

Genome Wild Analysis and Molecular Understanding of the Aquaporin Diversity in Olive Trees (Olea Europaea L.)

Mohamed Faize, Boris B. Fumanal, Francisco Luque, Jorge A Ramírez-Tejero, Zhi Zou, Xueying Qiao, Lydia Faize, Aurélie Gousset-Dupont, Patricia Roeckel-Drevet, Philippe Label, et al.

▶ **To cite this version:**

Mohamed Faize, Boris B. Fumanal, Francisco Luque, Jorge A Ramírez-Tejero, Zhi Zou, et al.. Genome Wild Analysis and Molecular Understanding of the Aquaporin Diversity in Olive Trees (Olea Europaea L.). International Journal of Molecular Sciences, 2020, 2020 (11), 10.3390/ijms21114183 . hal-02928619

HAL Id: hal-02928619
https://hal.inrae.fr/hal-02928619

Submitted on 2 Sep 2020

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

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



Distributed under a Creative Commons Attribution 4.0 International License

International Journal of
Molecular Sciences

MDPI

Article

Genome Wild Analysis and Molecular Understanding of the Aquaporin Diversity in Olive Trees (*Olea Europaea* L.)

Mohamed Faize [1,*], Boris Fumanal [2], Francisco Luque [3], Jorge A. Ramírez-Tejero [3], Zhi Zou [4], Xueying Qiao [4], Lydia Faize [5], Aurélie Gousset-Dupont [2], Patricia Roeckel-Drevet [2], Philippe Label [2] and Jean-Stéphane Venisse [2,*]

1 Laboratory of Plant Biotechnology, Ecology and Ecosystem Valorization, Faculty of Sciences, University Chouaib Doukkali, El Jadida 24000, Morocco
2 Université Clermont Auvergne, INRAE, PIAF, 63000 Clermont-Ferrand, France; boris.fumanal@uca.fr (B.F.); aurelie.gousset@uca.fr (A.G.-D.); patricia.drevet@uca.fr (P.R.-D.); philippe.label@inrae.fr (P.L.)
3 Department of Experimental Biology, Center for Advanced Studies in Olive Grove and Olive Oils, University of Jaén, 23071 Jaén, Spain; fjluque@ujaen.es (F.L.); jrtejero@ujaen.es (J.A.R.-T.)
4 Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, China; zouzhi2008@126.com (Z.Z.); roslinda@lgm.gov.my (X.Q.)
5 Group of Fruit Tree Biotechnology, Department of Plant Breeding, Murcia University, CEBAS CSIC, 30100 Murcia, Spain; lbremaud@cebas.csic.es
* Correspondence: faizemohamed@yahoo.fr (M.F.); j-stephane.venisse@uca.fr (J.-S.V.); Tel.: +212(0)6-36-87-60-48 (M.F.); +33(0)4-73-40-79-42 (J.-S.V.)

Received: 16 May 2020; Accepted: 3 June 2020; Published: 11 June 2020



Abstract: Cellular aquaporin water channels (AQPs) constitute a large family of transmembrane proteins present throughout all kingdoms of life, playing important roles in the uptake of water and many solutes across the membranes. In olive trees, AQP diversity, protein features and their biological functions are still largely unknown. This study focuses on the structure and functional and evolution diversity of AQP subfamilies in two olive trees, the wild species *Olea europaea* var. *sylvestris* (*Oeu*AQPs) and the domesticated species *Olea europaea* cv. Picual (*Oleur*AQPs), and describes their involvement in different physiological processes of early plantlet development and in biotic and abiotic stress tolerance in the domesticated species. A scan of genomes from the wild and domesticated olive species revealed the presence of 52 and 79 genes encoding full-length AQP sequences, respectively. Cross-genera phylogenetic analysis with orthologous clustered *Olea*AQPs into five established subfamilies: PIP, TIP, NIP, SIP, and XIP. Subsequently, gene structures, protein motifs, substrate specificities and cellular localizations of the full length *Olea*AQPs were predicted. Functional prediction based on the NPA motif, ar/R selectivity filter, Froger's and specificity-determining positions suggested differences in substrate specificities of Olea AQPs. Expression analysis of the *Oleur*AQP genes indicates that some genes are tissue-specific, whereas few others show differential expressions at different developmental stages and in response to various biotic and abiotic stresses. The current study presents the first detailed genome-wide analysis of the AQP gene family in olive trees and it provides valuable information for further functional analysis to infer the role of AQP in the adaptation of olive trees in diverse environmental conditions in order to help the genetic improvement of domesticated olive trees.

Keywords: aquaporin; *Olea europaea*; RNA-seq; functional diversity; *Oleaceae* evolution; plant domestication; *Verticillium dahliae*; cold and wound stress

1. Introduction

Olive trees (*Olea europaea* L.) are among the most important crops around the Mediterranean basin. They are important not only as plant cultivation from a commercial point of view across the millennia, largely due to their multiple uses (e.g., oil, wood, ornamental uses, canned fruit, medicinal applications), but also to date, as a model to study oil fruit physiology. Its remarkable aptitude to face scarce precipitations and high summer temperatures enables it to grow in very diverse environments [1]. Most cultivars of interest show a high intraspecific genetic diversity, resulting, notably, in a wide range of contrasted stress responses. In this respect, identifying the molecular mechanisms underlying various stress tolerances in this species is essential.

Fundamentally, the effects of adverse growth conditions on plants vary among species. Most of the time, they lead to typical symptoms of tissue water imbalances. These deviant physiological situations trigger several highly specific and rapid hydraulic responses to readjust the internal water status (or the water homeostasis). Interestingly, the underlying molecular regulations have been mainly attributed to the action of aquaporins [2]. Indeed, several lines of data indicate that the abiotic and biotic stresses regulate the transcript accumulation, post-translational modifications and localization of various aquaporins, which correlate with the modulation of several ecophysiological traits such as osmotic potentials or hydraulic conductivities [3].

Aquaporins (AQPs) are proteins that are part of all biomembranes, i.e., the plasma membrane and the intracellular membrane system with the morphoplasm [4]. Whatever the hydraulic models and their levels of complexity (from the cell to the organism), AQPs take part in many physiological processes as key elements for the plant development and the cell responses to a plethora of environment stresses. To achieve this, AQP are players of prime importance in the regulation of hydraulic conductances, aided by their quick reaction time (seconds to minutes). In that respect, AQPs maintain the cell water homeostasis of cell turgor, by drawing water either from the apoplasm or from the vacuole (the main cell water capacitor), to buffer any changes in volume and osmolarity of the cytoplasm.

The typical AQP proteins consist of six trans-alpha helical transmembrane regions (TMH1-TMH6) connected by five loops (Loops A–Loops E, LA–LE). The carboxylic and amino terminals (*C*-t, and *N*-t, resp.) on either side of the primary sequence lie on the cytoplasmic side. The AQP protein monomer structure is highly conserved and resembles a tridimensional "hourglass model" [4]. Given the general conservation in such a 3D design, several amino acids along the primary sequence interact together in the folded structure, creating specific motifs that are important for the organization of constricts and then pore selectivity.

The first constrict consists of two additional half-helices (HB and HE), which form the seventh TM helix by the opposite LB and LE dipping into the membrane. Each half-helix possesses a highly conserved "asparagine–proline–alanine" (NPA) motif, forming a narrow region in the middle of the pore. The two asparagine side chains pointing into the pore are located at the end of these two half-helices, creating a dipole water movement and a proton (H^+) exclusion environment. However, the general rule is that the size of this NPA constrict limits the size of the potentially transportable substrates [5]. The second constrict, known as "aromatic/arginine" or ar/R selectivity filter, consists of four residues (F58-H182-C191-R197 in AQP1). It is located towards the extracellular exit of the channel, approximately 8 Å away from the NPA regions. The ar/R constriction region forms the narrowest part of the pore, and it is therefore generally assumed that the discrepancies in its size and its hydrophobicity determine the solute transport specificity [6,7]. Similarly, five conserved Froger's residues (P1–P5 residues, T116-S196-A200-F212-W213 in AQP1) have been found to play critical roles in regulating substrate specificity [6]. Moreover, it has been predicted that AQPs have nine specificity-determining positions (SDPs) for non-aqua substrates for each unique group [6]. All these motifs or positions of particular amino acids are viewed as key elements for the selection and the transport of permeants across the cell membranes.

From a tridimensional viewpoint, AQPs are arranged as homo- and/or heterotetrameric structures, which usually provide greater stability and folding of the proteins, resulting in highly efficient water

transport across biological membranes [8]. A fifth channel forms through the middle of the tetramer array, which has been suggested to conduct gases, such as CO_2 [9]. In addition to water, AQPs facilitate the transport of a variety of solutes, including boron, silicon, arsenite, ammonia, glycerol, polyols and urea, some ions (nitrate, halide) and reactive oxygen species (hydrogen peroxide) [10,11]. This ability to transport different solutes makes it possible to clearly identify two AQP subfamilies, the ordinary or orthodox aquaporins (AQP, *stricto sensu*), which exclusively transport water, and the aquaglyceroporins (AGP or AQGP), which are additionally permeable for these small organic and inorganic compounds [12]. Aquaporins are a ubiquitous family of intrinsic membrane proteins present throughout all kingdoms of life, from archae to eukaryotes (plants, animals, fungi, insects and protists). The plant kingdom exhibits remarkable MIP diversity, which can be classified into eight subfamilies: the plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodulin 26-like intrinsic protein (NIPs), small basic intrinsic proteins (SIPs), X-intrinsic proteins (XIPs), GlpF-like intrinsic proteins (GIPs), hybrid intrinsic proteins (HIPs) and large intrinsic proteins (LIPs) [13–15]. To date, only three full-length aquaporin sequences belonging to PIP and TIP subfamilies have been described in olive trees: *Oe*PIP1;1, *Oe*PIP2;1 and *Oe*TIP1;1 [16]. As observed in other tree species, the transcript levels of these aquaporins have been reported to be differentially down-regulated in leaves, roots and twigs from olive trees (*O. europaea* L.) subjected to drought treatment. However, a comprehensive analysis of each aquaporin member and their responses in different tissues or under various abiotic and biotic stresses is lacking.

The availability of a whole genome sequence from the wild olive species *O.* var. *sylvestris* facilitates the genome-wide analysis to identify the complete set of AQPs in *Olea* (*Oeu*AQPs) and to understand the evolutionary relationship of *Olea*'s AQPs by comparing *Oeu*AQPs with their homologuous *Oleur*AQPs from the domesticated cultivar, *O. europaea* cv. Picual. Of Andalusian origin, cv. Picual is the leading olive cultivar planted in Spain, where it accounts for 50% of national production and 20% of world production. Extra virgin olive oil from cv. Picual offers exceptional organoleptic properties and high oxidative stability due to its elevated polyphenolic compounds content. Spurred by numerous health benefits, the world demand for such a precious vegetable oil is expected to increase in the forthcoming years.

