



Article

# Genome-Wide Identification of Epigenetic Regulators in *Quercus suber* L.

Helena G. Silva <sup>1</sup>, Rómulo S. Sobral <sup>1</sup>, Alexandre P. Magalhães <sup>2</sup>, Leonor Morais-Cecílio <sup>3,\*</sup>  
and M. Manuela R. Costa <sup>1,\*</sup>

<sup>1</sup> Biosystems and Integrative Sciences Institute, Plant Functional Biology Centre, University of Minho, 4710-057 Braga, Portugal; helenasgsilva@gmail.com (H.G.S.); romulosobral@bio.uminho.pt (R.S.S.)

<sup>2</sup> Plant Systems Biology, Technische Universität München, 85354 Freising, Germany; alexandre.magalhaes@wzw.tum.de

<sup>3</sup> Linking Landscape, Environment, Agriculture and Food, Institute of Agronomy, University of Lisbon, Tapada da Ajuda, 1349-017 Lisbon, Portugal

\* Correspondence: lmorais@isa.ulisboa.pt (L.M.-C.); manuela.costa@bio.uminho.pt (M.M.R.C.)

Received: 11 April 2020; Accepted: 25 May 2020; Published: 27 May 2020



**Abstract:** Modifications of DNA and histones, including methylation and acetylation, are critical for the epigenetic regulation of gene expression during plant development, particularly during environmental adaptation processes. However, information on the enzymes catalyzing all these modifications in trees, such as *Quercus suber* L., is still not available. In this study, eight DNA methyltransferases (DNA Mtases) and three DNA demethylases (DDMEs) were identified in *Q. suber*. Histone modifiers involved in methylation (35), demethylation (26), acetylation (8), and deacetylation (22) were also identified in *Q. suber*. In silico analysis showed that some *Q. suber* DNA Mtases, DDMEs and histone modifiers have the typical domains found in the plant model *Arabidopsis*, which might suggest a conserved functional role. Additional phylogenetic analyses of the DNA and histone modifier proteins were performed using several plant species homologs, enabling the classification of the *Q. suber* proteins. A link between the expression levels of each gene in different *Q. suber* tissues (buds, flowers, acorns, embryos, cork, and roots) with the functions already known for their closest homologs in other species was also established. Therefore, the data generated here will be important for future studies exploring the role of epigenetic regulators in this economically important species.

**Keywords:** epigenetic modifiers; cork oak; *Quercus*; DNA methylation/demethylation; Histone methylation/demethylation; Histone acetylation/deacetylation

## 1. Introduction

The transcriptional activation and/or silencing of key genes is crucial for essential biological processes. This is achieved by the activity of transcription factors, but also by DNA and histone modification enzymes that alter chromatin conformation (euchromatin or heterochromatin), in a process classically identified as epigenetic regulation. This control at the chromatin level is very dynamic and can be reversed, which means that inactive condensed regions of chromatin may be easily decondensed allowing for subsequent gene expression.

The best-known DNA modification that regulates gene expression is cytosine methylation that is catalyzed by several DNA methyltransferases. In plants, DNA methyltransferases methylate DNA in the carbon-5 of cytosine (5mC) in distinct sequence contexts such as CG and CHG, but also in the asymmetrical context CHH, in which H is any nucleotide but G [1]. While 5mC in gene promoter regions is often associated with gene silencing, in coding regions, although controversial, 5mC has been correlated with active gene transcription [2,3]. Methylation can be achieved by two mechanisms:

*de novo* and maintenance of the methylation status. *De novo* methylation implies the methylation of previously unmethylated cytosine residues without a “template”, resulting in new methylation patterns. The maintenance of the methylation status is the process by which pre-existing methylated residues serve as a template to the replication-coupled DNA methylation [4]. The methylation process is controlled by four families of DNA methyltransferases (Mtases), which are classified by their linear domain arrangement: Methyltransferase (MET), Chromomethylase (CMT), Domains Rearranged Methyltransferase (DRM) and DNA Methyltransferase homolog 2 (DNMT2). MET maintains CG methylation of heterochromatic regions enriched with transposable elements (TEs) and repeats, and in intragenic regions [5,6]. CMT and DRM mediate CHG (CMT3) and CHH (CMT2 and DRM2) methylation [7,8]. While the majority of the methyltransferases maintain the DNA methylation status, only DRM2 establishes *de novo* methylation in all three motif contexts [9]. *De novo* DNA methylation is triggered by the activation of RNA-directed DNA methylation (RdDM), in which small interfering RNAs (siRNAs) are generated, ending with a downstream methylation targeting phase that is mediated by DRM2 [7]. CMT family genes can initiate *de novo* DNA methylation at sites with specific histone modifications and target transposons, as well as heterochromatin during replication [8]. The role of the DNMT2 family in plant DNA methylation remains unclear.

DNA methylation can be reversed passively during DNA replication or actively through base excision repair mechanisms initiated by DNA glycosylases [9–12]. Functional studies of the DNA glycosylases REPRESSOR OF SILENCING 1 (ROS1) [9], and DEMETER (DME) [10,11] suggested roles in the inhibition of gene silencing by demethylating DNA. DEMETER-LIKE 2 and 3 (DML2 and 3) are also DNA demethylases (DDME) that, instead of reactivating gene expression, prevent the deleterious effect resulting from the accumulation of DNA methylation at or near some genes [12].

Epigenetic regulation is also accomplished by the acetylation and methylation of histones. Histone modifier enzymes add or remove chemical groups, particularly in lysine residues, in the amino-terminal tails of histone 3 (H3) and 4 (H4). The deposition of acetylation marks is mediated by histone acetyltransferases (HATs), a mark that can be removed by histone deacetylases (HDACs). Histone acetylation is frequently related to increased gene expression while deacetylation is associated with transcriptional repression [13]. HATs are grouped in four families: GNAT (Gcn5-related N-acetyltransferase); MYST (MOZ, Ybf2/Sas3, Sas2, and Tip60), CBP (cAMP-responsive element-binding protein-binding protein) and TAFII250 (TATA-binding protein-associated factor family) based on sequence homology and mode of action [14] while HDACs are divided in three families: RPD3/HDA1 (Reduced potassium dependency 3/Histone deacetylase 1), HD2 (Histone deacetylase 2) and SIR2 (Silent information regulator 2) based on sequence similarity and cofactor dependency [15]. Like acetylation, histone methylation catalyzed by histone methyltransferases (HMTs) is also a reversible process involving histone demethylases (HDMTs). HMTs are mainly divided in five classes, based on the amino acid conservation of the SET (Suppressor of variegation, Enhancer of zeste, and Trithorax in *Drosophila*) domain [16]. Each class of SET domain proteins has specificity for a particular histone residue. For example, while class I proteins transfer three methyl groups to lysine 27 of H3 (H3K27me<sub>3</sub>), which is correlated with gene silencing, class III proteins are responsible for the deposition of methyl groups associated with active gene expression like mono-, di- and trimethylation at the lysine 4 of H3 (H3K4me<sub>1/2/3</sub>) site. Histone methylation marks are removed by two types of HDMTs, histone KDM1/LSD1 (Lysine demethylase 1) and JmjC (Jumonji C) domain-containing proteins [17,18], depending on a cofactor required to act [19].

Plant development is controlled by endogenous but also by environmental cues, which requires a tight control of gene regulation. Adverse environmental factors, such as severe temperature, drought, saline, and biotic stress can alter the longevity of many species, especially of long living species such as *Quercus suber* L., able to live over the span of 200 years. *Quercus suber* L. belongs to the Fagaceae family and is one of the most economically and ecologically important forest species in the Mediterranean basin, being the dominant tree of the Portugal oak woodlands – *Montados*. The cork of this species is

harvested every nine years making this oak a very important natural resource of the semi-arid regions of Southern Europe.

To provide for their unique attributes, both their longevity and annual developmental transitions, *Q. suber* trees activate highly dynamic regulatory mechanisms in order to adapt and survive in a variable climate. So, the identification and characterization of chromatin modifier proteins is of great importance. In *Q. suber*, mutant lines are impossible to obtain, due to its long life cycle and its recalcitrant behavior to transformation; however, the relevance of the genes codifying for these enzymes could be studied by gene expression analysis.

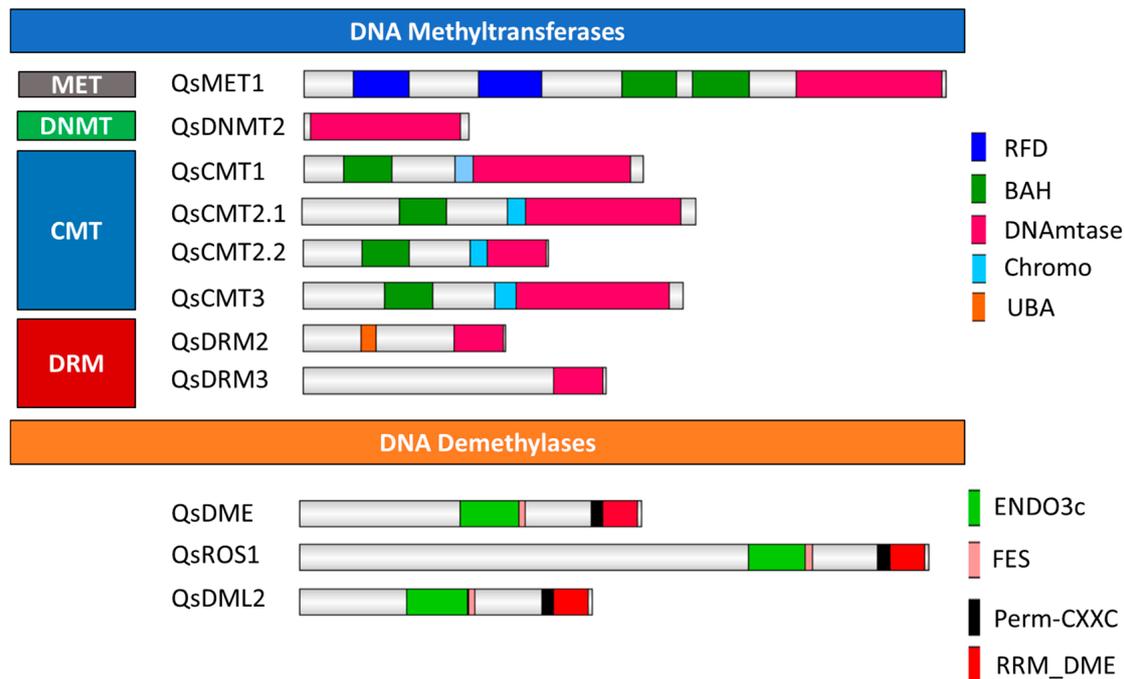
In this study, the genome-wide identification of DNA and histone modifier enzymes in *Q. suber* was obtained by data mining available genomes and transcriptomes through blast analysis. The phylogeny and composition of each gene family were identified and a brief overview of the transcriptional regulation of these newly identified genes in *Q. suber* was obtained from different tissues (buds, flowers, acorns, embryos, roots and cork) using publicly available data. Thus, this work provides a valuable source of information on genes potentially involved in epigenetic regulation and should greatly facilitate further studies in *Q. suber*.

## 2. Results

Epigenetic regulation plays an important role in the control of gene expression during plant growth and development. Although, significant advances in this field have been made in the model species *A. thaliana*, less has been reported in perennial tree species. In *Q. suber*, there is no knowledge about the genes that encode chromatin regulators. Therefore, in order to identify histone and DNA modifier enzymes with a potential role in epigenetic mark deposition, sequence-based searches and phylogenetic analysis were performed.

### 2.1. Identification and Classification of *Quercus suber* DNA Methyltransferases

Eight DNA Mtases proteins were identified based on a whole sequence similarity search and a C-5 cytosine methyltransferase (IPR001525) domain search. The DNA Mtases of each subfamily in *Q. suber* share the conserved catalytic domain with homologues from other plant species and harbor their characteristic small motifs (Figures S1–S4) [20–22]. A single *Q. suber* gene was identified within the MET subfamily (*QsMET1*), as well as in the DNMT2 subfamily (*QsDNMT2*). *QsMET1* contains the expected domains: replication foci domain RFD (PF12047), bromo adjacent homology (BAH) domain (PF01426) and DNA methyltransferase domain (PF00145) (Figure 1); a domain topology observed in the closest homologs of *Arabidopsis*, globe artichoke, strawberry, soybean, carrot, tomato, tobacco, pea, poplar, peach and rice [22–28]. *QsDNMT2* contains a single DNA mtase domain, as previously identified in the *A. thaliana* counterpart (Figure 1) [22]. In the CMT subfamily, four homolog proteins were identified, two different putative CMT2 proteins (named in this work as *QsCMT2.1* and *QsCMT2.2*), *QsCMT1* and *QsCMT3*. Members of the CMT subfamily have the C-terminal DNA Mtase domain, a chromodomain (CHR) (PF00385), which has been proposed as critical to transport these proteins to heterochromatic regions [29] and the BAH domain, crucial for DNA methylation maintenance during DNA replication [30] (Figure 1). Two *Q. suber* proteins were identified belonging to the DRM subfamily, *QsDRM2* and *QsDRM3* (Figure 1). Both *QsDRM2* and *QsDRM3* have the C-terminal DNA Mtase domain but only *QsDRM2* contains the UBA domain in their N-terminal region (PF00627), essential to its de novo methylation activity (Figure 1) [7,31,32]. *QsDRM3* does not possess a UBA domain (Figure 1), which is also described for other DRM-like proteins of maize, rice, and wild peanut [22,33].

