA methylation profiling techniques. Front Plant Sci. 2023;14:1181039.

23. Kumar S, Mohapatra T. Dynamics of DNA methylation and its functions in plant growth and development. Front Plant Sci. 2021;12:596236.

24. Lucibelli F, Valoroso MC, Aceto S. Plant DNA methylation: an epigenetic mark in development, environmental interactions, and evolution. Int J Mol Sci. 2022;23:8299.

25. Silva HG, Sobral RS, Magalhães AP, Morais-Cecílio L, Costa MMR. Genome-wide identification of epigenetic regulators in *Quercus suber* L. Int J Mol Sci. 2022;21:3783.

26. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell. 2002;110:33–42.

27. Ashapkin VV, Kutueva LI, Vanyushin BF. Plant DNA methyltransferase genes: multiplicity, expression, methylation patterns. Biochem (Mosc). 2016;81:141–51.

28. Markulin L, Škiljaica A, Tokić M, Jagić M, Vuk T, Bauer N. Leljak Levanić D. taking the wheel – *de novo* DNA methylation as a driving force of plant embryonic development. Front Plant Sci. 2021;12.

29. Valledor L, Meijón M, Hasbún R, Cañal MJ, Rodríguez R. Variations in DNA methylation, acetylated histone H4, and methylated histone H3 during *Pinus radiata* needle maturation in relation to the loss of in vitro organogenic capability. J Plant Physiol. 2010;167:351–7.

30. Mankessi F, Saya AR, Favreau B, Doulbeau S, Conéjéro G, Lartaud M, Verdeil JL, Monteuuis O. Variations of DNA methylation in *Eucalyptus urophylla×Eucalyptus grandis* shoot tips and apical meristems of different physiological ages. Physiol Plant. 2011;143:178–87.

31. Bartels A, Han Q, Nair P, Stacey L, Gaynier H, Mosley M, Huang QQ, Pearson JK, Hsieh TF, An YQC et al. Dynamic DNA methylation in plant growth and development. Int J Mol Sci. 2018;19.

32. Valledor L, Hasbún R, Meijón M, Rodríguez JL, Santamaría E, Viejo M, Berdasco M, Feito I, Fraga MF, Cañal MJ, et al. Involvement of DNA methylation in tree development and micropropagation. Plant Cell Tissue Organ Cult. 2007;91:75–86.

33. Meng FR, Li YC, Yin J, Liu H, Chen XY, Ni ZF, Sun QX. Analysis of DNA methylation during the germination of wheat seeds. Biol Plant. 2012;56:269–75.

34. Baránek M, Čechová J, Kovacs T, Eichmeier A, Wang S, Raddová J, Nečas T, Ye X. Use of combined MSAP and NGS techniques to identify differentially methylated regions in somaclones: a case study of two stable somatic wheat mutants. PLoS ONE. 2016;11:e0165749.

35. Chwialkowska K, Nowakowska U, Mroziewicz A, Szarejko I, Kwasniewski M. Water-deficiency conditions differently modulate the methylome of roots and leaves in barley (*Hordeum vulgare* L). J Exp Bot. 2016;67:1109–21.

36. Chwialkowska K, Korotko U, Kosinska J, Szarejko I, Kwasniewski M. Methylation sensitive amplification polymorphism sequencing (MSAP-Seq)-A method for high-throughput analysis of differentially methylated CCGG sites in plants with large genomes. Front Plant Sci. 2017;8:2056.

37. Guarino F, Heinze B, Castiglione S, Cicatelli A. Epigenetic analysis through MSAP-NGS coupled technology: the case study of white poplar monoclonal populations/stands. Int J Mol Sci. 2020;21:7393.

38. Pereira WJ, Pappas MCR, Grattapaglia D, Pappas GJ. A cost-effective approach to DNA methylation detection by Methyl Sensitive DArT sequencing. PLoS ONE. 2020;15:e0233800.

39. Bewick AJ, Schmitz RJ. Gene body DNA methylation in plants. Curr Opin Plant Biol. 2017;36:103–10.

40. Kato M, Takashima K, Kakutani T. Epigenetic Control of CACTA Transposon Mobility in *Arabidopsis thaliana*. Genetics. 2004;168:961–9.

41. Skipper KA, Andersen PR, Sharma N, Mikkelsen JG. DNA transposon-based gene vehicles - scenes from an evolutionary drive. J Biomed Sci. 2013;20:92.

42. Huang LC, Hsiao LJ, Pu SY, Kuo CI, Huang BL, Tseng TC, Huang HJ, Chen YT. DNA methylation and genome rearrangement characteristics of phase change in cultured shoots of *Sequoia sempervirens*. Physiol Plant. 2012;145:360–8.

43. Wang L, Leister D, Kleine T. Chloroplast development and genomes uncoupled signaling are independent of the RNA-directed DNA methylation pathway. Sci Rep. 2020;10:15412.

44. Ohta N, Sato N, Kawano S, Kuroiwa T. Methylation of DNA in the chloroplasts and amyloplasts of the pea, *Pisum sativum*. Plant Sci. 1991;78:33–42.

45. Sello S, MeneghessoA, Alboresi A, Baldan B, Morosinotto T. Plant biodiversity and regulation of photosynthesis in the natural environment. Planta. 2019;249:1217–28.

46. Sela A, Piskurewicz U, Megies C, Mène-Saffrané L, Finazzi G, Lopez-Molina L. Embryonic photosynthesis affects post-germination plant growth. Plant Physiol. 2020;182:2166–81.