Olea europaea cv. Picual displays a great capacity to adapt to a wide variety of soils and climatic conditions. It allows this cultivar to be cultivated in arid plains as well as in mountain areas and to withstand very cold winters. However, like most domesticated cultivar, cv. Picual is susceptible to various plant pathogenic agents, such as the Ascomycota *Verticillium dahliae*, the causal agent of Verticillium wilt [17]. The involvement of AQPs in cellular responses to tolerate abiotic (cold and wound) and biotic (*V. dalhiae* infestation) stresses needs to be resolved. In this respect, gene expression is a major step towards the gene elucidation of its biological function, and combined with the motif patterns and related predictive substrate permeability, the plant localization of aquaporin and expression profiles provide important clues to reveal their potential physiological roles in a plant.

In this work, we performed a genome-wide analysis of *Olea*AQP gene family members in wild (*Oeu*AQPs) and cultivated (*Oleur*AQPs) olive trees, and we characterized their phylogeny, chromosomal distribution, gene structure, protein motifs and potential substrate specificities. We investigated the transcriptional expression profile of *OleurAQP* genes using RNA-seq in various organs and tissues, during the early stages of olive-seedling development and the emergence of the juvenile plant and in response to biotic (*Verticillium dahliae* root inoculation) and abiotic (root wounding and leaf chilling) stresses. The present research will contribute to bringing novel and relevant information with regards to AQP evolution and their functional regulation in olive trees during various physiological processes of development and stress responses. Our exhaustive molecular understanding on *Olea* AQPs is expected to be helpful in elucidating the importance of the MIP superfamily in a commercially eminent oleaginous crop.

2. Results and Discussion

Climate changes have been highlighted as one of the most harmful events affecting the crop growth and productivity. Simultaneously, the present environmental conditions and agricultural practices create a fertile breeding ground for the spreading of a plethora phytopathogen agents. All these concomitant abiotic and biotic stresses result in significant agricultural losses worldwide. However, whatever the challenging stressors, plants must constantly adjust their water-balance control for normal development and maintenance of proper yield.

Olea europaea L. is part of the Lamiales, an Asterid order of core eudicots. To date, a complete picture of the AQP family from the Lamiales taxa is still missing. In this regard, the available draft genome sequence of olive trees provided us with an opportunity to analyze the AQP gene family in an oleaginous species and to assess their diversification in the *Olea* genus, and *vis-a-vis* its related Oleaceae common ash (*F. excelsior*) and some least related taxa including tomato, poplar and *Arabidopsis*.

2.1. Identification of the Complete Set of O. Europaea L. and F. Excelsior Aquaporins

The "Build your query" function including "Major Intrinsic Protein" and "Aquaporin" in Phytozome (V12) were used to identify the complete non-redundant AQP family of the wild olive *O. europaea* var. *sylvestris*. To investigate whether potential *Oeu*AQP genes were lacking or insufficiently annotated in the databases, the genomes of the cultivar olive *O. europaea* L. cv. Picual (OliveTreeDB website) and the Oleaceae related species *F. excelsior* (hardwoodgenomics website) were BLAST-searched using several AQP members belonging to each subfamily from *A. thaliana*, *L. esculentum* and *P. trichocarpa* as queries. Identified AQP sequences from each Oleaceae species were thereafter iteratively used as queries. Subsequently, every output putative AQP gene was manually inspected. This approach demonstrated that for the database-annotated AQP genes, output coding sequences and exon/intron selections sometimes resulted in a panel of non-satisfactory AQP gene models. With the exception of *Oeu*PIP2;11 and *Oeu*NIP5;2 sequences for which a manual annotation procedure generated very long AQP sequences, thirty-six non-satisfactory AQP gene models encode partial sequences due to an incomplete or miss-annotation procedure alongside the chromosomes or the scaffolds. These situations are linked to deletions or mutations that result in a shift in the reading frame leading to the synthesis of premature stop codons or non-functional proteins (e.g., *Oeu*NIP4;2). Wherever possible, each erroneous AQP sequence was manually corrected using orthologs from *F. excelsior* and *L. esculentum*, with the addition of AQPs from *P. trichocarpa* for the examination of the XIP subfamily. This last approach helped identify twenty new candidates encoding complete AQP sequences meeting the typical AQP features. Similarly, the alignment of the incomplete putative *Oeu*AQP sequences due to their scaffold edge location with orthologous counterparts gave three complete *Oeu*AQP sequences (*Oeu*PIP2;9, *Oeu*PIP2;10 and *Oeu*TIP4;1). Two *OeuAQP* genes were not detected and annotated in Phytozome databases (*OeuSIP2;2* on chromosome 11 and the incomplete sequence *OeuNIP4;2-related* on chromosome 8). Lastly, two raw *OeuPIP2*-related gene sequences initially emerged as two distinct 3′ and 5′ truncated AQP candidates (Oeu017254.1 and Oeu017256.1, respectively). By homology with orthologous PIP2s, these two sequences were gathered in one sequence to generate the putative *Oeu*PIP2;11 in its full-length version. However, the last intron (19421nt of size) enclosed the *Oeu*017255 gene sequence (uncharacterized protein). This sequence was retained for computational analysis since *Oleur*PIP2;11 from the domesticated olive species shares similar features.

In summary, 68 putative AQP sequences were identified in the wild variety, *O.* var. *sylvestris*. Thereof, the manual inspections and curating procedures showed that 52 sequences encode for typical full-length AQP proteins, namely, thereafter, *Oeu*AQPs. Sixteen AQP proteins had to be discarded for various reasons, and every identified *Oeu*AQP sequence is detailed in its 'original' version and its eventual corrected version in Figure S1 and Table S1.

In the domesticated olive cultivar 'Picual', 127 *Oleur*AQP protein sequences were identified. However, this high abundance needs to be put into perspective with 79 full-length and putative full-length candidates and 51 incomplete genes or pseudogenes sequences. Among the 79 full-length

candidates, 26 sequences are in single copy and 26 AQP-types are duplicated, the latter generating 27 full-length sequences and 3 pseudogenes. As a result, 53 full-length AQP-types are identified in the cultivated genome, a similar AQP level was recorded in the wild variety. The *Oleur*AQP superfamily expansion and diversity will be further discussed in a specific chapter. *Oleur*AQP sequences and their correspondence with the *Oeu*AQP sequences are detailed in Figure S2 and Table S2, respectively.

Concerning *F. excelsior*, 56 putative *Fraex*AQP sequences were extracted, including 54 full-length sequences used for comparative analyses of *Olea* AQP. Sequences were extracted and verified as for *Olea* AQP sequences and specifically annotated for this work (Figure S3).

These numbers are comparable to those of species exhibiting the largest number of members. In comparison, the basal viridiplantae *Selaginella moellendorffii* and *Physcomitrella patens* show 19 and 23 AQPs, respectively. The monocotyledons *Hordeum vulgare*, *Oryza sativa* and *Zea mays* show 40, 34 and 43 AQPs, respectively. In the Rosids, the Malvids *A. thaliana* and *Brassica rapa* show 35 and 47 AQPs, and the Fabids *P. trichocarpa*, *Glycine max* and *Gossypium hirsutum* show 58, 66 and 71 members, respectively. Lastly, among the Lamiids, the two Solanales *S. tuberosum* and *S. lycopersicum* show 41 and 47 members.

2.2. Diversity in OeuAQP Gene Families from the Wild Olive Tree Variety, O. Europaea Var. Sylvestris

A total of 52 full-length *Oeu*AQP encoding genes were identified in the genome of the wild olive. In accordance with most of the high plant genomes, the *Oeu*AQPs were classified in five distinct subfamilies: *Oeu*PIPs (19 members), *Oeu*TIPs (17 members), *Oeu*NIPs (9 members), *Oeu*XIPs (3 members) and *Oeu*SIPs (4 members) (Figure 1 and Figure S4). Furthermore, these subfamilies can be sub-grouped based on the analysis of various motifs (i.e., the aromatic/arginine selective filters and the Froger's residues, as detailed below) and the phylogenetic positioning that well corroborate each other (Figure 2a). Finally, the further classification of *Oeu*AQP subfamilies into subgroups shows that the var. *sylvestris* (as to the domesticated cultivar Picual or the related Oleaceae *F. excelsior*) does not contain additional phylogenetic AQP subgroups within the five subfamilies formed due to the divergence of the Asterids taxa within the core eudicots. Accordingly, *Oeu*AQP classification is consistent with AQPs from *Arabidopsis* or tomato, implying that AQP proteins have been highly conserved during evolution. In detail, the larger *Oeu*PIPs subfamily is divided in two distinct subgroups, *Oeu*PIP1 and *Oeu*PIP2, with eight and eleven members in each group, respectively.

Similarly, the *Oeu*TIP subfamily clusters in five subgroups (*Oeu*TIP1 to 5), where the *Oeu*TIP1 subgroup encompasses the greatest number of members. The subfamily *Oeu*NIPs consists of five subgroups, where the *Oeu*NIP1 subgroup encompasses the greatest number of members. No NIP2 subgroup was identified in the cultivated and wild olive trees. However, it should be noticed that the classification of this subgroup is relatively ambiguous, as observed with *Sl*NIP2;1 or *At*NIP2;1 that coalesce with various NIP4 or NIP1 members, respectively (Figure 1 and Figure S4). Using the recommendations proposed by Zou and collaborators [18], the NIP2 subgroup could specifically feature the ar/R selectivity filter (G/S/G/R), as can be observed in *S. lycopersicon* (*Sl*NIP2;1) [19], *P. trichocarpa* (*Pt*NIP2;1) [20], *Ricinus communis* (*Rc*NIP2;1) [21], *Cicer arietinum* (*Gm*NIP2;1) [22], *Jatropha curcas* (*Jc*NIP2;1) [18] and *Arachis duranensis* (*Adu*NIP2;1) [23]. This motif is distanced by 108 amino acids to the NPA domains, playing a crucial role in Si absorption in various plants species [24]. However, interesting enough, the *F. excelsior* genome has a NIP2 candidate, *Fraex*NIP2;1. This suggests that a functional divergence occurred in the Oleaceae NIP subfamily, notably by a loss of the NIP2 subgroup during the evolution of the *Olea* clade.