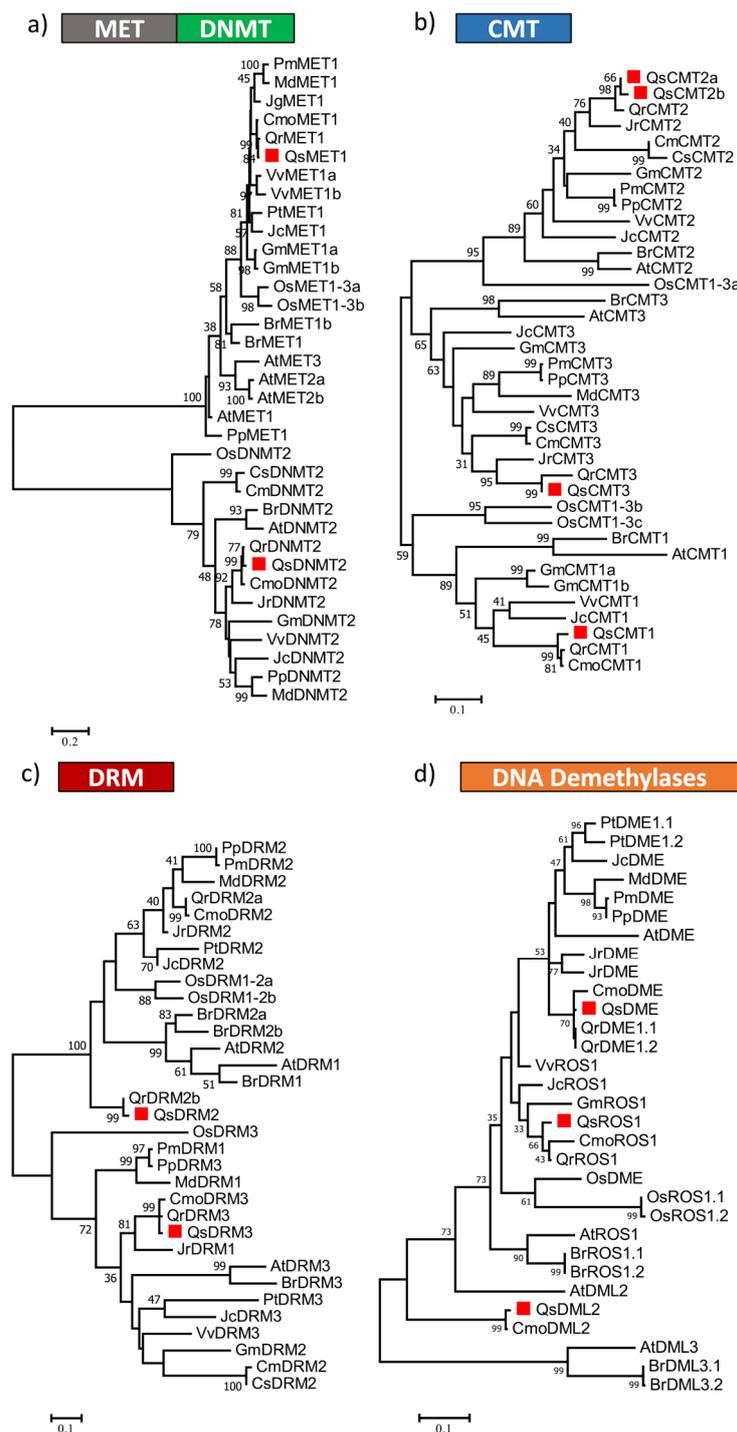


**Figure 1.** Domain organization of the *Quercus suber* L. DNA methyltransferases and demethylases proteins. Schematic diagrams show the domain organization of the enzymes. The proteins and the family they belong to are indicated on the left side of each corresponding schematic diagram. The different conserved protein domains are depicted in different colors as indicated in the legend on the right. C-5 cytosine methyltransferase (IPR001525) is the conserved domain common to all proteins. Replication foci domain RFD (PF12047) and bromo adjacent homology (BAH) domain (PF01426) are present in QsMET1, while chromodomain (CHR) (PF00385) are conserved in chromomethylase proteins. The UBA domain (PF00627) is identified in QsDRM2. The ENDO3c (PF00730), FES (SM000525), Perm-CXXC (PF15629), and RRM\_DME (PF15628) domains are present in all DDME.

A phylogenetic analysis was performed (Figure 2) using the amino acid sequences containing the conserved C-5 Mtase domain of the predicted DNA Mtases of *Q. suber* and of other plant species. QsMET1, as well as QsDNMT2, clustered together with homologs of the closely related Fagaceae species, *Quercus robur* and *Castanea mollissima* (Figure 2a). Phylogenetic trees were constructed separately for the CMT (Figure 2b) and DRM (Figure 2c) subfamilies. The analysis suggests that the *Q. suber* CMT2 paralogues might represent duplicated genes. However, only QsCMT2.1 contains the complete catalytic domain (Figure 1). The *Q. suber* genome lacks a DRM1 homolog like the other Fagaceae family members, *C. mollissima* and *Q. robur* (Figure 2c). QsDRM2 groups together with the *A. thaliana* DRM1 and DRM2 proteins in a clade supported with a bootstrap value of 100 (Figure 2c).

## 2.2. Identification and Classification of *Quercus suber* DNA Demethylases

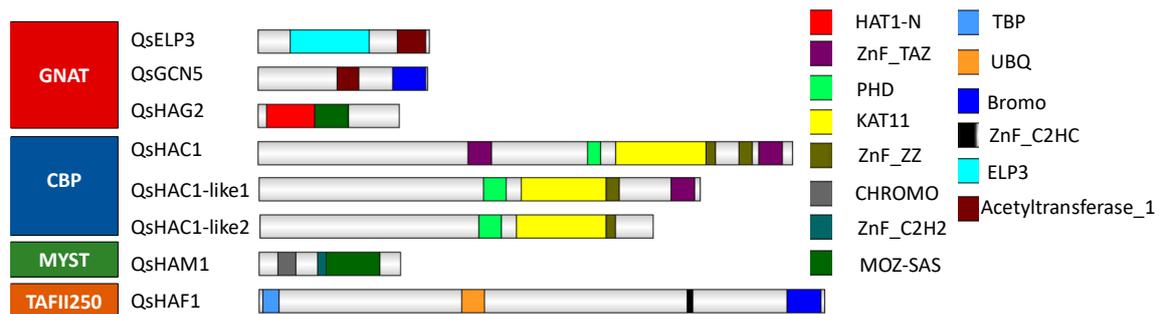
Three *Q. suber* DDME were identified according to their conserved domains: QsDME, QsROS1 and QsDML2 (Figure 1). Four characteristic domains ENDO3c (PF00730), FES (SM000525), Perm-CXXC (PF15629) and RR\_DME (PF15628) are found in all of these three putative DDME proteins (Figure 1), which are involved in base excision DNA repair [9]. A phylogenetic tree was constructed using all DDME proteins from *Q. suber* and from other species (Figure 2d). QsDME and QsROS1 cluster with the respective DME homologs from other Fagaceae, all being more closely related to AtROS1, whereas more distantly related to the other DDME members such as QsDML2, AtDML2, and AtDML3 (Figure 2d).



**Figure 2.** Phylogeny of DNA methyltransferase proteins of (a) Methyltransferase (MET), DNA Methyltransferase homolog 2 (DNMT), (b) Chromomethylase (CMT) family, (c) DRM family, and of (d) DNA Demethylase proteins. The DNA Mtase domain and all DNA demethylase conserved domains of *Quercus suber* L. (Qs), *Arabidopsis thaliana* (L.) Heynh. (At), *Brassica rapa* L. (Br), *Glycine max* (L.) Merr. (Gm), *Jatropha curcas* L. (Jc), *Populus trichocarpa* Torr. & A.Gray ex Hook. (Pt), *Prunus persica* (L.) Batsch (Pp), *Prunus mume* (Siebold) Siebold & Zucc. (Pm), *Vitis vinifera* L. (Vv), *Cucumis melo* L. (Cm), *Cucumis sativus* L. (Cs), *Juglans regia* L. (Jr), *Castanea mollissima* Blume (Cmo), *Quercus robur* L. (Qr) and *Oryza sativa* L. (Os) were aligned using ClustalW and used to infer the evolutionary history using the Maximum-likelihood method. The evolutionary distances (left side scale bar) were computed using the Jones-Taylor-Thornton (JTT) correction model. The numbers at the nodes represent bootstrap values from 1000 replicates. The *Q. suber* proteins are indicated with a red square.

### 2.3. Identification and Classification of *Quercus suber* Histone Acetyltransferases

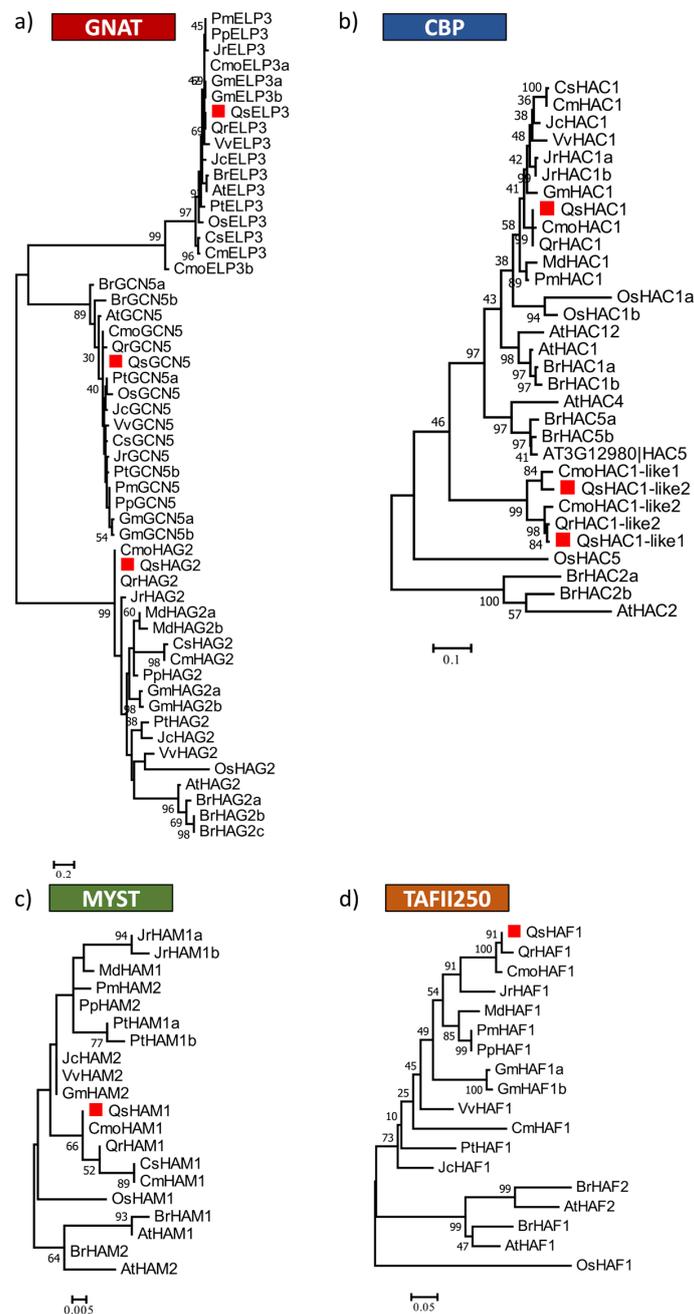
Only eight HATs were identified in the *Q. suber* genome based on homology analysis and protein domain identification (Figure 3). Members of the GNAT family are characterized by an Acetyltransf\_1 domain (PF00583) [15]. All the members of the GNAT family in Arabidopsis have a homolog in *Q. suber*: QsELP3, QsGCN5, and QsHAG2. QsELP3, and QsGCN5 present two domains each, the ELP3 and BROMO domain, respectively, and the Acetyltransf\_1 domain (PF00583). QsHAG2 contains a N-terminal domain HAT1-N (Figure 3) like other homologs in rice, Arabidopsis, soybean [15,34] and a MOZ\_SAS domain (PF01853) (Figure 3), a combination not seen in AtHAG2 but already reported in tomato and litchi [34,35]. The *Q. suber* genome contains genes that code for three CBP class proteins, QsHAC1, QsHAC1-like1, and QsHAC1-like2. Domain analysis reveals that the three CBP present the expected KAT11 (PF08214) domain along with several zinc finger type domains (Figure 3). However, the type and the number of zinc fingers differ in the *Q. suber* CBP members. QsHAC1-like2 has not a Transcription Adaptor putative Zinc finger (TAZ) domain (Figure 3). QsHAM1 has the characteristic C-terminal MOZ\_SAS domain of MYST family members [15,35,36] (Figure 3). Additionally, QsHAM1 contains an N-terminal Chromodomain (PF00385), described as able to recognize and bind specific histone residues [37]. One *Q. suber* protein was associated to the HAF family. QsHAF1 protein contains the characteristic TATA box binding protein (TBP) -binding (PF09247), Ubiquitin (UBQ) (PF00240), ZnF\_C2HC (PF01530) [34,35,38] and the BROMO (PF00439) domain in the C-terminal region (Figure 3) that, similarly to the chromodomain, is known to bind to acetylated histone lysine residues [39].



**Figure 3.** Domain organization of the *Quercus suber* L. histone acetyltransferases proteins. Schematic diagrams show the domain organization of the enzymes. The proteins and the family they belong to are indicated on the left side of each corresponding schematic diagram. The different conserved protein domains are depicted in different colors as indicated in the right side of the legend. Acetyltransferase\_1 (PF00583) and C-terminal Bromo (PF00439) domains are conserved domains of QsGCN5, N-terminal ELP3 (IPR006638), and C-terminal Acetyltransferase\_1 are domains of QsELP3; N-terminal Hat1\_N (PF10394) along with MOZ\_SAS (PF01853) are domains of QsHAG2. KAT11 (PF08214), PHD-finger (PF00628), and TAZ (PF02135) are conserved domains of the histone acetyltransferases (HAT) proteins QsHAC1 and QsHAC1-like1. N-terminal kinase TBP (PF09247), ubiquitin UBQ (PF00240), zinc-finger C2HC (PF01530), and C-terminal bromo are conserved domains of QsHAF1. N-terminal Chromo (PF00385), zinc-finger C2H2 (PF00096), and C-terminal MOZ\_SAS domains are typical of QsHAM1.