47. Xie SS, Duan CG. Epigenetic regulation of plant immunity: from chromatin codes to plant disease resistance. Abiotech. 2023;4:124–39.

48. Arora H, Singh RK, Sharma S, Sharma N, Panchal A, Das T, Prasad A, Prasad M. DNA methylation dynamics in response to abiotic and pathogen stress in plants. Plant Cell Rep. 2002;41:1931–44.

49. Chen X, Hu J, Zhang H, Ding Y. DNA methylation changes in photoperiod-thermo-sensitive male sterile rice PA64S under two different conditions. Gene. 2014;537:143–8.

50. Saito T, Whatmore P, Taylor PF, Fernandes JMO, Adam AC, Tocher DR, Espe M, Skjærven KH. Micronutrient supplementation affects transcriptional and epigenetic regulation of lipid metabolism in a dose-dependent manner. Epigenetics. 2021;16:1217–34.

51. Kabelitz T, Brzezinka K, Friedrich T, Górka M, Graf A, Kappel C, Bäurle I. A JUMONJI protein with E3 ligase and histone H3 binding activities affects transposon silencing in *Arabidopsis*. Plant Physiol. 2016;171:344–58.

52. Kraft E, Bostick M, Jacobsen SE, Callis J. ORTH/VIM proteins that regulate DNA methylation are functional ubiquitin E3 ligases. Plant J Cell Mol Biol. 2008;56:704–15.

53. Pan Y, Wang W, Zhao X, Zhu L, Fu B, Li Z. DNA methylation alterations of rice in response to cold stress. Plant Omics J. 2011;4:364–9.

54. Taylor NG. A role for Arabidopsis dynamin related proteins DRP2A/B in endocytosis; DRP2 function is essential for plant growth. Plant Mol Biol. 2011;76:117–29.

55. Ekanayake G, Smith JM, Jones KB, Stiers HM, Robinson SJ, LaMontagne ED, Kostos PH, Cornish PV, Bednarek SY, Heese A. DYNAMIN-RELATED PROTEIN DRP1A functions with DRP2B in plant growth, flg22-immune responses, and endocytosis. Plant Physiol. 2021;185:1986–2002.

56. Cheng K, Zhang C, Lu Y, Li J, Tang H, Ma L, Zhu H. The Glycine-rich RNA-Binding protein is a vital Post-transcriptional Regulator in crops. Plants. 2023;12.

57. Zhao D, Hamilton JP, Pham GM, Crisovan E, Wiegert-Rininger K, Vaillancourt B, DellaPenna D, Buell CR. *De novo* genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. GigaScience. 2017;6:gix065.

58. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

59. Han B, Wang L, Xian Y, Xie XM, Li WQ, Zhao Y, Zhang RG, Qin X, Li DZ, Jia KH. A chromosome-level genome assembly of the Chinese cork oak (*Quercus variabilis*). Front Plant Sci. 2022;13.

60. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. InterPro: the integrative protein signature database. Nucleic Acids Res. 2009;37:D211–5.

61. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38:3022–7.

62. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;gkae268.

63. Devesa Alcaraz JA, Martínez Sagarra G. Herbario COFC De La Universidad De Córdoba: colección general de Plantas vasculares. Department of Plant Biology (Botany), University of Cordoba; 2021.

64. Sghaier-Hammami B, Hammami SBM, Baazaoui N, Gómez-Díaz C, Jorrín-Novo JV. Dissecting the seed maturation and germination processes in the non-orthodox *Quercus ilex* species based on protein signatures as revealed by 2-DE coupled to MALDI-TOF/TOF proteomics strategy. Int J Mol Sci. 2020;21.

65. Xiong LZ, Xu CG, Saghai Maroof MA, Zhang Q. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. Mol Gen Genet. 1999;261:439–46.

66. Pérez-Figueroa A. Msap: a tool for the statistical analysis of methylation-sensitive amplified polymorphism data. Mol Ecol Resour. 2013;13:522–7.

67. Carøe C, Bohmann K, Tagsteady. A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. Mol Ecol Resour. 2020;20:1620–31.

68. Andrews S, FastQC. A Quality Control Tool for High Throughput Sequence Data [Online] (2010) Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

69. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17.
70. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
71. Li H, Handsake B, Wysokerv A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G. Durbin R, and 1000 Genome Project DatabProcessing Subgroup. The sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
72. Mayne BT, Leemaqz SY, Buckberry S, Rodriguez Lopez CM, Roberts CT, Bianco-Miotto T, Breen J. msgbsR: an R package for analysing methylation-sensitive restriction enzyme sequencing data. Sci Rep. 2018;8:2190.
73. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLOS Comput Biol. 2013;9:e1003118.
74. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
75. Echevarría-Zomeño S, Abril N, Ruiz-Laguna J, Jorrín-Novo J, Maldonado-Alconada AM. Simple, rapid and reliable methods to obtain high quality RNA and genomic DNA from Quercus ilex L. leaves suitable for molecular biology studies. Acta Physiol. Plant. 2012;34:793–805.
76. Pedersen S, Amtssygehus A. Multiplex relative gene expression analysis by real-time RT-PCR using the iCycler iQTM detection system BioRad 2001;107:10–11.
77. Guerrero-Sánchez VM, Castillejo MÁ, López-Hidalgo C, Maldonado-Alconada AM, Jorrín-Novo JV, Rey MD. Changes in the transcript and protein profiles of *Quercus ilex* seedlings in response to drought stress. J Proteom. 2021;243:104263.

Publisher's note