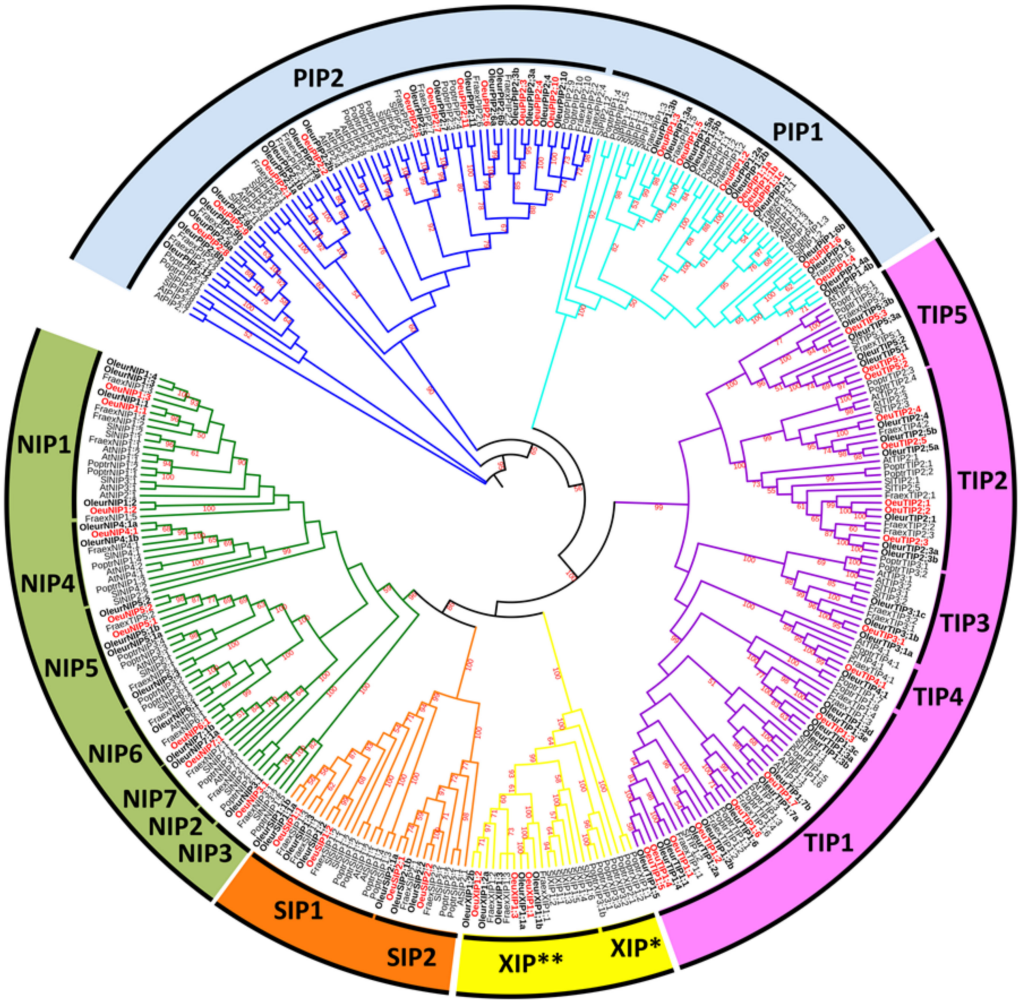


Figure 1. Phylogenetic analysis of protein sequences of the 52 *Oeu*AQPs and the 79 OleurAQPs with *Fraxinus excelsior*, *Arabidopsis thaliana* and *Populus trichocarpa* hortologs. Deduced amino acid sequences were aligned using ClustalW, and the phylogenetic tree was constructed using the maximum parsimony method. Maximum parsimony analysis was conducted using the subtree-pruning-regrafting algorithm. The number next to the branches represents bootstrap values ≥50% based on 5000 resampling. The distance scale denotes the number of amino acid substitutions per site. The AQP sequences from *O. europaea* var. *sylvestris*, *O. europaea* cv. Picual, *F. excelsior*, *A. thaliana* and *P. trichocarpa* are preceded by the prefixes Oeu, Oleur, Fraex, At and Poptr, respectively. *Oeu*AQPs are mentioned in red, and *Oleur*AQP in bold. The name of each subfamily and subgroup is indicated next to the corresponding group. XIP* and XIP** correspond to XIP sequences from Rosids and Asterids, respectively. *Oeu*AQP accession numbers and sequences are listed in and Figures 3 and S1. *Oleur*AQP and *Fraex*AQP accession numbers and sequences are listed in Figures S2 and S3, respectively. This Figure 1 is also presented in Supplemental Figure S4 with high definition for ease of reading.

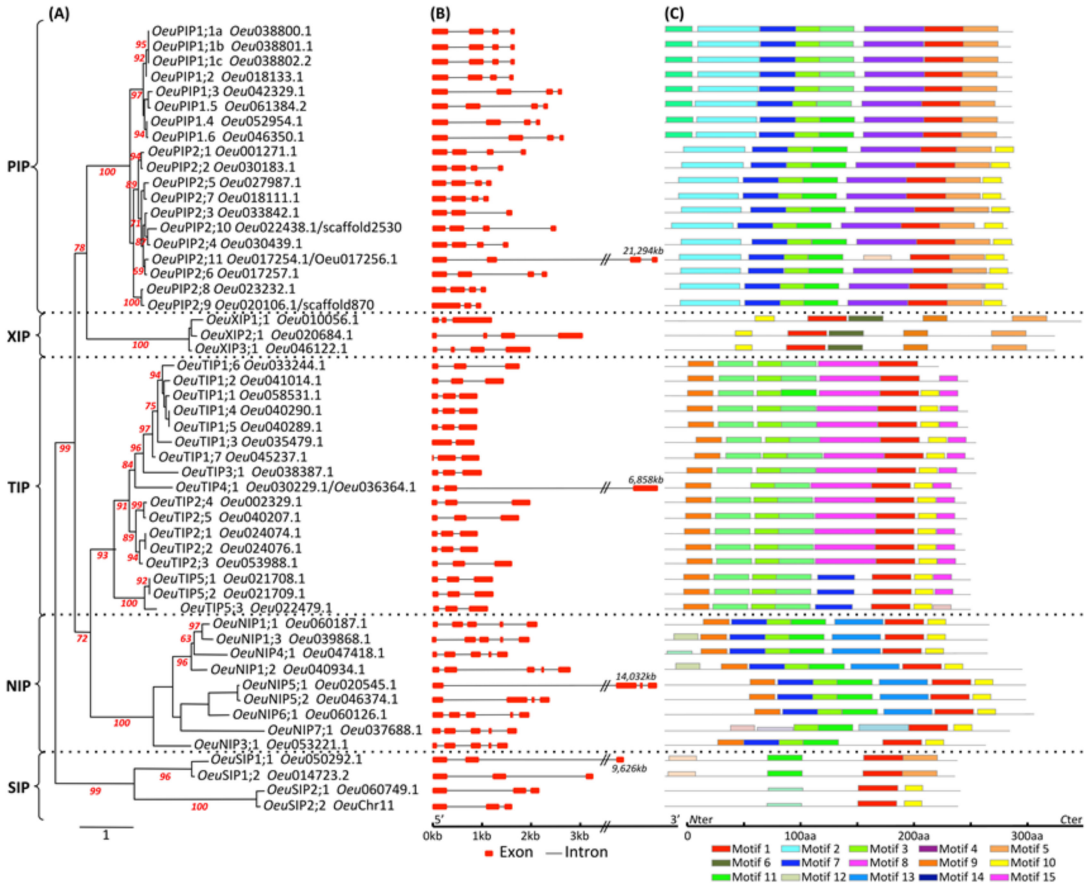


Figure 2. (**A**) Phylogenetic relationship, (**B**) Exon/Intron genomic structure, and (**C**) protein motif organization of the 52 full-length *Olea europaea* var. *sylvestris* aquaporin sequences. (**A**) *Oleu*AQP Sequences were aligned with MUSCLE, and the unrooted phylogenetic tree was reconstructed using the maximum likelihood method. The numbers at the nodes represent the percentage bootstrap values (only those 50% were represented). The reliability for the internal branch was assessed using 1000 bootstrap replicates. The distance scale denotes the number of amino acid substitutions per site. *Oleu*AQPs clustered into five AQP subfamilies: *Oleu*PIPs, *Oleu*XIPs, *Oleu*TIPs, *Oleu*NIPS and *Oleu*SIPs. (**B**) Exons and introns of the *Oleu*AQP genes are represented by red boxes and black lines, respectively. Gene structures were compared using GSDS software. Gene orientations are indicated (5′—3′) in the x-axis. (**C**) Distribution of the conserved motifs among the *Oleu*AQP proteins. Motif analysis was performed by using the MEME web server. Fifteen conserved motifs were identified, and the different motifs are identified using different colored boxes, as indicated at the bottom of the figure. Each color block in the different proteins indicates a specific motif, for which the amino acids are detailed in Figure S6. Protein orientations are indicated (*N*ter-Cter) on the x-axis.

The *Oeu*SIP subfamily clusters the four *Oeu*SIP homologues in two subgroups, *Oeu*SIP1 and *Oeu*SIP2, with two members each. (Figure 1, Figures S3 and S4). This modest number of *Oeu*SIP members is comparable to the SIP number identified in other plant species. However, the *Oeu*XIP1 subgroup features a pseudogenization of its SIP1;3-type, unlike the domesticated cv. Picual (*Oleur*SIP1;3) or *F. excelsior* (*Fraex*XIP1;3).

Concerning the *Oeu*XIP members, they are composed of three sequences with no specific partitioning. The presence of XIPs in olive trees is particularly interesting. This XIP subfamily represents the last clade of AQPs that was discovered, mostly in plants and fungi and more marginally in the Amoeboza *Dyctyostelium discoideum*. Although this subfamily is already present in basal plant lineage such as the moss *Physcomitrella patens*, it was lost in several recent plant lineages after the divergence of the basal dicots and core dicots [25,26], where all monocots, coniferophytes and

several dicots taxa (such as *Brassicaceae*) are characterized by its complete absence [27,28]. In olive trees, the phylogenetic positioning of *Olea*XIPs is in line with the diversity observed in Asterids (tomato, potato), which clearly points to that of Rosids, which exhibits at least two or three distinct subgroups [18,19,29]. In this respect, the phylogenetic organization of this XIP subclass clearly differs to that of the other AQP subclasses, clustering the XIP members in distinct taxon-specific clades, i.e., Rosids vs. Asterids, and for this last clade, Lamiales vs. Solanales. Although our comparative analysis was carried out on a small panel of XIP sequences, our results demonstrate that the evolutionary pattern of this XIP subfamily is completely unique in comparison to the other AQP subfamilies [26–30].

Among the *Oeu*AQPs and per subfamily, *Oeu*NIP was the most diverse subfamily with 38% sequence similarity between members at the amino acid level, whereas *Oeu*PIP1 and OeuPIP2s were the most conserved subfamilies with 89% and 79% sequence similarities, respectively (Table S3). As for *Oeu*XIP, *Oeu*TIP and *Oeu*SIP subfamilies, they exhibit 78%, 55% and 45% sequence similarities between members, respectively. Overall, *Oeu*AQP subfamilies show an average of 29% sequences similarity between them. These values are commonly observed in most of the plant MIP superfamilies.