A phylogenetic tree was generated for each of the four distinct HAT classes, GNAT, CBP, MYST, and HAFs, using the Acetyltransf\_1 domain (PF00583), the KAT11 domain (PF08214), the MOZ\_SAS domain (PF01853) and the UBQ (PF00240) together with the BROMO (PF00439) domain, respectively (Figure 4). All the members of GNAT in Arabidopsis were clearly distinguished in three different clades (Figure 4a). Each homolog is closely related to the homologs of each GNAT subfamily in other species (Figure 4a). Regarding the CBP family, QsHAC1 is more similar to AtHAC1 and OsHAC1 and more distantly related to QsHAC1-like1 (Figure 4b), that is more closely related to AtHAC2 due to the lack of the N-terminal TAZ domain. A high number of HAC1 copies appear to be common in the majority of the species used to construct the phylogenetic tree (Figure 4b). In contrast, the existence of

only one gene encoding the proteins QsHAM1 and QsHAF1 seems to be a common feature also for the other Fagaceae species represented on the phylogenetic tree (Figure 4c,d).



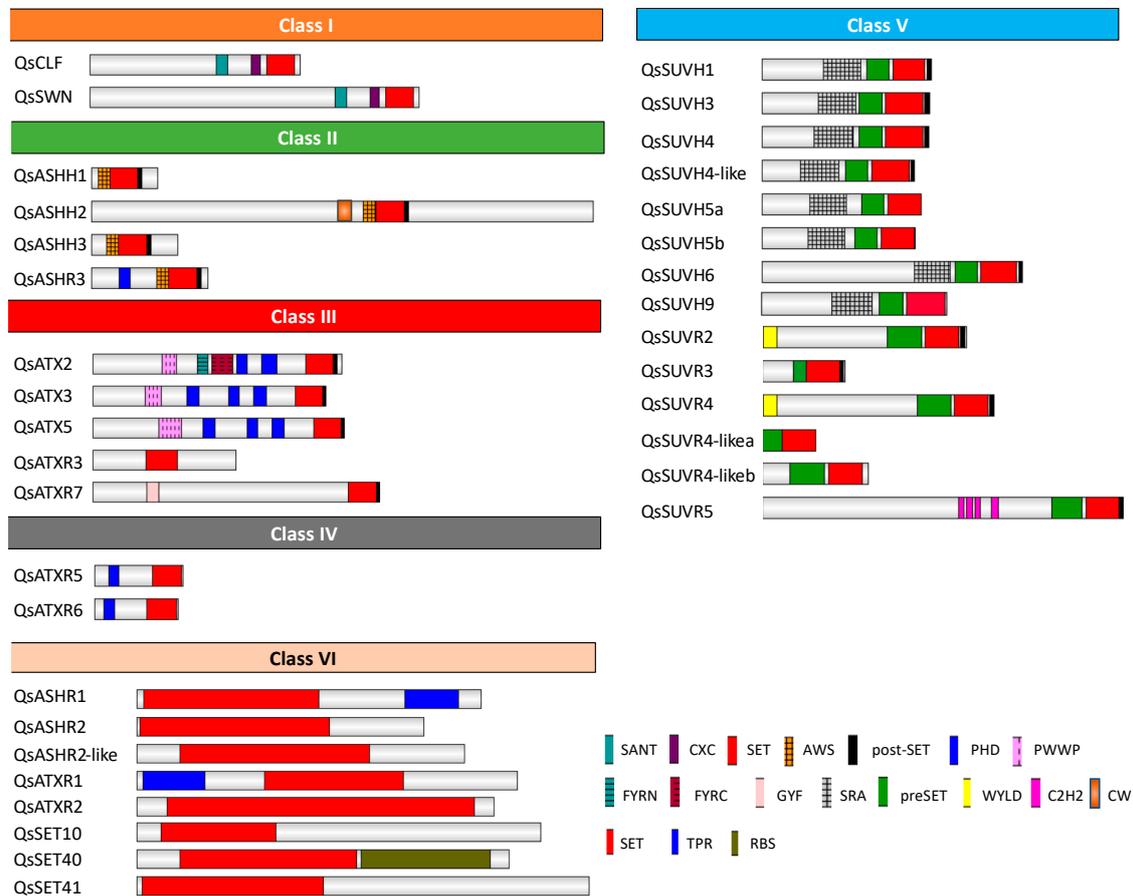
**Figure 4.** Phylogeny of HAT proteins of GNAT (a), CBP (b), MYST (c), and TAFII250 (d) family. The Acetyltransf\_1 domain (GNAT) or the KAT11 domain (CBP), the MOZ\_SAS domain (MYST) and the UBQ and BROMO domains (TAFII250) of *Quercus suber* L. (Qs), *Arabidopsis thaliana* (L.) Heynh. (At), *Brassica rapa* L. (Br), *Glycine max* (L.) Merr. (Gm), *Jatropha curcas* L. (Jc), *Populus trichocarpa* Torr. & A.Gray ex Hook. (Pt), *Prunus persica* (L.) Batsch (Pp), *Prunus mume* (Siebold) Siebold & Zucc. (Pm), *Vitis vinifera* L. (Vv), *Cucumis melo* L. (Cm), *Cucumis sativus* L. (Cs), *Juglans regia* L. (Jr), *Castanea mollissima* Blume (Cmo), *Quercus robur* L. (Qr) and *Oryza sativa* L. (Os) were aligned using ClustalW and used to infer the evolutionary history using the Maximum-likelihood method. The evolutionary distances (left side scale bar) were computed using the Jones-Taylor-Thornton (JTT) correction model. The numbers at the nodes represent bootstrap values from 1000 replicates. The *Q. suber* proteins are indicated with a red square.

#### 2.4. Identification and Classification of *Quercus suber* Histone Methyltransferases

Thirty-five proteins of the SET Domain Group (SDG) were identified in *Q. suber*, according to the presence of the SET domain (PF00856) (Figure 5). Although in *A. thaliana* three class I proteins were described (AtCLF, AtSWN and AtMEA), in *Q. suber*, a homolog of AtMEA could not be identified. QsSWN and QsCLF showed a similar domain architecture to Arabidopsis class I proteins since they contained the Swi3, Ada2, N-Cor, the TFIIB (SANT) domain, C-X(6)-C-X(3)-C-X-C (CXC) domain and SET domain. However, both QsSWN and QsCLF did not show the Enhancer of Zeste Domains (EZDs) (Figure 5) [40] similar to the Litchi class I proteins [35]. In *Q. suber*, only four proteins were identified as class II: QsASHH1, QsASHH2, QsASHH3, and QsASHR3. All the proteins of *Q. suber* identified have the N-terminal AWS domain (SM00570) in common, preceded by a SET and Post-SET domain (SM00508) (Figure 5), like the Arabidopsis homologs [40,41]. In addition, QsASHR3 has an extra N-terminal PHD domain (PF00628), as reported for other species [34,35,42], and QsASHH2 possess the zinc finger domain CW (PF07496) (Figure 5), which has been reported to bind monomethylated Lysine 4 of Histone 3 (H3K4me1) [43,44]. Five proteins (QsATX2, QsATX3, QsATX5, QsATXR3, and QsATXR7) were classified as class III (Figure 5). All these proteins have the same domain architecture of class III Arabidopsis counterparts [40,41], with the exception of QsATXR7 (Figure 5). QsATX3, QsATX5, QsATX2 and QsATXR7 contain SET and post-SET domains but QsATXR3 contains only the SET domain. The SET domain of QsATX2, QsATX3, QsATX5 proteins is detected along with PHD (PF00628) and PWWP (PF00855) domains in their N-terminal part and in the case of QsATXR7 with a GYF domain (SM00444) (Figure 5). AtATXR3 has also a GYF domain before the SET and Post-SET domains [41], but it was not possible to detect this domain in QsATXR3 because the protein sequence is the only one in this study that is not complete (only identified in the EST database and not in the available genome). In addition, QsATX2 contains the FYRN (SM000541) and FYRC (SM000542) domains (Figure 5). QsATXR5 and QsATXR6, belong to the class IV of HMTs. Both proteins have the N-terminal PHD and the C-terminal SET domains (Figure 5), similarly to the Arabidopsis homologs [41]. Fourteen *Q. suber* proteins were identified as class V HMTs, six SUVH (SU(VAR 3-9), and eight SUVR (SU(VAR) 3-9 related) (Figure 5). All SUVH proteins of *Q. suber* include the preSET (SM000468) and the 5mC-binding motif RING finger-associated (SRA) (SM000466). The SUVR proteins of *Q. suber* contain the predicted domains and the N-terminal WIYLD (PF10440) or C2H2 (PF00096) (Figure 5), that are specific of plants [40,45,46]. In addition, eight other proteins (QsSET41, QsASHR1, QsASHR2-like, QsASHR2, QsATXR2, QsATXR1, QsSET10, QsSET40) containing the SET domain that do not belong to the above-mentioned classes were also analyzed and were termed as class VI in this work (Figure 5). QsASHR1 and QsATXR1 possess a TPR (SM00028) domain in the C- and N-terminal part of the protein, respectively. QsSET40 presents a C-terminal RBS (PF09273) domain (Figure 5).

A phylogenetic analysis of the five HMT classes was performed. The class I proteins QsCLF and QsSWN clustered in two different subclades (Supplementary Figure S5A). The class IV proteins QsATXR5 and QsATXR6 grouped in different clades (Supplementary Figure S5B). The class II proteins QsASHH1, QsASHH2, QsASHH3, and QsASHR3 were separated in different clades supported by bootstrap values of 100, 100, 80, and 99, respectively (Supplementary Figure S6A). The class III proteins QsATX3, QsATX5 clustered in the same clade, whereas QsATX2, QsATXR3, and QsATXR7 clustered in different clades each (Supplementary Figure S6B). Regarding class V, QsSUVH1 and QsSUVH3 belong to the clade containing the SUVH1/3/7/8 homolog proteins, positioned in different branches closely related to *Q. robur* and *C. mollissima* homologs, supported by a 98 bootstrap (Supplementary Figure S7). QsSUVH9 grouped together with SUVH2 and SUVH9 homologs in another well supported clade (bootstrap value: 99). One *Q. suber* protein homolog to SUVH6 and two proteins homologs to SUVH5 were positioned in another clade. QsSUVH5 is duplicated in *Q. suber* and no homolog was found in the other Fagaceae trees *C. mollissima* and *Q. robur* (Supplementary Figure S7). Two *Q. suber* proteins cluster in the SUVH4 clade. SUVH4 is duplicated in the two *Quercus* species but also in the annual plant *Glycine max*. Apart from these eight proteins belonging to the SUVH clade of class V, another six proteins belong to the SUVR clade (Supplementary Figure S7). QsSUVR5 and QsSUVR6 were

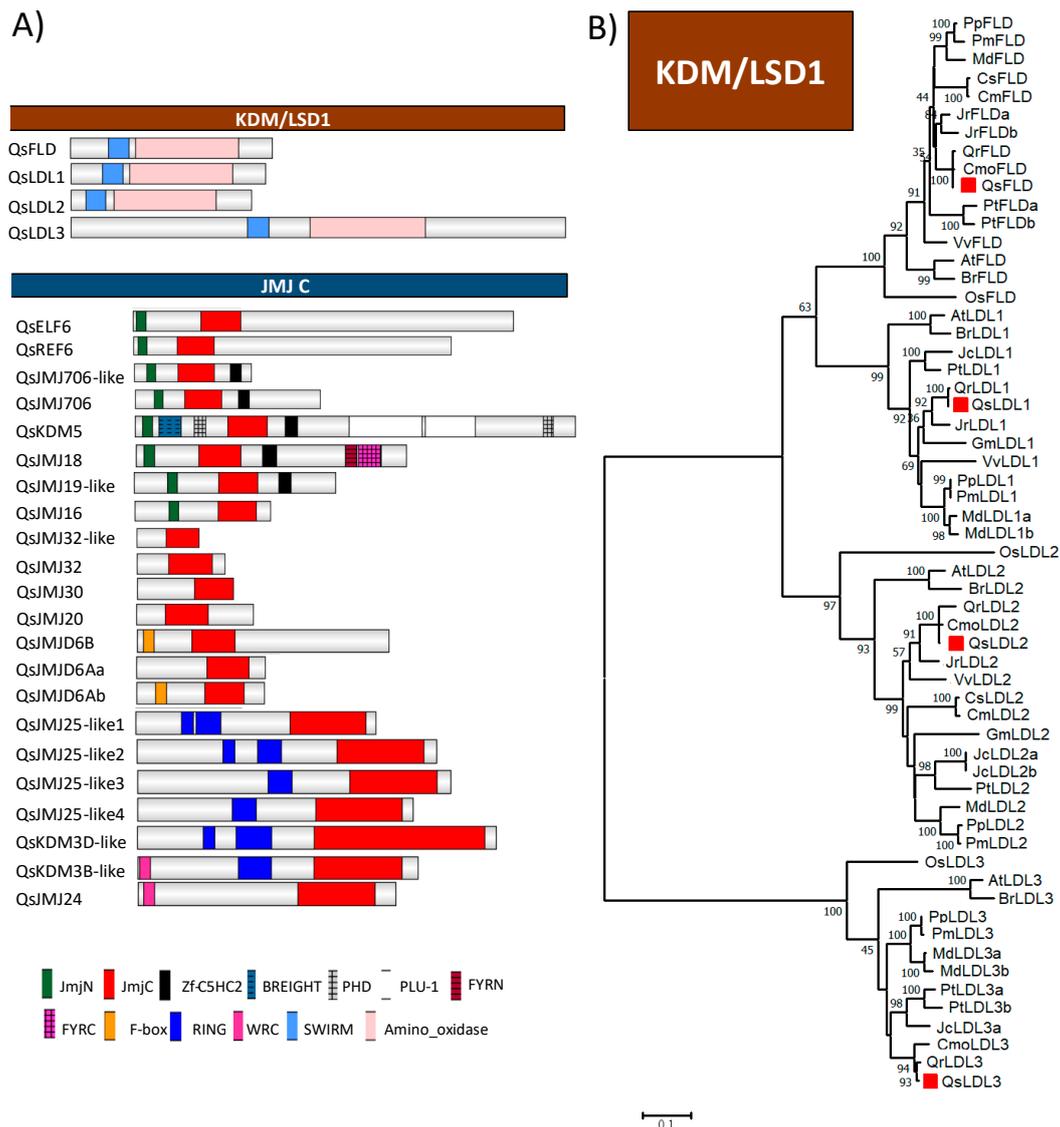
positioned with their closest homologs in distinct clades, while one QsSUVR2 and three SUVR4-like grouped together with their respective homologs in the same clade (Supplementary Figure S7). Each class VI protein was cluster together with their closest homologs in other species (Supplementary Figure S8).



**Figure 5.** Domain organization of the *Quercus suber* L. histone methyltransferases proteins. Schematic diagrams show the domain organization of the enzymes. The protein names and the family they belong to are indicated on the left side and above each corresponding schematic diagram, respectively. The different conserved protein domains are shown in different colors as indicated in the legend on the right side. SANT (SM00717), CXC (PF03638), and SET (PF00856) are conserved domains of Class I. N-terminal AWS (SM00570), SET and Post-SET (SM00508) are conserved domains of Class II. N-terminal PWWP (PF00855), PHD, SET, and Post-SET are conserved domains of some members of Class III. N-terminal GYF (PF02213) is characteristic of QsATXR7. FYRN (SM000541) and FYRC (SM000542) domains are characteristic of QsATX2. N-terminal PHD along with the C-terminal SET are the conserved domains of Class IV. SRA (SM000466), Pre-SET(SM000468), SET, and Post-SET are conserved domains of the SUVH (SU(VAR) 3-9) group of Class V. Pre-SET, SET, and Post-SET alongside with the N-terminal WIYLD (PF10440), or C2H2 (PF00096) or absence of domain are characteristic of Class V SUVH (SU(VAR) 3-9 related) group. Some members of the other SET domain-containing proteins have N-terminal or C-terminal TPR domain, while others have the C-terminal RBS domain.

### 2.5. Identification and Classification of *Quercus suber* Histone Demethylases

Twenty-six *Q. suber* proteins were identified as HDMTs (Figure 6A). HDMTs are divided in two main families: Lysine specific demethylase 1 (KDM/LSD1) and Jumonji C domain-containing proteins (JM).



**Figure 6.** *Quercus suber* L. Histone Demethylases proteins with their (A) domain organization and (B) phylogeny of the KDM/LSD1 family. (A) Schematic diagrams show the domain organization of the enzymes. The proteins and the family they belong to are indicated on left side and above each corresponding schematic diagram, respectively. The different conserved protein domains are depicted in different colors as indicated in the legend at the bottom. RING-finger (IPR001841), WRC (PF08879) and JmjC (PF02373) are conserved domains of the KDM3/JMJD1 class proteins. N-terminal JmjN (PF02375), JmjC, and the C-terminal C5HC2 (PF02928) domains are conserved typical of KDM4/JMJD2 proteins. N-terminal F-box (PF00646), and C-terminal JmjC are conserved domains of the JMJD6 class. KDM/LSD1 proteins contain only the Amino oxidase (IPR002937) and SWIRM (IPR007526) domains. (B) The amino-oxidase and SWIRM domains of *Quercus suber* L. (*Qs*), *Arabidopsis thaliana* (L.) Heynh. (*At*), *Brassica rapa* L. (*Br*), *Glycine max* (L.) Merr. (*Gm*), *Jatropha curcas* L. (*Jc*), *Populus trichocarpa* Torr. & A.Gray ex Hook. (*Pt*), *Prunus persica* (L.) Batsch (*Pp*), *Prunus mume* (Siebold) Siebold & Zucc. (*Pm*), *Vitis vinifera* L. (*Vv*), *Cucumis melo* L. (*Cm*), *Cucumis sativus* L. (*Cs*), *Juglans regia* L. (*Jr*), *Castanea mollissima* Blume (*Cmo*), *Quercus robur* L. (*Qr*) and *Oryza sativa* L. (*Os*) were aligned using ClustalW and used to infer the evolutionary history using the Maximum-likelihood method. The evolutionary distances (left side scale bar) were computed using the Jones-Taylor-Thornton (JTT) correction model. The numbers at the nodes represent bootstrap values from 1000 replicates. The *Q. suber* proteins are indicated with a red square.

The amino-oxidase (PF01593) and SWIRM (PF04433) domains characteristic of the four proteins that belong to KDM1/LSD1 family: QsLDL1, QsLDL2, QsLDL3, and QsFLD (Figure 6A) were used to generate a phylogenetic tree for the KDM1/LSD1 family (Figure 6B). The four proteins were separated in four distinct clades together along with their closest homologs in other species (Figure 6B).

All the *Q. suber* JMJ proteins contain the JmjC domain (PF02373) (Figure 6). QsELF6, QsREF6, QsJMJ706, QsJMJ706-like, QsKDM5, QsJMJ18, QsJMJ19-like, and QsJMJ16 have in common the JmjC and JmjN domains (PF02375) (Figure 6A). QsJMJ25, QsKDM3D-, and QsKDM3B-like present an N-terminal RING finger domain (IPR001841). Both QsJMJ24 and QsKDM3B possess an N-terminal WRC (PF08879) domain. QsJMJ32, QsJMJ32-like, QsJMJ30, and QsJMJ20 proteins do not have any other domain other than the JmjC domain (Figure 6A), while QsJMJD6B and QsJMJD6Ab contain an N-terminal F-box domain (Figure 6A).

The amino acid sequence of the JmjC domain (PF02373) was used to generate the phylogenetic tree for the JMJ family (Supplementary Figure S9). In Arabidopsis, the JMJ family is divided into four well-known classes: KDM4/JMJ2, KDM5/JARID, KDM3 and KDM2 [47]. The first clade (KDM4/JMJ2) contains members of the sub-clades PKDM9, PKDM8: QsREF6, QsELF5, QsJMJ706, and QsJMJ706-like. The second clade (KDM5/JARID) is represented by KDM5 and PKDM7 members: QsKDM5, QsJMJ18, QsJMJ19-like, and QsJMJ16 (Supplementary Figure S9). The *Q. suber* genome encodes four JMJ25-like proteins, one JMJ24-, one KDM3D- and one QsKDM3B-like proteins belonging to the KDM3/JMJ21 family, clustering in the same clade (Supplementary Figure S9). The same number of JMJ25 copies found in the *Q. suber* genome was also found in the genome of the biennial plant *B. rapa*. The fourth clade contains the remaining sub-clades PKDM12 (QsJMJ32, QsJMJ32-like, and QsJMJ30), PKDM11 (QsJMJ20) and JMJD6 (QsJMJD6A and QsJMJD6B) (Supplementary Figure S9). JMJ32 is duplicated in the two *Quercus* species as well as in the perennial *P. trichocarpa* (Supplementary Figure S9).

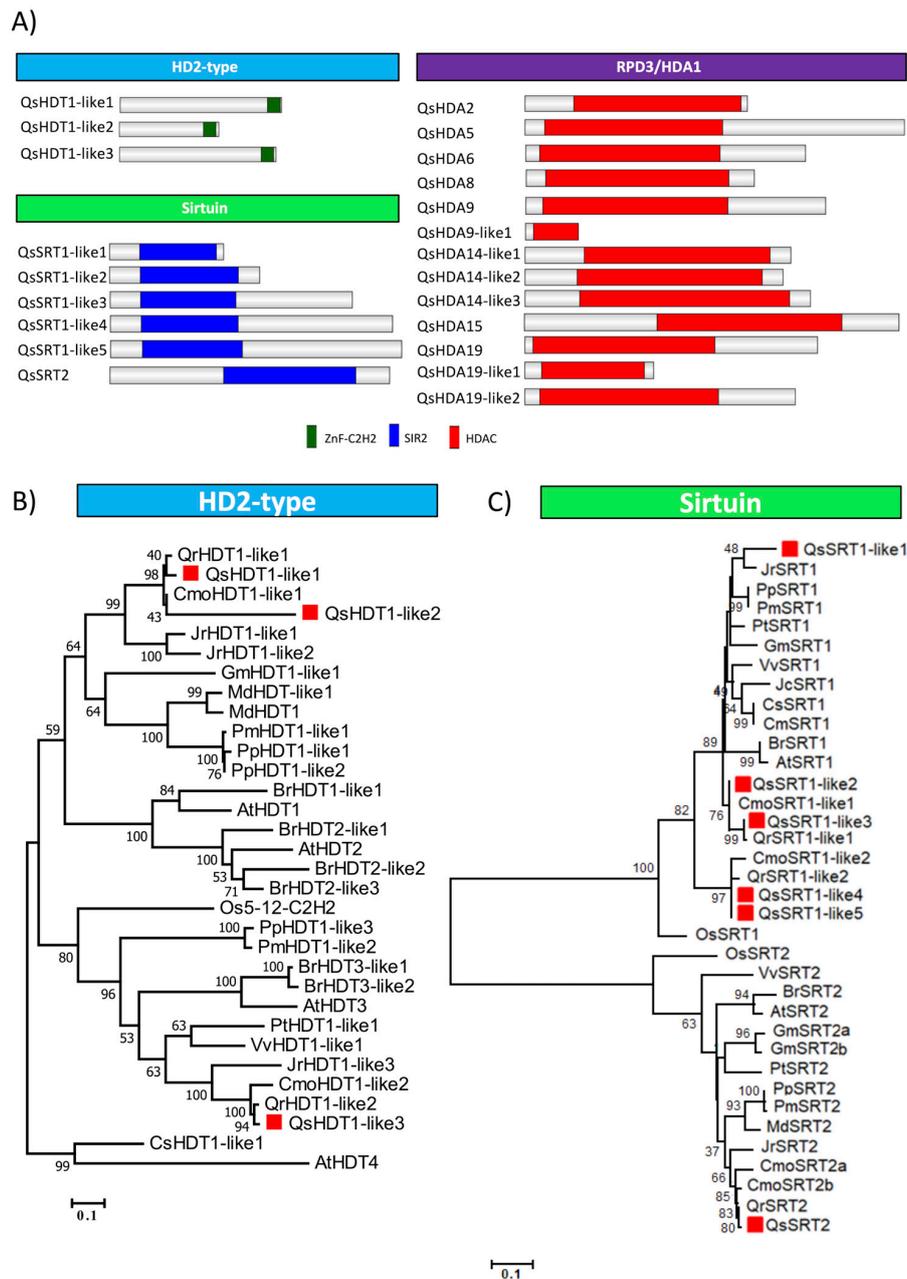
## 2.6. Identification and Classification of *Quercus suber* Histone Deacetylases

HDACs are divided into three main groups: HD2-type, Sirtuin and RPD3/HDA1. A total of 22 *Q. suber* proteins were identified as HDACs-like: three HD2-like proteins, 6 sirtuin-like proteins, and 13 RPD3/HDA1 group proteins (Figure 7A).

Sirtuin-like group members are characterized by the presence of the SIR2 domain (PF02146), while RPD3/HDA1 members show only the HDAC domain (PF00850). Five genes code for QsSRT1-like proteins while only one code for QsSRT2 (Figure 7A).

The phylogenetic tree of HD2-type proteins was generated using the alignment of the conserved ZnF-C2H2 regions (Figure 7B). QsHDT1-like1 and QsHDT1-like2 group in the same clade that AtHDT1 and AtHDT2, while QsHDT1-like3 is more phylogenetically related to AtHDT3. Figure 7C illustrates the divergence of the Sirtuin family proteins based on a phylogenetic tree generated through the SIR2 domain alignment. A higher number of SRT1 copies were found in the Fagaceae genomes, with *Q. robur* and *C. mollissima* having two copies, and the other species having a single copy gene.

Supplementary Figure S10 shows a phylogenetic tree illustrating the relationship among the RPD3/HDA1 superfamily proteins, produced by aligning their HDAC domains. In the first clade it was noticeable a higher number of copies of HDA19 homologs (three) that are not exclusive of *Quercus* species as other plants such as *V. vinifera*, *M. domestica*, *O. sativa* and *P. trichocarpa* have also a high number of copies. In contrast, three copies of HDA14 were exclusively found in *Q. suber*.