2.3. Diversity in OleurAQP Gene Families from the Domesticated Olive Tree cv. Picual'

*Oleur*AQPs are particularly abundant, with a total of 127 putative *AQP* gene sequences, including 79 full-length (Table S2) that group into the five AQP subfamilies of PIP, TIP, XIP, NIP and SIP. Furthermore, the number of *Oleur*AQP-types in each subfamily is relatively similar to that of the wild variety (Figure 1 and Figure S4). Only two AQP-types supplement those observed in the wild variety: *Oleur*PIP2;12 (a PIP2;8/9-types analogue, also present in *F. excelsior* genome, but a priori lost in the wild variety) and *Oleur*NIP1;4.

The *Oleur*AQP superfamily expansion is linked to the fact that the Olea genome was subjected to two rounds of tetraploidy (62 and 25 million years ago) and the cultivated olive to a nearly present partial duplication event [31,32]. Ultimately, no AQP-type loss after duplication appears to have occurred. As a putative consequence to these genome rearrangements, the number of full-length AQPs in both olive varieties (79 *Oleur*AQPs for 52 *Oeu*AQPs) was well correlated with the genome size and the total number of genes, i.e., (cv. Picual vs. var. *sylvestris*) genome sizes of 1.63/1.81Gbp vs. 1.14 Gbp, encompassing 78079 vs. 50684 protein-coding transcripts. Similar abundances were also reported in other cultivated plant species that evolved with polyploidization, e.g., *Brassica napus* with 120 AQPs or *G. hirsutum* with 113 AQPs [33,34].

Gene duplication concerned all *Oleur*AQP subfamilies, in which 21 full-length *Oleur*AQP-types are represented by two full-length paralogues, except for *Oleur*TIP1;3 with 5 duplicates and *Oleur*TIP3;1 with 3 duplicates (Figure S7; Table S2), generating 27 full-length duplicated genes sequences. On the other hand, 41 *Oleur*AQP incomplete sequences or pseudogenes add to them. Among them, six *Oleur*AQPs are related to *Oeu*AQP incomplete sequences and four pseudogene *Oleur*AQP sequences by introduction of deleterious mutations, probably during the tetraploidization process of the cultivated olive, derived from full-length *Oeu*AQP sequences.

2.4. Structural Features of OeuAQP and OleurAQP Proteins

Plant AQPs present different functional characteristics that are driven by specific biochemical properties [6]. The biochemical properties of the 52 *Oeu*AQPs were predicted, i.e., the molecular weight (MW), protein Isoelectric point (*p*I), grand average of hydropathy (GRAVY) and subcellular localization (Figure 3). *Oeu*AQP proteins range from 240 to 335 (average 272) amino acids in length, molecular weights range from 25.45 to 37.27 (average 28.86) kD, and *p*I values range from 5.12 to 10.04 (average 7.56) (Figure 3). Among the *Oeu*AQPs subfamilies, the lower average *p*I value concerns *Oeu*TIPs (average 5.78). This more acidic pH is mainly linked to the absence of basic residues in the *C*-terminal domains of the proteins, as also observed in *Arabidopsis* [35]. As for the GRAVY parameters, scores were all positive, ranging from 0.332 to 1.035. Positive and negative scores for protein GRAVY reflect protein hydrophobicity and hydrophilicity, respectively. As far as we are concerned here, the positive GRAVY

scores of the *Oeu*AQPs indicate their hydrophobic nature, which, for an aquaporin, is a key property that facilitates high water and substrate permeability across membranes [4]. In detail, the higher GRAVY values concern the *Oeu*TIP subfamily (average 0.82). This indicates that among the *Oeu*AQP subfamilies, the *Oeu*TIPs would present a better interaction with water molecules, with the exception of *Oeu*TIP3;1 and *Oeu*TIP5;1 (0.56 and 0.64, resp.). Interestingly, the two *Oeu*SIP1 members share high GRAVY scores (0.798 and 0.83), which noticeably contrast with *Oeu*SIP2 members that show lower GRAVY values (0.417 and 0.505).

[a]Loci Locus / Proposed gene name	Size (aa)	[b]MW (kDa)	[b]pI	[b]GRAVY	[c]TMH	[d]SubCL	ar/R selectivity filters H2-H5-LE1-LE2	Froger's residues P1-P2-P3-P4-P5	[e]NPA1 (LB)	[e]NPA2 (LE)
Plasma membrane intrinsic proteins (PIPs)										
Oeu038800.1 / *Oeu*PIP1;1a	285	30,46	9,18	0,408	6	PM	F-H-T-R	G-S-A-F-W	NPA	NPA
Oeu038801.1 / *Oeu*PIP1;1b	285	30,46	9,18	0,408	6	PM	F-H-T-R	G-S-A-F-W	NPA	NPA
Oeu038802.2 / *Oeu*PIP1;1c	285	30,46	9,18	0,408	6	PM	F-H-T-R	G-S-A-F-W	NPA	NPA
Oeu018133.1 / *Oeu*PIP1;2	286	30,58	9,2	0,405	6	PM	F-H-T-R	G-S-A-F-W	NPA	NPA
Oeu042329.1 / *Oeu*PIP1;3	287	30,94	8,97	0,404	6	PM-ER	F-H-T-R	M-S-A-F-W	NPA	NPA
Oeu052954.1 / *Oeu*PIP1;4	285	30,48	9,02	0,398	6	PM	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu061384.2 / *Oeu*PIP1;5	289	31,13	8,77	0,349	6	PM	F-H-T-R	M-S-A-F-W	NPA	NPA
Oeu046350.1 / *Oeu*PIP1;6	285	30,48	9,02	0,382	6	PM	F-H-T-R	G-S-A-F-W	NPA	NPA
Oeu030183.1 / *Oeu*PIP2;1	287	30,82	6,29	0,429	6	PM	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu001271.1 / *Oeu*PIP2;2	286	30,72	7,08	0,437	6	PM	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu033842.1 / *Oeu*PIP2;3	289	30,78	8,58	0,498	6	PM	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu030439.1 / *Oeu*PIP2;4	289	30,68	8,28	0,509	6	PM-ER	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu027987.1 / *Oeu*PIP2;5	280	29,89	8,19	0,522	6	PM	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu017257.1 / *Oeu*PIP2;6	288	30,73	8,28	0,488	6	PM	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu018111.1 / *Oeu*PIP2;7	283	30,17	7,61	0,539	6	PM-Vac	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu023232.1 / *Oeu*PIP2;8	283	30,35	8,3	0,527	6	PM	F-H-T-R	M-S-A-F-W	NPA	NPA
Oeu020106.1/Scaffold870 / *Oeu*PIP2;9	283	30,26	8,83	0,511	6	PM-Vac	F-H-T-R	M-S-A-F-W	NPA	NPA
Oeu022438.1/scaffold2530 / *Oeu*PIP2;10	274	29,59	9,48	0,533	6	PM-Vac-ER	F-H-T-R	Q-S-A-F-W	NPA	NPA
Oeu017254.1/Oeu017256.1 / *Oeu*PIP2;11	284	30,7	8,8	0,513	6	PM-ER	F-H-T-R	Q-S-A-F-W	NPA	NPA
Tonoplast intrinsic proteins (TIPs)										
Oeu058531.1 / *Oeu*TIP1;1	252	26,15	5,37	0,865	6 (7)*	PM-Vac	H-I-A-V	T-S-A-Y-W	NPA	NPA
Oeu041014.1 / *Oeu*TIP1;2	252	25,99	5,15	0,837	6	PM-Vac	H-I-A-V	T-S-A-Y-W	NPA	NPA
Oeu035479.1 / *Oeu*TIP1;3	258	26,86	6,13	0,79	6	Vac-PM	H-I-A-V	T-S-A-Y-W	NPA	NPA
Oeu040290.1 / *Oeu*TIP1;4	252	26,18	5,41	0,841	6	PM-Vac	H-I-A-V	T-S-A-F-W	NPA	NPA
Oeu040289.1 / *Oeu*TIP1;5	252	26,17	5,42	0,852	6	PM-Vac	H-I-A-V	T-S-A-F-W	NPA	NPA
Oeu033244.1 / *Oeu*TIP1;6	227	23,32	7,87	0,746	6	Vac	H-I-A-V	T-S-A-Y-W	NPA	NPA
Oeu045237.1 / *Oeu*TIP1;7	257	26,58	5,36	0,7	6 (7)*	Vac	H-I-G-R	T-S-A-Y-W	NPA	NPA
Oeu024074.1 / *Oeu*TIP2;1	248	25,06	5,58	1,035	6 (7)*	PM-Vac	H-I-G-R	T-S-A-Y-W	NPA	NPA
Oeu024076.1 / *Oeu*TIP2;2	248	25,06	5,58	1,035	6 (7)*	PM-Vac	H-I-G-R	T-S-A-Y-W	NPA	NPA
Oeu053988.1 / *Oeu*TIP2;3	248	25,34	5,33	1,008	6	PM-Vac	H-I-G-R	T-S-A-Y-W	NPA	NPA
Oeu002329.1 / *Oeu*TIP2;4	250	25,36	5,8	0,952	6 (7)*	Vac	H-I-G-R	T-S-A-Y-W	NPA	NPA
Oeu040207.1 / *Oeu*TIP2;5	250	25,45	5,12	0,936	6	Vac	H-I-G-R	T-S-A-Y-W	NPA	NPA
Oeu038387.1 / *Oeu*TIP3;1	258	27,25	6,43	0,565	6 (5)*	PM-Vac	H-V-A-R	T-A-A-Y-W	NPA	NPA
Oeu030229.1 / ***Oeu*036364.1** / OeuTIP4;1	247	25,93	5,57	0,843	5	Vac-PM	H-V-A-R	T-S-A-Y-W	NPA	NPA
Oeu021708.1 / *Oeu*TIP5;1	253	26,29	6,64	0,644	6	PM-Vac	N-V-G-Y	T-S-A-Y-W	NPA	NPA
Oeu021709.1 / *Oeu*TIP5;2	253	26,29	6,64	0,644	6	PM-Vac	N-V-G-Y	T-S-A-Y-W	NPA	NPA
Oeu022479.1 / *Oeu*TIP5;3	252	25,61	6,63	0,729	6	PM-Vac	N-V-G-Y	T-A-A-Y-W	NPA	NPA
Uncharacterized *X* intrinsic proteins (XIPs)										
Oeu010056.1 / *Oeu*XIP1;1	345	37,27	6,82	0,709	6 (7)*	PM	I-L-A-R	V-C-A-F-W	NP***V***	NPA
Oeu020684.1 / *Oeu*XIP1;2	324	34,9	5,88	0,648	6 (7)*	PM-Vac	A-L-A-R	V-C-P-F-W	NP***V***	NPA
Oeu046122.1 / *Oeu*XIP1;3	323	34,79	6,54	0,665	6 (7)*	PM-Vac	A-L-A-R	V-C-A-F-W	NP***V***	NPA
Nodulin-26 like intrinsic proteins (NIPs)										
Oeu060187.1 / *Oeu*NIP1;1	269	28,69	9,37	0,506	6	PM-Vac	W-V-A-R	F-S-A-Y-I	NPA	NPA
Oeu040934.1 / *Oeu*NIP1;2	297	32,13	9,63	0,339	6	PM-Vac	C-I-A-R	F-S-A-Y-I	NPA	NPA
Oeu039868.1 / *Oeu*NIP1;3	267	28,57	9,4	0,637	6	PM-Vac	W-V-A-R	F-S-A-Y-I	NPA	NP***V***
Oeu053221.1a / *Oeu*NIP3;1	266	28,96	8,32	0,332	6	PM-ER	W-/-A-R	F-S-A-Y-V	NPA	NPA
Oeu047418.1 / *Oeu*NIP4;1	267	28,55	5,94	0,713	6	Vac-PM	W-V-A-R	F-S-A-Y-I	NPA	NPA
Oeu020545.1 / *Oeu*NIP5;1	298	31,01	6,58	0,541	6	PM-Vac	A-I-G-R	F-T-A-Y-L	NP***S***	NP***V***
Oeu046374.1 / *Oeu*NIP5;2	298	31,02	8,63	0,517	6	PM-Vac	A-I-G-R	F-T-A-Y-L	NP***S***	NP***V***
Oeu060126.1 / *Oeu*NIP6;1	306	31,85	9,11	0,552	6	PM-Vac	T-I-A-R	F-T-A-Y-L	NP***S***	NP***V***
Oeu037688.1 / *Oeu*NIP7;1	285	30,3	6,51	0,587	6	PM-Vac	A-V-G-R	Y-S-A-Y-L	NPA	NPA
Small basic intrinsic proteins (SIPs)										
Oeu050292.1 / *Oeu*SIP1;1	241	25,34	9,12	0,798	6 (7)*	PM-Vac	V-T-P-N	F-A-A-Y-W	NP***T***	NPA
Oeu014723.2 / *Oeu*SIP1;2	240	25,45	9,1	0,83	6 (7)*	PM-Golgi	V-T-P-N	F-A-A-Y-W	NP***T***	NPA
Oeu060749.1 / *Oeu*SIP2;1	245	26,87	9,78	0,505	6 (5)*	PM-Nucl-Chlo-ER	S-K-G-S	I-V-A-Y-W	NP***L***	NPA
OeuChr11 / *Oeu*SIP2;2	241	26,58	10,04	0,417	6 (5)*	PM-Vac-Nucl	S-K-G-S	I-V-A-Y-W	NP***L***	NPA