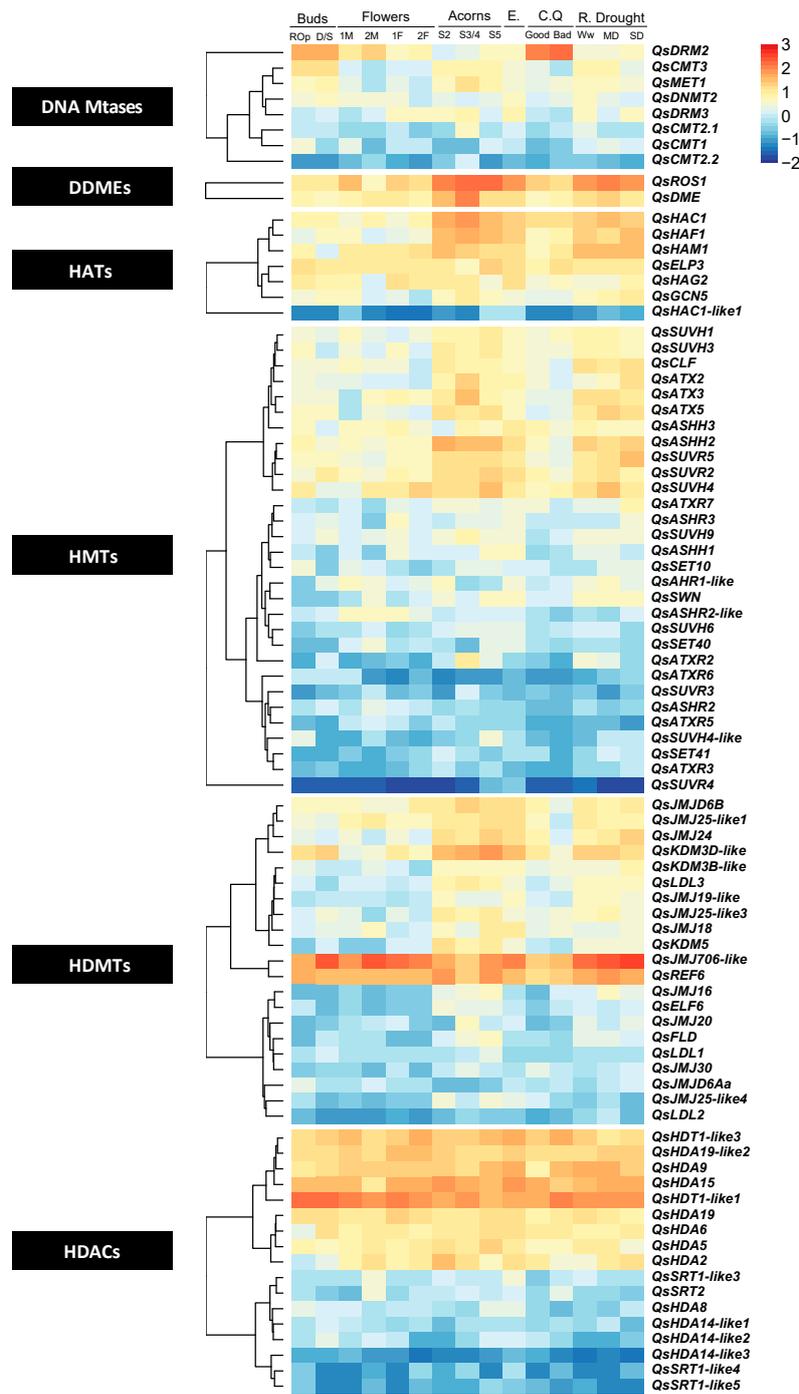


**Figure 7.** Domain organization (A) of HDAC proteins in *Quercus suber* and the phylogeny of (B) HD2 and (C) Sirtuin family. (A) The proteins and the family they belong to are indicated on left side and above each corresponding schematic diagram, respectively. The different conserved protein domains are depicted in different colors as indicated in the right-side legend. The HDAC domain (PF00850) is the conserved domain of RPD3/HDA1 proteins. C-terminal zinc finger domain C2H2 (PF00096) is typical in the HD2-type proteins. The SIR2 proteins contain an SIR2 domain (PF02146). (B) All the protein of HD2-type (HD2 family) and (C) SIR2 domain (Sirtuin family) of *Quercus suber* L. (Qs), *Arabidopsis thaliana* (L.) Heynh. (At), *Brassica rapa* L. (Br), *Glycine max* (L.) Merr. (Gm), *Jatropha curcas* L. (Jc), *Populus trichocarpa* Torr. & A.Gray ex Hook. (Pt), *Prunus persica* (L.) Batsch (Pp), *Prunus mume* (Siebold) Siebold & Zucc. (Pm), *Vitis vinifera* L. (Vv), *Cucumis melo* L. (Cm), *Cucumis sativus* L. (Cs), *Juglans regia* L. (Jr), *Castanea mollissima* Blume (Cmo), *Quercus robur* L. (Qr) and *Oryza sativa* L. (Os) were aligned using ClustalW and used to infer the evolutionary history using the Maximum-likelihood method. The evolutionary distances (left side scale bar) were computed using the Jones-Taylor-Thornton (JTT) correction model. The numbers at the nodes represent bootstrap values from 1000 replicates. The *Q. suber* proteins were indicated with a red square.

## 2.7. Gene Expression of Epigenetic Regulators during *Quercus suber* Plant Development

To infer the putative function of the enzymes under study, their transcript levels were analyzed in RNAseq experiments representing five *Q. suber* tissues. The tissues used were: acorns (from developmental stage 2, 3/4 and 5) [48]; a pool of embryos (collected from the acorns of stage 1 to 8) [48]; cork (bad and good quality) [49]; flowers (male and female) [50]; roots (well-watered, with medium and severe drought) [51] and buds (red or opened buds and dormant or swollen buds) [52]. The transcriptomic analysis of six tissues (root, eco-dormant bud, swelling bud, leaf, in vitro dedifferentiated callus and secondary differentiating xylem), previously done by Lesur and collaborators [53] for the phylogenetically close species *Q. robur* were also mined to detect the expression of these regulators (Table S1).

A heatmap with the normalized expression data (Table S2) was generated for each family of epigenetic modifier enzymes in *Q. suber*. The transcripts of DNA mtases coding genes *QsMET1* and *QsCMT3* were detected mainly in buds, acorns, and roots (Figure 8, Table S2). The closest homolog to *QsMET1* (*QrMET1*) was highly expressed in swelling buds when compared with the other tissues (Table S1). *QsDRM2* showed highest expression in cork and buds. The DDME *QsROS1* was more expressed in roots and acorns while *QsDME* was preferentially expressed in acorns at S3/S4 stages. The HAT genes *QsHAC1*, *QsHAF1*, and *QsHAM1* showed high expression in acorns and in roots when comparing with flowers and buds. The class I HMT gene *QsCLF* was more expressed in roots (Figure 8). The class III HMT genes *QsATX2*, *QsATX3*, and *QsATX5* were expressed in different tissues (Figure 8). *QsATX2* and *QsATX3* had higher expression in the intermediate stages (S3/S4) of fruit development, but was also detected in roots, where *QsATX2* expression was higher during severe drought stress (Figure 8). The class II gene *QsASHH2* was highly expressed during different development stages of fruit development and in embryos (Figure 8). The class V gene *QsSUVH4* expression is higher in red and opened buds, in female flowers and acorns (S5 stage), and in bad quality cork when compared with the other HMT genes (Figure 8). Comparing the different conditions of watering in *Q. suber* roots the other class III HMT gene *QsATXR7* was more expressed in roots under severe drought. However, *QrATXR7* appears to be more expressed in *Q. robur* ecodormant buds (Table S1). The HDAC *QsJMJ706-like* and *QsREF6* transcripts were highly expressed in almost all tissues, when comparing to all the other genes (Figure 8). All *Q. suber* RPD3/HDA1-like members were more expressed in all tissues than the sirtuin family genes (Figure 8). The HD2-type *QsHDT1-like1* was expressed at high levels in all the tissues analyzed in *Q. suber* when compared with the other putative HDAC genes (Figure 8). *Q. robur* homolog *QrHDT1-like1* is also the HDAC gene with highest expression (Table S1).



**Figure 8.** Expression profiles of the homologs of DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone methyltransferases, histone acetylases, and histone demethylases in *Quercus suber* L. The expression levels of the epigenetic regulators identified in diverse *Q. suber* published databases were analyzed after mapping the 454 reads derived from the RNA sequencing of several tissues (categorized on the top) to the *Q. suber* genome and the corresponding read count normalization. The tissues with available libraries include: acorns in 3 development stages (S2, S3/4, S5) [48], embryos (E) [48], good and bad quality cork (C.Q) [49], early (1F and 1M) and late stages (2F and 2M) of female (F) and male flower (M) development [50], roots (R.) with medium (MD), severe (SD) or without drought stress (Ww) [51], samples with red and open buds (ROp) and samples with dormant and swollen buds (D/S) [52]. The results were plotted in an heatmap after Rlog transformation followed by Z-score computation. For interpretation of the expression patterns color, the chart is next to each heatmap. The gene name code was indexed in the heatmap, based on the phylogenetic analysis results.

### 3. Discussion

Taking advantage of the recent publication of the *Q. suber* genome [54], it is now possible to identify the complete list of genes encoding chromatin packaging regulators in this species (Table S3). In this study we were able to identify 8 DNA Mtases, 3 DNA demethylases, 35 HMTs, 26 HDMTs, 8 HATs, and 22 HDACs. In general, all the phylogenetic trees generated in this work are consistent with previous works in other species [15,22,23,25,26,33–35,38,55–58].

*Q. suber*, like other perennials, and in contrast to annual species, must adapt its vegetative and reproductive growth every year to fluctuating environmental conditions that occur over the annual growth. Individuals sense the decreasing photoperiod and temperature and anticipate the winter period by adjusting their own physiology to initiate a rest period. The bud dormancy period during the winter is a good example of a process that has been shown to be dependent on epigenetic gene silencing in several species [59,60]. The cold-induced establishment of trimethylation of lysine 27 of histone 3 (H3K27me3) in known regulators of dormancy, such as *DORMANCY-ASSOCIATED MADS-BOX* (*DAM*) genes, have been reported in several perennial species [59,60], which in turn, may affect the activity of bud burst and flowering inducing genes, such as *FLOWERING LOCUS T* (*FT*) [61–64]. The HMT members of the polycomb repressive complex (PRC) 2, responsible for H3K27me3 deposition such as *CLF*, are likely to regulate growth-dormancy transitions [65,66]. In *Q. suber* buds, it was not possible to differentiate the expression of *QsCLF* between flowering inductive and non-inductive conditions because dormant and swollen buds were pooled together in the data of the transcriptomic study analyzed in this work [52].

Gene silencing by H3K27 methylation and by DNA methylation is essential in order to regulate genome stability by blocking transposable elements [67,68]. In *A. thaliana* the establishment of these epigenetic marks is a result of the redundant action of *ATXR5* and *ATXR6* (H3K27me) and specific DNA Mtases. In *Q. suber*, *QsATXR5* and *QsATXR6* transcripts are less abundant in roots, acorns, and embryos when compared with the other genes encoding HMTs (Figure 8, Table S2). The transcripts of *QsMET1* and *QsCMT3* genes were detected mainly in actively proliferating cell tissues such as buds (with flower meristems inside), acorns and roots (Figure 8, Table S2). The phylogenetic relationship and the conserved domain architecture of all the *QsCMTs* and *QsMET1* proteins (Figures 1 and 2) suggests similar functions to the Arabidopsis homologs. So, *QsMET1* and *QsCMT3* may play a role in maintaining DNA methylation during DNA replication ensuring their correct transmission during subsequent cell divisions [31,68]. Maintaining DNA methylation is a decisive mechanism for *Q. suber* that develops new shoots every year to reinforce the canopy and to produce flowers and fruits ensuring reproductive success. Interestingly, in *Q. suber* the lower expression level of the presumed DNA methylation maintenance genes *QsCMT-* and *QsMET-* like is counterbalanced by the higher expression of the putative *de novo* DNA methylation gene *QsDRM2*. This result was already observed in previous works with this species [69,70]. *QsDRM2* higher expression is well noticed in cork and buds (containing vegetative and reproductive tissues) (Figure 8, Table S2). Previous studies have proven that *AtDRM2* exhibited *de novo* methylation activity through its UBA domain [7,31,32]. *QsDRM2* is phylogenetically close to *AtDRM2* (Figure 2) and contains the UBA domain (Figure 1). So, the activation of putative *de novo* methylation gene *QsDRM2* may be associated with the establishment of new DNA methylation patterns in these tissues [71]. DNA methylation is in close association with H3K9 methylation, through SUVH proteins [72–75] due to the SRA domain that works as a 5mC-binding motif [40,41]. In *Q. suber*, eight proteins were identified as belonging to the SUVH clade (Supplementary Figure S7), and all of them contain the domain structure reported for the Arabidopsis proteins (Figure 5). In *A. thaliana*, SUVH4 is the best-known protein of the SUVH group and is involved mainly in the maintenance of cytosine methylation in a non-CG context [76,77]. DNA methylation variability may contribute to cork cell characteristics linked to quality [78]. Inácio et al. (2018) [70] emphasized the presence of both DNA and H3K9 methylation silencing pathways in cork. Here, we show that *QsSUVH4* is expressed in cork (Figure 8) but also in the other tissues analyzed. The deacetylation of histones has also a negative role in gene expression. The *Q. suber* homologs of HDACs, *QsHDA9* and *QsHDA15* were

highly expressed in several tissues but less in good quality cork and in the last stages of male flower development. *QsHDA5* and *QsHDA2* are less expressed in the overall tissues when compared with the previous ones but were more expressed in the last and early stages of fruit development, respectively. In *A. thaliana* HDA9 and HDA5 modulate flowering time [79,80] while HDA15 has been shown to be a crucial element of photomorphogenesis [81]. The impact of histone deacetylation catalyzed by AtHDA2 is not fully understood. It is possible that these enzymes in *Q. suber* may have the same roles as the Arabidopsis counterparts but also could function in other developmental processes specific of *Q. suber*, such as in cork formation and acorn production.

Lysine methylation of histones is also linked to gene expression activation, which is catalyzed by class II proteins, by methylation of Lysine 36 or 4 of histone 3 (H3K36/H3K4). AtASHH2 has been reported to control *A. thaliana* flowering induction leading to increased transcription of the flowering-repressor *FLOWERING LOCUS C (FLC)* by both H3K36 methylation [82,83] and H3K4 trimethylation [84]. AtASHH2 is essential for proper expression of pivotal genes throughout reproductive development stages, such as floral organ identity or embryo sac and pollen development [85]. *QsASHH2* is highly expressed during different development stages of fruit development and embryos (Figure 8). Since *QsASHH2* groups together with the AtASHH2 (Supplementary Figure S6A) and contains similar domains (Figure 5), it likely performs similar functions. The class III HMT proteins like AtATX1, AtATX2, AtATXR3 and AtATXR7 are also involved in gene activation pathways [86–88]. In *A. thaliana*, AtATXR3 is required for appropriate root growth [89]; *QsATXR3* expression was higher in roots subjected to severe drought (Figure 8). *QsATXR7* and *QsATX2* transcripts were also detected in roots and its expression were higher during severe drought stress (Figure 8), suggesting a putative role in dehydration stress. In contrast, the expression of *QsATXR2* in roots was higher in well watering conditions decreasing with the increase of drought severity. The crosstalk between gene activation pathways and the environment is necessary to promote the expression of genes involved in abiotic stress resistance. The Mediterranean climate, in which *Q. suber* subsists, is very sensitive to hydrological changes such as drought and, as a consequence, the roots should adapt to supply the tree with its water and nutrient demands. In Arabidopsis, AtATXR2 promotes accumulation of H3K36me3 enhancing transcription and it is mainly related to the control of molecular components implicated in root organogenesis [90]. The expression differences of *QsATXR3*, *QsATXR7*, *QsATX2* and *QsATXR2* in contrastive watering condition of *Q. suber* roots suggest a putative role in root development, particularly during dehydration stress. Other genes whose proteins are involved in the deposition of euchromatic marks such as the MYST-like gene *QsHAM1* and the TAFII250 *QsHAF1* have also a strong expression in roots. High levels of expression for both genes were also observed in acorns at different developmental stages (Figure 8). Interestingly, in tomato *HAF1* homolog also has a strong expression in fruit and roots [34] and the *HAM1* homolog is more expressed in flowers and in very immature fruits [34]. In Arabidopsis, single *ham1* or *ham2* mutant plants have a wild-type phenotype, whilst *ham1ham2* double mutation caused mitotic defects in the mega and microgametophyte development [36]. So, a possible role for *QsHAM1* and *QsHAF1* in reproductive development should be considered.

Phylogenetic analysis allowed us to evidence the diversification of epigenetic regulators among the *Q. suber* genome, showing several cases of duplications such as the case of CMT2, SUVH5, SUVH4 and JM32 with two copies, HDA19, HAC1, and HDA14 with three copies, JM25 with four copies, and remarkably, SRT1 with five. The phylogenetic analysis showed that JM32 is not only duplicated in *Q. suber* but also in *Q. robur* and in *P. trichocarpa* in contrast with the other species used in the phylogenetic analysis that contain only JM32 single-copy genes (Supplementary Figure S9). Qian et al. (2015) [91] reported the specificity of JM32 as a single copy gene in plants with the exception of a duplication of the *P. trichocarpa* JM32. Like oaks, *P. trichocarpa* is a woody perennial that requires several years to reach maturity, when cyclical transitions between vegetative and flowering phases occur. In Arabidopsis, AtJM32 mediates the demethylation of H3K27me3 in the flowering-repressor *FLC* locus, preventing premature early flowering under warm temperatures [92]. It is tempting to

suggest that JM32 duplications may have evolved to overcome special requirements of perennials regarding flowering time control.

More than one *SRT1* gene copy was found only in Fagaceae genomes (Figure 7). *SRT1* is required for transposon repression and to regulate the expression of genes involved in plant stress-response and programmed cell death [93,94]. Unfortunately, it was not possible to obtain any clue about *QsSRT1* functional divergence since only *QsSRT1-like3* was significantly expressed in the *Q. suber* tissues analyzed in this work (more expressed in last stages of male flower development, in embryos and in well-watered roots) (Figure 8, Table S2). The closest homolog of *QsSRT1* in *Q. robur* (*QrSRT1-like1*) was not detected in any tissue analyzed but *QrSRT1-like2* had higher expression in roots when compared to other tissues. The higher number of *SRT1* proteins in oaks could be related to genus-specific traits such as their high longevity and essential plasticity. These trees need to have the competence to modulate their own growth (by gene transcription regulation) during more than two hundred years in order to face multiple biotic and abiotic stresses. The duplication of *CMT2*, *SUVH5*, and *HDA14* was only found in *Q. suber* (Figure 2B, Supplementary Figure S7 and Supplementary Figure S10). Moreover, no *SUVH5* homologs were found in *C. mollissima* and *Q. robur*. *AtCMT2* acts in the protection against Arabidopsis genome instability by silencing transposons located in the pericentromeric regions of genomes [32,72,95,96]. *AtHDA14* is involved in Arabidopsis photosynthesis by regulating the expression of RuBisCO activase under low-light conditions [97]. *AtSUVH5* acts in transposon silencing by methylation of histone H3 at Lysine 9 (H3K9me) and by mediating DNA methylation in non-CG sequences [98]. If the function of *CMT2* and *SUVH5* is indeed conserved in *Q. suber*, their duplication may have evolved due the important role of these enzymes in transposon silencing.

The high number of proteins here described and their expression profile in distinct developmental stages suggest an important role for these regulators in cork oak development. The transcriptomic data here represented is, however, not a full representation of the *Q. suber* transcriptome. In the future, the expression should be analyzed using a whole set of plant organs, at different developmental stages and collected at the same conditions to be accurately compared. The identification of a vast set of *Q. suber* epigenetic regulators in this study should greatly facilitate this type of expression analysis. Here, their likely roles were predicted based on the expression levels but also based on the phylogenetic analysis and domain architecture. The domain annotation is an important and valuable tool for discovering the homologs of protein families in plant species in which the functional characterization by mutant analysis is impossible, like in *Q. suber*. The next challenge should be to know if these genes have a similar function to the well characterized Arabidopsis homologs, and to study whether changes in their expression is the cause of an adaptation process.

## 4. Materials and Methods

### 4.1. Identification of *Quercus suber* DNA (De)Methyltransferases and Histone Modifiers

The protein sequences of *Arabidopsis thaliana* DNA Mtases and DDME, as well as some histone modifiers (HATs, HDACs, HMTs, and HDMTs) were retrieved from The Arabidopsis Information Resource (TAIR) database (<http://www.arabidopsis.org/>). These sequences were used as queries to search for *Q. suber* homologs using BLASTp. To ensure that all *Q. suber* DNA Mtases genes were identified, an additional search was performed in the CorkOakDB (<http://www.corkoakdb.org/search>) using the protein domains identification (InterPro ID) common to DNA Mtases (C-5 cytosine methyltransferase IPR001525), HATs (histone acetyltransferase domain, MYST-type IPR002717 and histone acetyltransferase Rtt109/CBP type IPR013178), HDACs (Histone deacetylase IPR000286 and Sirtuin IPR003000), HMTs (SET domain IPR001214) and HDMTs (JmjC domain IPR003347, Amine oxidase IPR002937 and SWIRM domain IPR007526). Proteins identified by these approaches were recorded and redundancy removed. Each corresponding *Q. suber* EST sequence and Arabidopsis protein were used to perform a search in the *Q. suber* genome (CorkOak1.0 version released on January 2018 under the RefSeq assembly: GCF\_002906115.1) using the BLASTn and BLASTp algorithm, and the

complete protein sequences were retrieved. All the epigenetic modifier enzymes identified in *Q. suber* are listed in Table S3.

#### 4.2. Prediction of Protein Domains

*Q. suber* protein sequences were analyzed for recognizable domains using NCBI Batch Conserved Domain search tool (<http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>). The presence of specific domains and their organization was also verified using the Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de>). Schematic diagram of protein domain structures with its functional domains were constructed using Illustrator for Biological Sequences (IBS) (version IBS 1.0.1) (<http://ibs.biocuckoo.org/index.php>).

#### 4.3. Phylogenetic Analysis

*Arabidopsis thaliana* (At) protein sequences were used as queries to search for homologs in other species such as *Brassica rapa* (Br), *Glycine max* (Gm), *Jatropha curcas* (Jc), *Populus trichocarpa* (Pt), *Prunus persica* (Pp), *Prunus mume* (Pm), *Vitis vinifera* (Vv), *Cucumis melo* (Cm), *Cucumis sativus* (Cs) and *Juglans regia* (Jr) using the BLASTp tool in NCBI (<https://blast.ncbi.nlm.nih.gov/>). Protein homologs from other Fagaceae species (*Castanea mollissima* (Cmo) and *Quercus robur* (Qr)) were also included by performing a local BLASTp (prfactBLAST 2.0) against the Draft of Whole Genome Sequence (v1.1) of *C. mollissima* (<http://www.hardwoodgenomics.org/chinese-chestnut-genome>), and a tBLASTn (prfactBLAST 2.0) search against *Q. robur* data (Oak assembly V3\_OCV3-91k). Protein data from *Oryza sativa* (Os) was retrieved from RGAP release 7 (<http://rice.plantbiology.msu.edu/>) to be used as an out-group. All the amino acid sequences were aligned using the alignment tool CLUSTAL W [99], using the Gonnet Protein Weight Matrix, and changing the default parameters of the Multiple Alignment Gap Opening penalty to 3 and the Multiple Alignment Gap Extension penalty to 1.8, more appropriate to proteins [100]. Phylogenetic trees were constructed using the Maximum-likelihood algorithm with the Jones-Taylor-Thornton (JTT) correction model (MEGA 7.0 software). The bootstrap consensus tree was inferred from 1000 replicates to obtain a support value for each branch.

#### 4.4. Transcriptome Data Analysis

RNA sequencing data were acquired from the Sequencing Read Archive of NCBI (<https://www.ncbi.nlm.nih.gov/sra/>) using the experiments provided by the high-throughput sequencing of the *Q. suber* transcriptome [101]. The *Q. suber* transcriptome was obtained from published libraries using tissues at different stages/conditions of development such as acorns from developmental stage 2 (ERR490202), 3 and 4 (ERR490203) and 5 (ERR490204) [48]; a pool of embryos collected from the acorns of stage 1 to 8 (ERR490207) [48]; cork of bad (SRR1009171) and good (SRR1009172) quality [49]; male (SRR1609152) and female (SRR1609153) flowers [50]; roots of plants with different degrees of watering: medium (SRR1812375), low (SRR1812376) and abundant (SRR1812377) [51]; red and opened buds (SRR5345606) and dormant and swollen buds (SRR5345607) [52]. The libraries were first trimmed to remove SMART adapters, present in the 454 libraries, using AlienTrimmer [102] and then filtered by QTRIM [103] using default quality parameters. Alignment against the *Q. suber* published genome CorkOak1.0 was performed with Burrows–Wheeler Aligner (BWA) using the BWA-MEM algorithm [104]. The alignment files were used to quantify gene expression using feature Counts [105].

#### 4.5. Differential Expression Analysis

Gene expression was analyzed in different *Q. suber* libraries using the read counts generated after the mapping of 454 reads into the CorkOak1.0 genome. Each expression pattern was normalized and estimated by DESeq2, an R package for differential expression analysis [106]. The Rlog function of DESeq2 was also used to transform the count data to the log<sub>2</sub> scale in order to minimize variances between samples with fewer counts, considering each library size. The resulting values were Z scored, scaled, and used for visualization and clustering. The R package NMF was used to plot heatmaps

using the *heatmap* function that is based on the Euclidian algorithm. Gene expression was also studied in the phylogenetically closed species *Q. robur* using transcriptomic data obtained by Lesur and colleagues [53], in which the number of normalized read counts for root, eco-dormant bud, swelling bud, leaf, in vitro dedifferentiated *calli*, and secondary differentiating xylem are available.

**Supplementary Materials:** Supplementary Materials can be found at <http://www.mdpi.com/1422-0067/21/11/3783/s1>.

**Author Contributions:** Conceptualization—H.G.S., M.M.R.C., L.M.-C.; Data curation—H.G.S., A.P.M.; Formal analysis—H.G.S., A.P.M.; Funding acquisition—M.M.R.C., L.M.C.; Investigation—H.G.S.; Project administration—M.M.R.C., L.M.C.; Software—A.P.M.; Supervision—M.M.R.C., L.M.C.; Visualization—H.G.S., R.S.S., A.P.M.; Writing—original draft—H.G.S.; Writing—review & editing—H.G.S., R.S.S., A.P.M., M.M.R.C., L.M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the FEDER funds through the Operational Competitiveness Programme—COMPETE and by National Funds through FCT—Fundação para a Ciência e a Tecnologia under the projects PTDC/AGR-GPL/118508/2010, “Characterization of Reproductive Development of *Quercus suber*” and PTDC/AGR-FOR/3356/2014, “Characterization of cork formation and reproductive biology in a cork oak hybrids population” and the PhD grant ref. SFRH/BD/111529/2015. Work supported by UIDP/04046/2020 and UIDB/04046/2020 Centre grants from FCT, Portugal to BioISI and by UID/AGR/04129/2020 Centre grants from FCT, Portugal to LEAF.

**Acknowledgments:** We acknowledge Ana Raposo for preliminary analysis of some epigenetic regulator families.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Gruenbaum, Y.; Naveh-Many, T.; Cedar, H.; Razin, A. Sequence specificity of methylation in higher plant DNA. *Nature* **1981**, *292*, 860–862. [[CrossRef](#)] [[PubMed](#)]
2. Bewick, A.J.; Schmitz, R.J. Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* **2017**, *36*, 103–110. [[CrossRef](#)] [[PubMed](#)]
3. Zhang, X.; Yazaki, J.; Sundaresan, A.; Cokus, S.; Chan, S.W.-L.; Chen, H.; Henderson, I.R.; Shinn, P.; Pellegrini, M.; Jacobsen, S.E.; et al. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **2006**, *126*, 1189–1201. [[CrossRef](#)]
4. Chen, T.; Li, E. Structure and Function of Eukaryotic DNA Methyltransferases. *Curr. Top. Dev. Biol.* **2004**, *60*, 55–89.
5. Su, C.; Wang, C.; He, L.; Yang, C.; Wang, Y. Shotgun Bisulfite Sequencing of the *Betula platyphylla* Genome Reveals the Tree’s DNA Methylation Patterning. *Int. J. Mol. Sci.* **2014**, *15*, 22874–22886. [[CrossRef](#)]
6. Lister, R.; Omalley, R.; Tonti-Filippini, J.; Gregory, B.D.; Berry, C.C.; Millar, A.H.; Ecker, J.R. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* **2008**, *133*, 523–536. [[CrossRef](#)]
7. Law, J.A.; Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **2010**, *11*, 204–220. [[CrossRef](#)]
8. Naumann, U.; Daxinger, L.; Kanno, T.; Eun, C.; Long, Q.; Lorkovic, Z.; Matzke, M.; Matzke, A.J.M. Genetic evidence that DNA methyltransferase DRM2 has a direct catalytic role in RNA-directed DNA methylation in Arabidopsis thaliana. *Genetics* **2011**, *187*, 977–979. [[CrossRef](#)]
9. Gong, Z.; Morales-Ruiz, T.; Ariza, R.R.; Roldan-Arjona, T.; David, L.; Zhu, J. ROS1, a Repressor of Transcriptional Gene Silencing in Arabidopsis, Encodes a DNA Glycosylase/Lyase. *Cell* **2002**, *111*, 803–814. [[CrossRef](#)]
10. Xiao, W.; Gehring, M.; Choi, Y.; Margossian, L.; Pu, H.; Harada, J.J.; Goldberg, R.B.; Pennell, R.I.; Fischer, R.L. Imprinting of the MEA Polycomb Gene Is Controlled by Antagonism between MET1 Methyltransferase and DME Glycosylase. *Dev. Cell* **2003**, *5*, 891–901. [[CrossRef](#)]
11. Gehring, M.; Huh, J.H.; Hsieh, T.-F.; Penterman, J.; Choi, Y.; Harada, J.J.; Goldberg, R.B.; Fischer, R.L. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **2006**, *124*, 495–506. [[CrossRef](#)] [[PubMed](#)]
12. Bauer, I. Zur Bestimmung der Isocitronensäure neben Citronensäure. *Anal. Bioanal. Chem.* **1966**, *219*, 207–208. [[CrossRef](#)]
13. Hebbes, T.R.; Thorne, A.W.; Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **1988**, *7*, 1395–1402. [[CrossRef](#)]