Figure 3. Nomenclature and protein properties of the wild species *Olea europaea* var. *sylvestris* aquaporins (complete sequences). [a] Loci, gene IDs and AQP location are based on Phytozome assembly v1.0; [b] MW, protein molecular wight; *p*I, protein isoelectric point; GRAVY, ground average of hydropathy; [c] TMH, number of transmembrane helices predicted by TMHMM and SOSUI analysis tools; * regions adjusted according to alignment with characterized orthologs from Arabidopsis, poplar and Tomato; [d] DubCl, predicted subcellular localization: PM, plasma membrane; Vac, vacuolar; ER, endoplasmic reticulum; Chlo, chloroplast; Nucl, nuclear. [e] NPA: Asparagine, Proline, Alanine; Bold italic letters denote usual amino acids in NPA motifs.

Prediction of TMHs (transmembrane helical regions) showed that the majority of identified putative *Oeu*AQPs contained six TMHs (Figure 3). However, these predictive tools did not correctly identify all TMHs, predicting six TMHs for most AQPs, but not all AQPs for which predictions are in

a few cases contradicting where five or seven TMHs are predicted. That concerns all *Oeu*XIPs and *Oeu*SIPs and six *Oeu*TIP members (marked with an asterisk in Figure 3). Inspection of hydrophobicity plots (data not shown) and the amino acid sequence alignments for each AQP subfamily have therefore been necessary to remove any ambiguity (Figure S5). Given the high degree of sequence conservation between *Oeu*AQPs and different well-characterized orthologs, it is reasonable to suggest that all *Oeu*AQPs feature six TMHs (with the exception of *Oeu*TIP1;6 which presents a truncated 3′ end).

The protein motifs are highly conserved amino acid residues that are considered to possibly have key functional and/or structural roles in active proteins. Concerning aquaporins, the hydrophobicity degree and the size of the side chain of each amino acid engaging within the pores determine substrate specificity. To seek further insights into motif structures and organizations that design the *Oeu*AQPs, the 52 full-length *Oeu*AQPs proteins were analyzed using the MEME program. Fifteen conserved motifs, designated as motif 1 to motif 15, were identified (Figure 2C and Figure S6). Most *Oeu*AQPs proteins belonging to the same subfamily generally harbor similar motifs, but notably differ between subfamilies. This organization is consistent with the phylogenetic distribution (Figures 1 and 2A). Similar patterns of motif conservation and/or deletion between subfamilies/subgroups have been observed in recent genome-wide studies of AQPs in other plants. Of these patterns, the *Oeu*PIP1 and *Oeu*PIP2 members present a high motif similarity but have a specific *N*- and *C*-terminal domain. These regions modulate water transport activities differentially in oocyte experiments [36,37]. Motif 4 is specifically detected in the *Oeu*PIPs subfamily, whereas motif 5 is shared between *Oeu*PIPs and *Oeu*SIP1s. Motif 3 was found in all the *Oeu*AQPs, except for *Oeu*XIPs and *Oeu*SIP2s, and motif 6 is specific to *Oeu*XIPs. In line with previous motif analysis in other plant models, the XIP and SIP subfamilies exhibit relatively few motifs. Based on analysis of conserved protein motifs between different Oleaceae and the tomato AQP, gain and loss of certain motifs can be observed within orthologous sequences/subgroups (data not shown). This highlights plausible functional divergences of some Asterid orthologues.

Some of these motifs contain specific amino acid residues (or signatures) that are crucial for the transportation function of the aquaporins (Figure 3; Figure S5), i,e., the dual Asn-Pro-Ala (NPA) (NPA1 in motif 1 and NPA2 in motif 3), the aromatic/arginine (ar/R) selectivity filter (H2 in motif 1, H5 in motif 6 and LE1 and LE2 in motif 10) and the Froger's residues (P1 in motif 1 and P2-P3-P4-P5 in motif 7). To understand the possible physiological role and substrate specificity of olive tree aquaporins, all these residue positions were identified and analyzed.

The association of the two highly conserved "NPA" motifs initially placed in the loops LB and LE forms the main central constrict of the channel and creates an electrostatic repulsion of protons. Different point mutations of amino acids at these positions have been found to strongly impact the substrate specificities of aquaporins [38]. Every *Oeu*AQP contains these motifs in loops LB and LE, but some amino acid triplets show residue variations (Figure 3 and Figure S5). *Oeu*PIPs and *Oeu*TIP contain two NPA residue triplets. The *Oeu*XIPs members contain a conserved second NPA motif, but all of the first NPA motifs show a replacement of the alanine (A) by a valine (V). In the case of the *Oeu*SIP subfamily, members from each subgroup have dual NPA motifs, but the third NPA1 residue is replaced by a threonine (T) (*Oeu*SIP1s) or a leucine (L) (*Oeu*SIP1s). As for the *Oeu*NIP subfamily, the first NPA motif shows an alanine (A) to serine (S) substitution in three *Oeu*NIPs (*Oeu*NIP5;1, *Oeu*NIP5;2 and *Oeu*NIP6;1), and the second NPA motif shows an alanine (A) to valine (V) substitution in four *Oeu*NIPs (*Oeu*NIP1;3, *Oeu*NIP5;1, *Oeu*NIP5;2 and *Oeu*NIP6;1). Each of these substitutions partly determines the function of transporting water [39].

The aromatic/Arg (ar/R) selectivity filter is essential for selective transport of substrate molecules [2]. Compared to the two NPA motifs, the ar/R positions are characterized by high subfamily- and/or subgroup-specific residues (Figure 3 and Figure S5). The *Oeu*PIP members all contain the more hydrophilic ar/R selectivity filter (F-H-T-R), as observed in the pure water channel *AqpZ*. This motif is a hallmark of water and solute transporting aquaporins compared to other subfamilies [5]. By contrast, these residues are highly variable in the other subfamilies, but their diversity matches each subgroup identified. In detail, among the *Oeu*TIPs, *Oeu*TIP1 members harbor residues (H/I/A/V), forming a

more hydrophobic ar/R filter compared to the *Oeu*TIP2 or *Oeu*TIP3 and *Oeu*TIP4 subgroups, for which members contain an ar/R filter with (H/I/G/R) or (H/V/A/R) residues, respectively. Lastly, the three *Oeu*TIP5 members harbor the most hydrophobic (N/V/G/Y) residues. The residues present in the ar/R selectivity filter in the different *Oeu*TIP subgroups are similar to TIPs from other plant species. Among the *Oeu*NIPs, the *Oeu*NIP1, -3 and -4 subgroups which harbor the (W/V/A/R) residues would be more hydrophobic compared to the *Oeu*NIP5, -6 and -7 subgroups with (A/I/G/R), (T/I/A/R) and (A/V/G/R) residues, respectively. This suggests that each subgroup of NIPs has its own function. For example, it was reported that NIPs exhibiting a (H/V/A/R) signature present a water permeability decrease, while enhancing transport of uncharged solutes such as glycerol and formamide [40]. In addition, the NIPs that combine the ar/R signature (A/I/G/R) with the dual NPS/NPV motifs transport arsenite, boric acid and silicon in rice [41,42]. *Oeu*NIP5 members possess this combination, suggesting that the olive *Oeu*NIP5s have the ability to absorb these particular substrates. These OeuNIP5s may substitute the NIP2 subgroup that is absent in Olea genomes. *Oeu*XIPs contain a quite hydrophobic ar/R selectivity filter composed of (I-A/L/A/R). This feature is accentuated with the isoleucine and leucine at the H2 and H5 positions. The ar/R selectivity filter in XIPs from different plants is relatively hydrophobic in nature, facilitating the transport of bulky and hydrophobic molecules such as glycerol, urea and boric acid in plants [30,43–45].