14. Sterner, D.E.; Berger, S. Acetylation of Histones and Transcription-Related Factors. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 435–459. [[CrossRef](#)]
15. Pandey, R. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res.* **2002**, *30*, 5036–5055. [[CrossRef](#)] [[PubMed](#)]
16. Ng, D.W.-K.; Wang, T.; Chandrasekharan, M.B.; Aramayo, R.; Kertbundit, S.; Hall, T.C. Plant SET domain-containing proteins: Structure, function and regulation. *Biochim. Biophys. Acta (BBA) Bioenerg.* **2007**, *1769*, 316–329. [[CrossRef](#)]
17. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J.R.; Cole, P.A.; Casero, R.A.; Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1 Instance, Histone H3 K9 (H3-K9) Methylation Is Associ-Ated with Heterochromatin Formation (Nakayama et Al and Also. *Sidney Kimmel Compr. Cancer Cent.* **2004**, *119*, 941–953. [[CrossRef](#)]
18. Tsukada, Y.I.; Fang, J.; Erdjument-Bromage, H.; Warren, M.E.; Borchers, C.H.; Tempst, P.; Zhang, Y. Histone Demethylation by a Family of JmjC Domain-Containing Proteins. *Nature* **2006**, *439*, 811–816. [[CrossRef](#)]
19. Klose, R.J.; Zhang, Y. Regulation of histone methylation by demethylation and demethylation. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 307–318. [[CrossRef](#)]
20. Kumar, S.; Cheng, X.; Klimašauskas, S.; Sha, M.; Pósfai, J.; Roberts, R.J.; Wilson, G.G. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **1994**, *22*, 1–10. [[CrossRef](#)]
21. Pósfai, J.; Bhagwat, A.S.; Pósfai, G.; Roberts, R.J. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **1989**, *17*, 2421–2435. [[CrossRef](#)] [[PubMed](#)]
22. Pavlopoulou, A.; Kossida, S. Plant Cytosine-5 DNA Methyltransferases: Structure, Function, and Molecular Evolution. *Curr. Top. Dev. Biol.* **2007**, *90*, 530–541. [[CrossRef](#)]
23. Cao, D.; Ju, Z.; Gao, C.; Mei, X.; Fu, D.; Zhu, H.; Luo, Y.; Zhu, B. Genome-wide identification of cytosine-5 DNA methyltransferases and demethylases in *Solanum lycopersicum*. *Gene* **2014**, *550*, 230–237. [[CrossRef](#)] [[PubMed](#)]
24. Chang, L.; Zhang, Z.; Han, B.; Li, H.; Dai, H.; He, P.; Tian, H. Isolation of DNA-methyltransferase genes from strawberry (*Fragaria × ananassa* Duch.) and their expression in relation to micropropagation. *Plant Cell Rep.* **2009**, *28*, 1373–1384. [[CrossRef](#)] [[PubMed](#)]
25. Garg, R.; Kumari, R.; Tiwari, S.; Goyal, S. Genomic Survey, Gene Expression Analysis and Structural Modeling Suggest Diverse Roles of DNA Methyltransferases in Legumes. *PLoS ONE* **2014**, *9*, e88947. [[CrossRef](#)] [[PubMed](#)]
26. Gianoglio, S.; Moglia, A.; Acquadro, A.; Comino, C.; Portis, E. The genome-wide identification and transcriptional levels of DNA methyltransferases and demethylases in globe artichoke. *PLoS ONE* **2017**, *12*, e0181669. [[CrossRef](#)] [[PubMed](#)]
27. Sharma, R.; Singh, R.K.M.; Malik, G.; Deveshwar, P.; Tyagi, A.K.; Kapoor, S.; Kapoor, M. Rice cytosine DNA methyltransferases-gene expression profiling during reproductive development and abiotic stress. *FEBS J.* **2009**, *276*, 6301–6311. [[CrossRef](#)] [[PubMed](#)]
28. Xu, J.; Xu, H.; Xu, Q.; Deng, X. Characterization of DNA Methylation Variations during Fruit Development and Ripening of Sweet Orange. *Plant Mol. Biol. Rep.* **2014**, *33*, 1–11. [[CrossRef](#)]
29. Eissenberg, J.C. Molecular biology of the chromo domain: An ancient chromatin module comes of age. *Gene* **2001**, *275*, 19–29. [[CrossRef](#)]
30. Callebaut, I.; Courvalin, J.-C.; Mornon, J.-P. The BAH (bromo-adjacent homology) domain: A link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* **1999**, *446*, 189–193. [[CrossRef](#)]
31. Cao, X.; Aufsatz, W.; Zilberman, D.; Mette, M.; Huang, M.S.; Matzke, M.; Jacobsen, S.E. Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. *Curr. Biol.* **2003**, *13*, 2212–2217. [[CrossRef](#)] [[PubMed](#)]
32. Cao, X.; Jacobsen, S.E. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* **2002**, *12*, 1138–1144. [[CrossRef](#)]
33. Wang, P.; Gao, C.; Bian, X.; Zhao, S.; Zhao, C.; Xia, H.; Song, H.; Hou, L.; Wan, S.; Wang, X.-J. Genome-Wide Identification and Comparative Analysis of Cytosine-5 DNA Methyltransferase and Demethylase Families in Wild and Cultivated Peanut. *Front. Plant Sci.* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]

34. Cigliano, R.A.; Sanseverino, W.; Cremona, G.; Ercolano, M.R.; Conicella, C.; Consiglio, F.M. Genome-wide analysis of histone modifiers in tomato: Gaining an insight into their developmental roles. *BMC Genom.* **2013**, *14*, 57. [[CrossRef](#)] [[PubMed](#)]
35. Peng, M.; Ying, P.; Liu, X.; Li, C.; Xia, R.; Li, J.; Zhao, M. Genome-Wide Identification of Histone Modifiers and Their Expression Patterns during Fruit Abscission in Litchi. *Front. Plant Sci.* **2017**, *8*, 57. [[CrossRef](#)] [[PubMed](#)]
36. Latrasse, D.; Benhamed, M.; Henry, Y.; Domenichini, S.; Kim, W.; Zhou, D.-X.; Delarue, M. The MYST histone acetyltransferases are essential for gametophyte development in Arabidopsis. *BMC Plant Biol.* **2008**, *8*, 121. [[CrossRef](#)]
37. De La Cruz, X.; Lois, S.; Sánchez-Molina, S.; Martínez-Balbás, M.A. Do protein motifs read the histone code? *BioEssays* **2005**, *27*, 164–175. [[CrossRef](#)]
38. Liu, X.; Luo, M.; Zhang, W.; Zhao, J.; Zhang, J.; Wu, K.; Tian, L.; Duan, J. Histone acetyltransferases in rice (*Oryza sativa* L.): Phylogenetic analysis, subcellular localization and expression. *BMC Plant Biol.* **2012**, *12*, 145. [[CrossRef](#)]
39. Ghasroddashti, E.; Pantarotto, J.; Macrae, R.; Gerig, L. SU-FF-T-412: The Reliability of Surrogates in Predicting Tumour Motion: A Comparison of Surrogate Based and Non-Surrogate Based Approach. *Med. Phys.* **2007**, *34*, 2496. [[CrossRef](#)]
40. Pontivianne, F.; Blevins, T.; Pikaard, C.S. Arabidopsis Histone Lysine Methyltransferases. *Adv. Bot. Res.* **2010**, *2296*, 1–18. [[CrossRef](#)]
41. Emmanuel, A.; Kamm, M.; Cohen, R. Preface. *Best Pr. Res. Clin. Gastroenterol.* **2002**, *16*, 527. [[CrossRef](#)]
42. Huang, Y.; Liu, C.; Shen, W.; Ruan, Y. Phylogenetic analysis and classification of the Brassica rapa SET-domain protein family. *BMC Plant Biol.* **2011**, *11*, 175. [[CrossRef](#)] [[PubMed](#)]
43. Hoppmann, V.; Thorstensen, T.; Kristiansen, P.E.; Veiseth, S.V.; Rahman, M.A.; Finne, K.; Aalen, R.B.; Aasland, R. The CW domain, a new histone recognition module in chromatin proteins. *EMBO J.* **2011**, *30*, 1939–1952. [[CrossRef](#)]
44. Liu, Y.; Huang, Y. Uncovering the mechanistic basis for specific recognition of monomethylated H3K4 by the CW domain of Arabidopsis histone methyltransferase SDG8. *J. Biol. Chem.* **2018**, *293*, 6470–6481. [[CrossRef](#)] [[PubMed](#)]
45. Thorstensen, T.; Fischer, A.; Sandvik, S.V.; Johnsen, S.S.; Grini, P.; Reuter, G.; Aalen, R.B. The Arabidopsis SUV4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. *Nucleic Acids Res.* **2006**, *34*, 5461–5470. [[CrossRef](#)]
46. Baumbusch, L.O. The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.* **2001**, *29*, 4319–4333. [[CrossRef](#)]
47. Zhou, X.; Ma, H. Evolutionary history of histone demethylase families: Distinct evolutionary patterns suggest functional divergence. *BMC Evol. Biol.* **2008**, *8*, 294. [[CrossRef](#)]
48. Miguel, A.; De Vega, J.J.; Marum, L.; Chaves, I.; Santo, T.; Leitão, J.; Varela, M.C.; Miguel, C.M. Characterization of the cork oak transcriptome dynamics during acorn development. *BMC Plant Biol.* **2015**, *15*, 158. [[CrossRef](#)]
49. Teixeira, R.T.; Fortes, A.M.; Pinheiro, C.; Pereira, H. Comparison of good- and bad-quality cork: Application of high-throughput sequencing of phellogenic tissue. *J. Exp. Bot.* **2014**, *65*, 4887–4905. [[CrossRef](#)]
50. Rocheta, M.; Sobral, R.; Magalhães, J.; Amorim, M.I.; Ribeiro, A.T.; Pinheiro, M.; Egas, C.; Morais-Cecã, L.; Costa, M.M.R.; Sobral, R.; et al. Comparative transcriptomic analysis of male and female flowers of monoecious *Quercus suber*. *Front. Plant Sci.* **2014**, *5*, 599. [[CrossRef](#)]
51. Magalhães, A.; Verde, N.; Reis, F.; Martins, I.; Costa, D.; Lino-Neto, T.; Castro, P.H.; Tavares, R.; Azevedo, H. RNA-Seq and Gene Network Analysis Uncover Activation of an ABA-Dependent Signalosome During the Cork Oak Root Response to Drought. *Front. Plant Sci.* **2016**, *6*, 3523. [[CrossRef](#)] [[PubMed](#)]
52. Usié, A.; Simões, F.; Barbosa, P.; Meireles, B.; Chaves, I.; Alves, E.; Folgado, A.; Almeida, M.H.; Matos, J.; Ramos, A.M. Comprehensive Analysis of the Cork Oak (*Quercus suber*) Transcriptome Involved in the Regulation of Bud Sprouting. *Forests* **2017**, *8*, 486. [[CrossRef](#)]
53. Lesur, I.; Le Provost, G.; Bento, P.; Da Silva, C.; Leple, J.-C.; Murat, F.; Ueno, S.; Bartholomé, J.; Lalanne, C.; Ehrenmann, F.; et al. The oak gene expression atlas: Insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC Genom.* **2015**, *16*, 112. [[CrossRef](#)] [[PubMed](#)]

54. Ramos, A.P.; Usié, A.; Barbosa, P.; Barros, P.M.; Capote, T.; Chaves, I.; Simões, F.; Abreu, I.A.; Carrasquinho, I.; Faro, C.; et al. The draft genome sequence of cork oak. *Sci. Data* **2018**, *5*, 180069. [[CrossRef](#)]
55. Aquea, F.; Timmermann, T.; Arce-Johnson, P. Analysis of histone acetyltransferase and deacetylase families of *Vitis vinifera*. *Plant Physiol. Biochem.* **2010**, *48*, 194–199. [[CrossRef](#)]
56. Liew, L.C.; Singh, M.B.; Bhalla, P.L. An RNA-Seq Transcriptome Analysis of Histone Modifiers and RNA Silencing Genes in Soybean during Floral Initiation Process. *PLoS ONE* **2013**, *8*, e77502. [[CrossRef](#)]
57. Papaefthimiou, D.; Likotrafiti, E.; Kapazoglou, A.; Bladenopoulos, K.; Tsaftaris, A. Epigenetic chromatin modifiers in barley: III. Isolation and characterization of the barley GNAT-MYST family of histone acetyltransferases and responses to exogenous ABA. *Plant Physiol. Biochem.* **2010**, *48*, 98–107. [[CrossRef](#)]
58. Malik, G.; Dangwal, M.; Kapoor, S.; Kapoor, M. Role of DNA methylation in growth and differentiation in *Physcomitrella patens* and characterization of cytosine DNA methyltransferases. *FEBS J.* **2012**, *279*, 4081–4094. [[CrossRef](#)]
59. Horvath, D.; Sung, S.; Kim, D.; Chao, W.; Anderson, J. Characterization, expression and function of DORMANCY ASSOCIATED MADS-BOX genes from leafy spurge. *Plant Mol. Biol.* **2010**, *73*, 169–179. [[CrossRef](#)]
60. Leida, C.; Conesa, A.; Llácer, G.; Badenes, M.L.; Ríos, G. Histone modifications and expression of DAM6 gene in peach are modulated during bud dormancy release in a cultivar-dependent manner. *New Phytol.* **2011**, *193*, 67–80. [[CrossRef](#)]
61. Hao, X.; Chao, W.; Yang, Y.; Horvath, D. Coordinated Expression of Flowering Locus T and Dormancy Associated Mads-Box-Like Genes in Leafy Spurge. *PLoS ONE* **2015**, *10*, e0126030. [[CrossRef](#)] [[PubMed](#)]
62. Ito, A.; Saito, T.; Sakamoto, D.; Sugiura, T.; Bai, S.; Moriguchi, T. Physiological differences between bud breaking and flowering after dormancy completion revealed by DAM and FT/TFL1 expression in Japanese pear (*Pyrus pyrifolia*). *Tree Physiol.* **2015**, *36*, 109–120. [[CrossRef](#)] [[PubMed](#)]
63. Niu, Q.; Li, J.; Cai, D.; Qian, M.; Jia, H.; Bai, S.; Hussain, S.; Liu, G.; Teng, Y.; Zheng, X. Dormancy-associated MADS-box genes and microRNAs jointly control dormancy transition in pear (*Pyrus pyrifolia* white pear group) flower bud. *J. Exp. Bot.* **2015**, *67*, 239–257. [[CrossRef](#)] [[PubMed](#)]
64. Singh, R.K.; Maurya, J.P.; Azeez, A.; Miskolczi, P.; Tylewicz, S.; Stojkovič, K.; Delhomme, N.; Busov, V.; Bhalerao, R.P. A genetic network mediating the control of bud break in hybrid aspen. *Nat. Commun.* **2018**, *9*, 4173. [[CrossRef](#)] [[PubMed](#)]
65. Lomax, A.; Woods, D.P.; Dong, Y.; Bouché, F.; Rong, Y.; Mayer, K.; Zhong, X.; Amasino, R.M. An ortholog of Curly Leaf/Enhancer of Zeste like-1 is required for proper flowering in *Brachypodium distachyon*. *Plant J.* **2018**, *93*, 871–882. [[CrossRef](#)]
66. Ruttink, T.; Arend, M.; Morreel, K.; Storme, V.; Rombauts, S.; Fromm, J.; Bhalerao, R.P.; A Boerjan, W.; Rohde, A. A Molecular Timetable for Apical Bud Formation and Dormancy Induction in Poplar. *Plant Cell* **2007**, *19*, 2370–2390. [[CrossRef](#)]
67. Jacob, Y.; Feng, S.; LeBlanc, C.; Bernatavichute, Y.; Stroud, H.; Cokus, S.; Johnson, L.; Pellegrini, M.; Jacobsen, S.; Michaels, S. ATXR5 and ATXR6 Are Novel H3K27 Monomethyltransferases Required for Chromatin Structure and Gene Silencing. *Nat. Struct. Mol. Biol.* **2009**, *16*, 763–768. [[CrossRef](#)]
68. Saze, H.; Scheid, O.M.; Paszkowski, J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **2003**, *34*, 65–69. [[CrossRef](#)]
69. Ramos, M.J.N.; Rocheta, M.; Carvalho, L.C.; Inácio, V.; Graça, J.; Cecilio, L.M. Expression of DNA methyltransferases is involved in *Quercus suber* cork quality. *Tree Genet. Genomes* **2013**, *9*, 1481–1492. [[CrossRef](#)]
70. Inácio, V.; Martins, M.T.; Graça, J.; Cecilio, L.M. Cork Oak Young and Traumatic Periderms Show PCD Typical Chromatin Patterns but Different Chromatin-Modifying Genes Expression. *Front. Plant Sci.* **2018**, *9*. [[CrossRef](#)]
71. Cao, X.; Jacobsen, S.E. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16491–16498. [[CrossRef](#)] [[PubMed](#)]
72. Jackson, J.P.; Lindroth, A.M.; Cao, X.; Jacobsen, S.E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **2002**, *416*, 556–560. [[CrossRef](#)] [[PubMed](#)]
73. Johnson, L.M.; Bostick, M.; Zhang, X.; Kraft, E.; Henderson, I.R.; Callis, J.; Jacobsen, S.E. The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Curr. Biol.* **2007**, *17*, 379–384. [[CrossRef](#)] [[PubMed](#)]

74. Johnson, L.M.; Law, J.A.; Khattar, A.; Henderson, I.R.; Jacobsen, S.E. SRA-Domain Proteins Required for DRM2-Mediated De Novo DNA Methylation. *PLoS Genet.* **2008**, *4*, e1000280. [[CrossRef](#)]
75. Johnson, L.M.; Cao, X.; Jacobsen, S.E. Interplay between Two Epigenetic Marks. *Curr. Biol.* **2002**, *12*, 1360–1367. [[CrossRef](#)]
76. Malagnac, F.; Bartee, L.; Bender, J. An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **2002**, *21*, 6842–6852. [[CrossRef](#)]
77. Du, J.; Johnson, L.M.; Groth, M.; Feng, S.; Hale, C.J.; Li, S.; Vashisht, A.A.; Wohlschlegel, J.A.; Patel, D.J.; Jacobsen, S.E. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol. Cell* **2014**, *55*, 495–504. [[CrossRef](#)]
78. Inácio, V.; Barros, P.M.; Costa, A.; Roussado, C.; Gonçalves, E.M.F.; Costa, R.L.; Graça, J.; Oliveira, M.M.; Cecilio, L.M. Differential DNA Methylation Patterns Are Related to Phellogen Origin and Quality of *Quercus suber* Cork. *PLoS ONE* **2017**, *12*, e0169018. [[CrossRef](#)]
79. Kim, Y.J.; Wang, R.; Gao, L.; Li, N.; Xu, C.; Mang, H.; Jeon, J.; Chen, X.; Zhong, X.; Kwak, J.M.; et al. POWERDRESS and HDA9 interact and promote histone H3 deacetylation at specific genomic sites in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14858–14863. [[CrossRef](#)]
80. Luo, M.; Tai, R.; Yu, C.-W.; Yang, S.; Chen, C.-Y.; Lin, W.-D.; Schmidt, W.; Wu, K. Regulation of flowering time by the histone deacetylase HDA 5 in Arabidopsis. *Plant J.* **2015**, *82*, 925–936. [[CrossRef](#)]
81. Liu, X.; Chen, C.-Y.; Wang, K.-C.; Luo, M.; Tai, R.; Yuan, L.; Zhao, M.; Yang, S.; Tian, G.; Cui, Y.; et al. PHYTOCHROME INTERACTING FACTOR3 associates with the histone deacetylase HDA15 in repression of chlorophyll biosynthesis and photosynthesis in etiolated Arabidopsis seedlings. *Plant Cell* **2013**, *25*, 1258–1273. [[CrossRef](#)] [[PubMed](#)]
82. Xu, L.; Zhao, Z.; Dong, A.; Soubigou-Taconnat, L.; Renou, J.-P.; Steinmetz, A.; Shen, W. Di- and Tri- but Not Monomethylation on Histone H3 Lysine 36 Marks Active Transcription of Genes Involved in Flowering Time Regulation and Other Processes in Arabidopsis thaliana. *Mol. Cell. Biol.* **2007**, *28*, 1348–1360. [[CrossRef](#)] [[PubMed](#)]
83. Zhao, Z.; Yu, Y.; Meyer, D.; Wu, C.; Shen, W. Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nature* **2005**, *7*, 1256–1260. [[CrossRef](#)] [[PubMed](#)]
84. Kim, S.Y.; He, Y.; Jacob, Y.; Noh, Y.-S.; Michaels, S.; Amasino, R.M. Establishment of the Vernalization-Responsive, Winter-Annual Habit in Arabidopsis Requires a Putative Histone H3 Methyltransferase. *Plant Cell* **2005**, *17*, 3301–3310. [[CrossRef](#)]
85. Grini, P.; Thorstensen, T.; Alm, V.; Vizcay-Barrena, G.; Windju, S.S.; Jørstad, T.S.; A Wilson, Z.; Aalen, R.B. The ASH1 HOMOLOG 2 (ASHH2) Histone H3 Methyltransferase Is Required for Ovule and Anther Development in Arabidopsis. *PLoS ONE* **2009**, *4*, e7817. [[CrossRef](#)]
86. Yun, J.-Y.; Tamada, Y.; Kang, Y.E.; Amasino, R.M. Arabidopsis Trithorax-Related3/Set Domain Group2 Is Required for the Winter-Annual Habit of Arabidopsis Thaliana. *Plant Cell Physiol.* **2012**, *53*, 834–846. [[CrossRef](#)]
87. Napsucially-Mendivil, S.; Alvarez-Venegas, R.; Shishkova, S.; Dubrovsky, J. Arabidopsis homolog of trithorax1 (ATX1) is required for cell production, patterning, and morphogenesis in root development. *J. Exp. Bot.* **2014**, *65*, 6373–6384. [[CrossRef](#)]
88. Saleh, A.; Alvarez-Venegas, R.; Yilmaz, M.; Hou, G.; Sadler, M.; Al-Abdallat, A.; Xia, Y.; Lu, G.; Ladunga, I.; Avramova, Z. The Highly Similar Arabidopsis Homologs of Trithorax ATX1 and ATX2 Encode Proteins with Divergent Biochemical Functions. *Plant Cell* **2008**, *20*, 568–579. [[CrossRef](#)]
89. Yao, X.; Feng, H.; Yu, Y.; Dong, A.; Shen, W. SDG2-Mediated H3K4 Methylation Is Required for Proper Arabidopsis Root Growth and Development. *PLoS ONE* **2013**, *8*, e56537. [[CrossRef](#)]
90. Lee, K.; Park, O.-S.; Seo, P.J. ATXR2 as a core regulator of de novo root organogenesis. *Plant Signal. Behav.* **2018**, *13*, e1449543–e1449605. [[CrossRef](#)]
91. Qian, S.; Wang, Y.; Ma, H.; Zhang, L. Expansion and Functional Divergence of Jumonji C-Containing Histone Demethylases: Significance of Duplications in Ancestral Angiosperms and Vertebrates. *Plant Physiol.* **2015**, *168*, 1321–1337. [[CrossRef](#)] [[PubMed](#)]
92. Gan, E.S.; Xu, Y.; Wong, J.-Y.; Goh, J.G.; Sun, B.; Wee, W.-Y.; Huang, J.; Ito, T. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. *Nat. Commun.* **2014**, *5*, 5098. [[CrossRef](#)] [[PubMed](#)]

93. Liu, X.; Wei, W.; Zhu, W.; Su, L.; Xiong, Z.; Zhou, M.; Zheng, Y.; Zhou, D.-X. Histone Deacetylase AtSRT1 Links Metabolic Flux and Stress Response in Arabidopsis. *Mol. Plant* **2017**, *10*, 1510–1522. [[CrossRef](#)] [[PubMed](#)]
94. Huang, L.; Sun, Q.; Qin, F.; Li, C.; Zhao, Y.; Zhou, D.X. Down-Regulation of a Silent Information Regulator2-Related Histone Deacetylase Gene, OsSRT1, Induces DNA Fragmentation and Cell Death in Rice. *Plant Physiol.* **2007**, *144*, 1508–1519. [[CrossRef](#)]
95. Zemach, A.; Kim, M.Y.; Hsieh, P.-H.; Coleman-Derr, D.; Eshed-Williams, L.; Thao, K.; Harmer, S.; Zilberman, D. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **2013**, *153*, 193–205. [[CrossRef](#)]
96. Hume Stroud, T.D.; Du, J.; Zhong, X.; Feng, S.; Johnson, L.; Patel, D.J.; Jacobsen, S.E. The Roles of Non-CG Methylation in Arabidopsis. *Nat. Struct. Mol. Biol.* **2014**, *21*, 64. [[CrossRef](#)]
97. Hartl, M.; Füll, M.; Boersema, P.J.; Jost, J.O.; Kramer, K.; Bakirbas, A.; Sindlinger, J.; Plöckinger, M.; Leister, D.; Uhrig, R.G.; et al. Lysine acetylome profiling uncovers novel histone deacetylase substrate proteins in Arabidopsis. *Mol. Syst. Biol.* **2017**, *13*, 949. [[CrossRef](#)]
98. Ebbs, M.L.; Bender, J. Locus-Specific Control of DNA Methylation by the Arabidopsis SUVH5 Histone Methyltransferase. *Plant Cell* **2006**, *18*, 1166–1176. [[CrossRef](#)]
99. Thompson, J.D.; Higgins, D.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)]
100. Hall, B.G. Building Phylogenetic Trees from Molecular Data with MEGA. *Mol. Biol. Evol.* **2013**, *30*, 1229–1235. [[CrossRef](#)]
101. Leal, J.P.; Abreu, I.A.; Alabaça, C.S.; Almeida, M.H.; Almeida, P.; Almeida, T.; Amorim, M.I.; Araújo, S.; Azevedo, H.; Badia, A.; et al. A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. *BMC Genom.* **2014**, *15*, 371. [[CrossRef](#)]
102. Criscuolo, A.; Brisse, S. AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genome* **2013**, *102*, 500–506. [[CrossRef](#)] [[PubMed](#)]
103. Shrestha, R.K.; Lubinsky, B.; Bansode, V.B.; Moinz, M.B.; McCormack, G.P.; Travers, S. QTrim: A novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform. *BMC Bioinform.* **2014**, *15*, 33. [[CrossRef](#)] [[PubMed](#)]
104. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Mass Genom.* **2009**, *25*, 1754–1760. [[CrossRef](#)]
105. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2013**, *30*, 923–930. [[CrossRef](#)]
106. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 31. [[CrossRef](#)]