Froger's positions concern five residues that have an important role in the selection of molecules across cell membranes, helping to discriminate between aquaglyceroporines (AQGP) and aquaporins (AQP *stricto sensus*) [46]. These Froger's residues concern an aromatic residue at P1, an acidic residue at P2, a basic residue at P3 and two non-aromatic residues at P4 and P5. A relative consensus can be observed between the *Oeu*AQP subfamilies, but some singularities emerge (Figure 3 and Figure S5). The *Oeu*PIP members exhibit identical amino acids in the four P2-P5 residues (S-A-F-W), but the first residue is highly variable with three amino acids, glycine (G), methionine (M) or glutamine (Q), which are alternatively encountered at this position. As for *Oeu*TIPs, the five residues (T/S/A/Y/W) are highly conserved, with the exception of *Oeu*TIP3;1 and *Oeu*TIP5;3, for which P2 is an alanine (A). The Froger's positions (as in the ar/R selectivity filter) are quite divergent in the *Oeu*TIP subfamily members compared with those in other subfamilies, indicating that *Oeu*TIPs may have different solute permeabilities. The three *Oeu*XIPs harbor identical residue positions, as well as members belonging to each *Oeu*SIP subgroup. Concerning the *Oeu*NIP members, a great variability appears between members but with subgroup-dependent patterns, e.g., *Oeu*NIP1s with (F-S-A-Y-I), *Oeu*NIP3 with (F-S-A-Y-V), *Oeu*NIP4 with (F-S-A-Y-I), *Oeu*NIP5s with (F-T-A-Y-L) and *Oeu*NIP6 with (Y-S-A-F-L).

The different steric hindrances of the side chains involved in the variable NPA motifs, ar/R selectivity filter and Froger's observed between *Oeu*AQPs form differential constriction zones. Their comprehensive comparison would indicate that *Oeu*PIPs, *Oeu*TIPs, *Oeu*XIPs, *Oeu*NIPs and *Oeu*SIPs facilitate the transport of distinct substrates across membranes. Interestingly, they coordinately play a crucial role in facilitating the transport of various permeants with some other amino acid residues, for which their own side chain is involved within the central pore [6,18,21]. To this respect, the prediction of *Oeu*AQP functions, based on key protein domains conservation (SPD) as proposed by Hove and Bhave [6], was recorded (Figure 4; Figure S8). *Oeu*PIPs have H_2O_2-type (with the exception of *Oeu*PIP2;11) and urea-type SDPs. In addition, all *Oeu*PIP1s have boric acid-type SDPs, and only two of them, *Oeu*PIP1;4 and *Oeu*PIP1;6 display CO_2-type SDPs. These results show the conservation of plant PIPs in the transport of urea and hydrogen peroxide [47,48] and that PIP1s—but not PIP2s—could be main boric acid channels and CO_2 [49]. Concerning CO_2, AQPs were shown to facilitate the diffusion of gases, and as such, AQP activity may affect CO_2 transport and carbon metabolism as a whole [50,51]. Compared to *Oeu*PIPs, the *Oeu*TIP subfamily that features specific and conserved selectivity filters between members all have urea-type SDPs, and four members (*Oeu*TIP1;7, *Oeu*TIP2;1, *Oeu*TIP2;2 and *Oeu*TIP5;3) harbor H_2O_2-type SDPs. The fact that the *Oeu*TIP subgroups harbor specific ar/R filters and SDP signatures suggests that *Oeu*TIPs play a crucial role in transporting a wide range of molecules, as experimentally observed in *Citrus sinensis* [6].

SD position / Aquaporin	SDP1	SDP2	SDP3	SDP4	SDP5	SDP6	SDP7	SDP8	SDP9
Typical boric acid transporter	**T/V**	**I/V**	**H/I**	**P**	**E**	**I/L**	**I/L/T**	**A/T**	**A/G/K/P**
***Oeu*PIP1;1a**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;1b**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;1c**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;2**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;3**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;4**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;5**	T	I	H	P	E	L	L	T	P
***Oeu*PIP1;6**	T	I	H	P	E	L	L	T	P
***Oeu*NIP5;1**	T	I	H	P	E	L	L	A	P
***Oeu*NIP5;2**	T	I	H	P	E	L	L	A	P
***Oeu*NIP6;1**	T	I	H	P	E	L	L	A	**E**
***Oeu*XIP1;1**	V	V	H	P	E	I	T	T	K
***Oeu*XIP1;2**	V	V	H	P	E	I	T	T	K
***Oeu*XIP1;3**	V	I	H	P	E	I	T	T	K
Typical CO_2 transporter	**I/L/V**	**I**	**C**	**A**	**I/V**	**D**	**W**	**D**	**W**
***Oeu*PIP1;4**	I	I	C	A	I	D	W	D	W
***Oeu*PIP1;6**	I	I	C	A	I	D	W	D	W
Typical H_2O_2 transporter	**A/S**	**A/G**	**L/V**	**A/F/L/T/V**	**I/L/V**	**H/I/L/Q**	**F/Y**	**A/V**	**P**
***Oeu*PIP1;1a**	A	G	V	F	I	H	F	V	P
***Oeu*PIP1;1b**	A	G	V	F	I	H	F	V	P
***Oeu*PIP1;1c**	A	G	V	F	I	H	F	V	P
***Oeu*PIP1;2**	A	G	V	F	I	H	F	V	P
***Oeu*PIP1;3**	A	G	V	F	I	Q	F	V	P
***Oeu*PIP1;4**	A	G	V	F	I	H	F	V	P
***Oeu*PIP1;5**	A	G	V	F	I	H	F	V	P
***Oeu*PIP1;6**	A	G	V	F	I	H	F	V	P
***Oeu*PIP2;1**	A	G	V	F	I	H	F	V	P
***Oeu*PIP2;2**	A	G	V	F	I	H	F	V	P
***Oeu*PIP2;3**	A	G	V	F	I	Q	F	V	P
***Oeu*PIP2;4**	A	G	V	L	I	Q	F	V	P
***Oeu*PIP2;5**	A	G	V	F	I	Q	F	V	P
***Oeu*PIP2;6**	A	G	V	F	I	H	F	V	P
***Oeu*PIP2;7**	A	G	V	F	I	Q	F	A	P
***Oeu*PIP2;8**	A	G	V	F	I	H	F	V	P
***Oeu*PIP2;9**	A	G	V	F	I	H	F	V	P
***Oeu*PIP2;10**	A	G	V	V	I	Q	F	V	P
***Oeu*TIP1;7**	S	A	L	A	I	H	Y	A	P
***Oeu*TIP2;1**	S	A	L	V	I	I	Y	V	P
***Oeu*TIP2;2**	S	A	L	V	I	I	Y	V	P
***Oeu*TIP5;3**	S	A	L	T	I	Q	Y	V	P
***Oeu*NIP1;1**	A	A	L	L	I	I	Y	V	P
***Oeu*NIP1;3**	A	A	L	L	I	I	Y	V	P
OeuXIP1;1	A	G	L	V	L	H	F	A	P
***Oeu*XIP1;3**	A	G	L	V	L	H	F	V	P
Typical urea transporter	**H**	**P**	**F/I/L/T**	**A/C/F/L**	**F/L/M**	**A/G/P/T**	**G/S**	**G/S**	**N**
***Oeu*PIP1;1a**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;1b**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;1c**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;2**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;3**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;4**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;5**	H	P	F	F	L	P	G	G	N
***Oeu*PIP1;6**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;1**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;2**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;3**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;4**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;5**	H	P	L	F	L	P	G	G	N
***Oeu*PIP2;6**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;7**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;8**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;9**	H	P	F	F	L	P	G	G	N
***Oeu*PIP2;10**	H	P	F	F	L	P	G	S	N
***Oeu*PIP2;11**	H	P	F	F	L	P	G	G	N
***Oeu*TIP1;1**	H	P	L	F	L	A	G	S	N
***Oeu*TIP1;2**	H	P	F	F	L	A	G	S	N
***Oeu*TIP1;3**	H	P	F	F	L	A	G	S	N
***Oeu*TIP1;4**	H	P	F	F	L	A	G	S	N
***Oeu*TIP1;5**	H	P	F	F	L	A	G	S	N
***Oeu*TIP1;6**	H	P	F	F	L	P	G	S	N
***Oeu*TIP1;7**	H	P	F	F	L	A	G	S	N
***Oeu*TIP2;1**	H	P	F	A	L	P	G	S	N
***Oeu*TIP2;2**	H	P	F	A	L	P	G	S	N
***Oeu*TIP2;3**	H	P	F	A	L	P	G	S	N
***Oeu*TIP2;4**	H	P	F	A	L	P	G	S	N
***Oeu*TIP2;5**	H	P	F	A	L	P	G	S	N
***Oeu*TIP3;1**	H	P	F	L	L	P	G	S	N
***Oeu*TIP4;1**	H	P	L	L	L	P	G	S	N
***Oeu*TIP5;1**	H	P	F	A	L	P	G	S	N
***Oeu*TIP5;2**	H	P	F	A	L	P	G	S	N
***Oeu*TIP5;3**	H	P	F	V	L	P	G	S	N
***Oeu*NIP1;1**	H	P	I	A	L	P	G	S	N
***Oeu*NIP1;2**	H	P	I	A	L	P	G	S	N
***Oeu*NIP1;3**	H	P	I	A	L	P	G	S	N
***Oeu*NIP3;1**	H	P	I	A	F	P	G	S	N
***Oeu*NIP4;1**	H	P	I	A	L	P	G	S	N
***Oeu*NIP5;1**	H	P	I	A	L	P	G	S	N
***Oeu*NIP5;2**	H	P	I	A	L	P	G	S	N
***Oeu*NIP7;1**	H	P	I	A	L	P	G	S	N
***Oeu*XIP1;2**	H	P	F	A	F	T	G	G	N
***Oeu*XIP1;3**	H	P	F	A	L	T	G	G	N
Typical Silicic acid transporter	**C/S**	**F/Y**	**A/E/L**	**H/R/Y**	**G**	**K/N/T**	**R**	**E/S/T**	**A/K/P/T**
Not found	-	-	-	-	-	-	-	-	-
Typical NH_3 transporter	**F/T**	**K/L/N/V**	**F/T**	**V/L/T**	**A**	**D:S**	**A/H/L**	**E/P/S**	**A/R/T**
Not found	-	-	-	-	-	-	-	-	-

Figure 4. Summary of typical specificity-determining positions (SDPs) and those identified in the complete sequences of *Oeu*AQPs from the wild olive species *Olea europaea* var. *sylvestris*. The red residue represents a putative novel site. Amino acid alignments of the *Oeu*AQPs transporting non-aqua substrates where SDP signatures were identified are shown in Figure S8.

The NIP subfamily has low intrinsic water permeability and the ability to transport solutes such as glycerol and ammonia [52]. All *Oeu*NIPs have urea-type SDPs, except for OeuNIP6s. In addition, *Oeu*NIP1;1 has the H_2O_2-type SDPs. Similarly, *Oeu*NIP5s and *Oeu*NIP6;1 members (with the presence of a glutamic acid (E) in SDP9 for which its regulatory function needs to be experimentally validated) have boric acid-type SDPs. This last transport prediction is in accordance with the NPS/NPV aqueous pore and (A-I-G-R) ar/R selectivity filters observed in these NIP members, which are known as boric acid transporters in the orthologous *At*NIP5;1 [42]. This suggests that NIPs could be involved in the transport of larger solutes. SDP analysis suggests that the *Oeu*XIP subgroup displays contrasted solute permeability between members: *Oeu*XIP1;1 is a transporter of boric acid and H_2O_2, *Oeu*XIP1;2 is a transporter of boric acid and urea and *Oeu*XIP1;3 is a transporter of boric acid, H_2O_2 and urea. These SDP signatures corroborate 3D modeling and functional analysis [43,44]. Notwithstanding that their functional role remains quite unclear in plants, it is possible that *Oeu*XIPs have different roles in olive trees.

It should be noted that we were unable to predict SDP signatures for the four *Oeu*SIP members. Similarly, no *Oeu*AQP held silicic acid-type and NH_3-type SDPs. We have already reported that the NIP2 subgroup is absent in the olive genome. However, it has been reported that the NIP3 subgroup is also a transporter of Si (and Ge, As and B) and that their expression can be a critical factor in determining the ability of a plant to absorb Si [24–53]. *Oeu*NIP3;1 could assume this role of Si transporter. Incidentally, it suggests that other types of SDPs could be involved in Si absorption (in particular for *Oeu*NIP3;1) during plant growth, development or stress responses in olive trees. As for the NH_3 element, no complete NH_3-type SDPs were identified in the two concerned analogous *Oeu*TIP and *Oleur*TIP subgroups, whereas they are present in their counterpart *Fraex*TIP2;1 (T/L/T/V/S/H/P/A), as confirmed with its orthologous *At*TIP2;1 [54] (data not shown). This functional divergence shows the nucleotide substitutions that played a crucial role in the diversification of the conserved residues that determine the substrate specificity. It supports the assertion that these evolutionary impetuses are particularly universal in the AQP superfamily, reflecting a wide variety of their complementary substrate transport capacity.

Aquaporin activity is regulated transcriptionally and/or post-translationally by pH, phosphorylation and vesicle trafficking. In this context, a panel of motifs and post-translational regulation signatures were identified in *Oeu*PIPs. It is admitted that PIP1s have null or lower water permeability than PIP2s and that a functional difference would be attributed to one residue just after H2 in the THM2: all PIP1s possess alanine (A) while the PIP2s have valine (V) or isoleucine (I) [55]. At this position, *Oeu*PIP1s and *Oeu*PIP2s feature this singular difference (Figure S5A), suggesting that the *Oeu*PIP2 subgroup is more permeable to water than the *Oeu*PIP1 subgroup. This hypothesis is supported by the observation of Secchi et al. [16] that *Oe*PIP2;1 (the *Oeu*PIP2;1 analogue) is highly permeable to water while *Oe*PIP1 has no water transport activity in oocytes. Several *Oeu*TIPs share this alanine residue with *Oeu*PIP1s, though they are considered (at least in theory) as being water permeable. However, previous data showed that *Oe*TIP1.1 (the *Oeu*TIP1;1 analogue) exhibits water transport activity in oocytes, though lower than *Oe*PIP2 [16]. It is evident that water permeability is defined by more than one residue. In this respect, some *Oeu*TIPs such as *Oeu*TIP1;3 and *Oeu*TIP2;1-2-3 harbor an alanine substituted with a glycine (G) (Figure S5B). Interestingly, these members show the highest levels of transcriptional expression (*cf* RNAseq expression item). With the exception of *Oeu*NIP7;1, *Oeu*NIP members also exhibit this glycine position (Figure S5D). The plausible contribution of this small glycine to the water permeability of *Oeu*TIP and *Oeu*NIP members has to be resolved.

Similarly, AQPs are known to function as tetramers. For this to happen, the highly conserved cysteine (C) (Cys80 in *Zm*PIP2;1) in loop LA of PIP1s and PIP2s is shown to be essential for formation of disulphide bonds between PIP monomers, increasing the oligomer stability [56]. Every *Oeu*PIP1 and *Oeu*PIP2 contains this cysteine (Figure S5A), suggesting possible oligomerization between members. However, this specific residue is absent in TIPs, XIPs, NIPs and SIPs, suggesting that other residue(s) could be involved in the structural relationship of monomers leading to AQP quaternary models.

Furthermore, we note the presence of various phosphorylation sites in all *Oeu*PIP1 and *Oeu*PIP2 members with, for example, the serine (S) residue in the loop LB that is located in domains which correspond to the recognition sequences of protein kinase A (R-K-x-S) and protein kinase C (R-K-x-S-x-x-K) in all *Oeu*PIP1s and *Oeu*PIP2s, with the exception of *Oeu*PIP2;10, which contains a calmodulin-like domain protein kinase (L-x-R-x-x-S) (Figure S5A). In addition, the *C*-termini of some OeuPIP2s have a serine located in the potential cAMP-dependent protein kinase site (S-x-K). Lastly, different sites can be observed, such as the *N*-t AEFxxT-box at the -60 amino acid position to the first NAP with unclear function, the histidine (H) residue in loop LD involved in gating, the *N*-terminal diacidic motif KD(V/I)E for putative conserved methylation sites and found to be important for the exit of newly synthesized PIPs from the ER to plasma membrane and the putative conserved blocking leucine residue (L) in TMH5. Regarding the NIP subfamily, different phosphorylation sites are also found in the *C*-termini of *Oeu*NIP1;1, *Oeu*NIP1;3 and *Oeu*NIP4;1 (Figure S5D). It would then be very interesting to specifically mutate some of these residues in order to validate the importance of these regulation points in the permeability control of these AQPs.

In Picual cultivar, 77% of *Oleur*AQP-types have >99% similarity with their *Oeu*AQP paralogues. For the other *Oleur*AQP-types, and despite several residue variations, all *Oleur*AQP-types adhere to the *Oeu*AQP structures and features. Most *Oleur*AQPs contain residue signatures (ar/R filter, Froger's position and the NPA and different substrate-type SDP motives) identical to those of their *Oeu*AQP paralogues. Only *Oleur*PIP2;2, *Oleur*PIP2;6, *Oleur*TIP5s, *Oleur*NIP6;1 and *Oleur*SIP1;1c diverge in some of the residues constituting the ar/R filter, Froger's positions and NPA motives. In addition, *Oleur*PIP2;2ab and *Oleur*NIP7;1ab display residue substitutions in their urea-type SDPs, annihilating the urea transport capacity of these members, a priori.

Although some variations in amino acids are seemingly fixed between paralogues during the evolution of the Oleaceae, strong sequence similarities and protein functionalities seem to be conserved between *Oleur*AQPs and *Oeu*AQPs, suggesting that each paralogue contributes to specific and common key biological functions in a domesticated cultivar. This hypothesis is strengthened by the purifying selections that predominated across the paralogue gene pairs between wild and domesticated olive trees. This evolution pressure of domestication might have resulted in the conservation of protein sequences and related biological functionalities.

In conclusion, the overall protein similarity between AQPs from the Oleaceae species studied here and a large panel of plant species, notably at the residues constituting the different key motifs designing the permease selectivity and ultimately the biological functions, suggests that the knowledge on channel selectivity which was revealed in various plant models might be easily transferable to Oleaceae crop isoforms. This would indicate that the diversity of AQPs in *O. europaea* L. have a crucial role in the cell responses to environmental stresses.

2.5. Localization of the OeuAQP Proteins

Investigations have been increasingly deployed to understand tissue and subcellular specificity of genome-wide AQPs in plants, especially as the subfamily names do not systematically indicate location [reviewed in 37,57]. To this respect, sub-cellular localizations of *Oeu*AQPs were predicted to ascertain expression at different cellular and organellar levels (Figure 3).

Plant AQPs exhibit a broader subcellular localization, where PIPs, NIPs and XIPs are usually found in the plasma membrane or chloroplast membranes, while TIPs and SIPs are generally localized to the tonoplast and endoplasmic reticulum and Golgi apparatus, respectively [2,43]. The subcellular localization prediction of *Oeu*AQPs showed that, as observed in other plants and suggested by their nomenclatures, basic *Oeu*PIPs and acidic *Oeu*TIPs are predicted to preferentially localize to plasma membranes and vacuolar membranes, respectively.

In addition, predictions situate some *Oeu*PIPs at the vacuolar membrane and/or endoplasmic reticulum, and *Oeu*TIPs to plasma membranes. All basic *Oeu*NIPs were predicted to preferentially target the plasma membrane, whereas *Oeu*TIP1;6–7 and *Oeu*NIP4;1 appeared localized in the

tonoplast. Their NIP homologs in several organisms were determined to localize to plasma membrane, endoplasmic reticulum or the peribacteroid membrane of root nodules [42–58].

Compared with the diverse predicted localizations using WoLF PSORT or Plant-mPLoc, all *Oeu*XIPs were predicted to localize predominantly to the plasma membrane, which is consistent with other experimental results [43,44].

Concerning the last *Oeu*SIP subfamily we studied, a wide range of subcellular localizations are reported, including plasma membrane, tonoplast, endoplasmic reticulum, Golgi apparatus, nuclear membrane and chloroplasts. It has also been reported that SIP1 proteins could transport water across endoplasmic reticulum membrane while SIP2 proteins may act as an endoplasmic reticulum channel for small molecules or ions [59–61].

Overall, many experimental observations report that AQP subcellular localizations are multiple, complex and finely regulated. For example, *Nt*AQP1, a member of the PIP group from tobacco, was found in the inner chloroplast membrane and the plasma membrane [51]. Dual localizations have also been observed for *Zm*PIP1;2 from maize in the plasma membrane and endoplasmic reticulum [36] and the seed specific *At*TIP3;1 and *At*TIP3;2 isoforms from *Arabidopsis*, which were observed in tonoplast and plasma membrane [62]. It is now clearly established that the subcellular localization, relocalization and redistribution of specific AQPs are influenced by fluctuating environmental conditions [63,64]. However, the molecular mechanisms underlying the multiple localizations of these genes are not yet clear.

Our in silico analyses are highly informative but not conclusive. Although our predicted subcellular localizations of most *Oeu*AQPs are consistent with the general features of plant AQPs, further investigations are required to better document the subcellular localizations of *Oeu*AQPs.

2.6. Chromosomal Distribution and Gene Structure of OeuAQPs

The Olea genome is constituted of 23 chromosomes. The physical mapping of the 68 identified *OeuAQP*-encoding genes (full-length and pseudogenes) indicates that they are distributed on nineteen chromosomes (Figure 5). With the exception of chromosome 4, which contains the largest number of *OeuAQP* genes with seven members, other chromosomes harbor one to three *OeuAQP* members. No *OeuAQPs* gene was found on chromosomes 7, 14, 15 and 23. To date, 22 *OeuAQPs* are positioned on 22 scaffolds not yet assigned to a chromosome, and that concerns particularly the *Oeu*PIP2 subfamily with seven members out of the eleven *OeuPIP2s* identified. *OeuAQP* gene density varies on individual chromosomes. However, globally, *OeuAQP* genes were dispersed in a single manner alongside the chromosomes. However, we note five clusters on the chromosomes 2, 4, 12, 19 and 20, which correspond to tandem duplication of *OeuTIP2;1*/*OeuTIP2;2* (chr2), *OeuTIP1;4-OeuTIP1;5* (chr4); *OeuTIP5;1-OeuTIP5;2* (chr12), *OeuPIP1;1a-b-c* (chr19) and *OeuTIP1;2-OeuTIP1-related* (chr20). In addition, we found two *OeuNIP3s* (*OeuNIP3;1-OeuNIP3-related*) tandemly duplicated in the unanchored scaffold 635. All these paralogues are characterized by the same open reading frame orientation on the same linkage group, and duplicates are relatively close to each other (*Oeu*PIP1a-b-c: 2897/7516nt, resp; *Oeu*TIP1;4-5: 1702nt; *Oeu*TIP2;1-2, 1159nt; and *Oeu*TIP5;1-2, 1022nt).

Compared to the 68 cv. Picual *Oleur*AQP duplicates, the number of *OeuAQP* gene pairs is relatively small. Nonetheless, gene duplication is considered as a major driving force for the evolution and the expansion of gene families, which may also contribute to stress tolerance via gene dosage increasing, avoiding some deleterious mutations and creating the opportunity for new function emergence [65]. In this regard, several duplication events have been recorded during the evolution of several plant species leading in particular to the expansion of the MIP superfamily [66,67]. The tandemly duplicated genes mostly concern the *Oeu*TIP subfamily, and particularly the *Oeu*TIP1 subgroup. That suggests that tandem duplications might have played a vital role in the expansion of the AQP gene family in olive trees, affecting in particular the TIP subfamily.

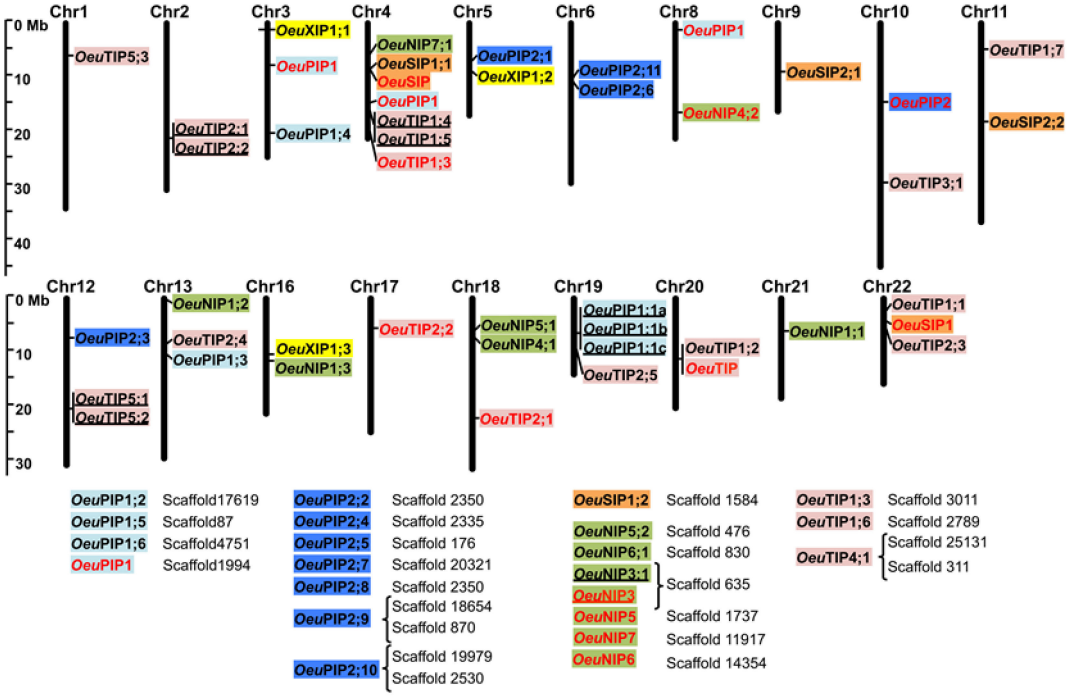


Figure 5. Genomic distribution of *Oeu*AQP genes on the 23 *Olea europaea* var. *sylvestris* chromosomes (except chromosomes 7, 14, 15 and 23). The different aquaporin subclasses are distinguished using different colored boxes that correspond to the color code in Table S2. The scale is in megabase pairs (Mb). Sixteen sequences could not be anchored to a specific chromosome and are specified in their scaffold of origin. Pseudogenes are specified in red. Tandemly duplicated genes are underlined.

Furthermore, the duplicated gene pairs might have undergone three alternative fates during their evolution, i.e., nonfunctionalization, neofunctionalization and subfunctionalization [68]. Our results show that each full-length *Oeu*AQP duplicate has a *Ka*/*Ks* ratio between 0,001 and 0,2201, suggesting that *Oeu*AQP genes in the wild variety might have experienced strong purifying selection pressure with limited functional divergence after duplication. Purifying selection supports the process of elimination of deleterious loss-of-function mutations, enhances fixation and preserves the function of a new duplicated gene [69]. Our results suggest that the duplicated *OeuTIP* genes in the wild olive species did not diverge significantly during subsequent evolution, where purifying selection might have contributed greatly to the maintenance of function in the TIP subfamily genes. However, the exact functional contribution for each paralogue needs to be established.

By contrast, it should be noted that the three *OeuXIPs* are dispersed on the chromosomes 3, 5 and 16. Although the chromosome-scale assembly is not yet available for the olive cultivar Picual or for *F. excelsior*, the unanchored scaffolds of these species are wide enough on both sides of the *XIP* sequences to confirm the absence of tandem duplicates (data not shown). This distribution contrasts radically with what is generally observed in plant genomes which contain this singular subfamily and where the *XIP* paralogues are systematically found next to each other on chromosomes (including tomato [19]). Would this mean that XIP genes differentially evolved in Lamiales? Only access to additional related genomes will answer this intriguing question.

The exon-intron structure organization is an additional lever to provide clues for understanding the evolutionary connection of a gene family. In the three Oleaceae families, the number of AQP exons ranged from two to five. Although the length of each exon is similar for most members in each subfamily, some deviations are noted. In *O. europaea* var. *sylvestris*, with a few exceptions, the number and the size of the exons are conserved within each AQP subfamily (Figure 2B). This finding further validates the nomenclature we proposed with the phylogenetic analysis (Figures 1 and 2A). However, these features differ notably between AQP subfamilies. Most *OeuPIP* members are characterized by

four exons, the exceptions being *OeuPIP2;3* and *OeuPIP2;9*, which feature only three exons. The genes assigned to the *Oeu*TIP subfamily are mostly composed of three exons. Only *OeuTIP1;3* contains two exons. As for *Oeu*TIP4;1 and its analogous *Oleur*TIP4;1, the second intron is remarkably long, with 6483nt and 6041nt, respectively. This particularity is not shared by *Fraex*TIP4;1 (315n) and seems specific to *Olea* genomes. In addition, the majority of the *OeuNIP* members contain five exons, except for the two *OeuNIP5s*, which harbor four exons due to the lack of the second intron. Another observation can be made with this NIP5 subgroup: while in *Oeu*NIP5;2 the exon and intron lengths are comparable to different orthologous NIP5, *Oeu*NIP5;1 isoforms possess longer introns, giving a *OeuNIP5;1* gene size of more than 14032 bp (the same is observed with *OleurNIP5;1a*, 7392nt; *FraexNIP5;1*, 5999nt). Despite these intron elongations, the sequence identity of *Oeu*NIP5;1 exons is similar to other NIP subgroups, resulting in highly similar NIP5 protein sequences. As observed with the TIP4 members, the molecular and evolutionary reasons for intron elongation tendencies have yet to be resolved. The subfamily of *Oeu*SIPs are characterized by a conserved three-exon structure. However, this apparent conservation is relatively unexpected, particularly with respect to the very low identity percentage between the two *Oeu*SIP1 subgroups. As for the last XIP subfamily, the *OeuXIP* members display three exons (*OeuXIP1;1*) or four exons (*OeuXIP1; -2* and *-3*). The exon structure of the XIP subfamily is relatively variable between plant species.

By contrast, the introns vary greatly in both length and position. This is particularly marked for the *Oeu*XIP, *Oeu*NIP and *Oeu*SIP subfamilies, and with the *Oeu*PIP exception of *OeuPIP2;11*, which is split into two parts by the integration of a putative gene. That results in a gene length within *Oeu*AQP that varies from 0.86Kb (*Oeu*TIP1;3) to 14.032Kb (*Oeu*NIP5;1), corresponding to the exon-intron patterns observed in other crop or fundamental plants belonging to distinctly different taxa such as *Arabidopsis* [35], soybean [70], orange [29], poplar [20], chickpea [22] or watermelon [71].

2.7. Gene Structure of OleurAQPs

The 'Picual' cultivar genome encompasses 79 full-length sequences, including 26 single sequences and 53 duplicated gene sequences (Table S2). Concerning the duplicates, only four sequences show 100% similarity with their paralogue (Figure S7). However, although other paralogues show several variations in nucleotides, all protein duplicates share identical residue signatures and motifs to their paralogues. This is consistent with a nearly present duplication of this cultivar. Thus, it has taken a very short time to diverge, suggesting that they retained the same biological function in the course of evolution. Only two paralogues form an exception, *Oleur*PIP1;3b and *Oleur*NIP7;1b, which display substituted residues at the P5 and P1 Forger's positions, respectively.

Similarly, the intron-exon organizations are highly conserved, although there are some differences in the intron size between the wild and domesticated species or between several *Oleur*AQP duplicates. In addition, some particular exon divisions for *Oleur*PIP1;3a, *Oleur*PIP2;9, *Oleur*TIP1;7, *Oleur*XIP1;1b and *Oleur*NIP6;1b are observable, without impacting the final protein structures. The exon-intron structural divergences commonly occurred in the evolution of duplicate and paralogous gene sequences, and notably, the number and the size of the introns are correlated with the gene expression, gene duplication and gene diversification in plants [28]. These intragenic reorganizations occur through possible mechanisms of gain/loss, exonization/pseudo-exonization and insertion/deletion events [72]. Nevertheless, this intron diversity in the Oleaceae clade suggests original avenues of investigation.

Overall, gene duplication is a major mechanism for acquiring new genes and creating genetic novelty in an organism, enabling plants to rapidly adapt to new environments [73]. As with *Oleur*NIP1;4, each *Oleur*AQP duplicate is to be considered as a new sequence. However, the analysis of phylogenetic relationships between functional domains, exon/intron structure and gene duplications showed the relatively high conserved evolution of olive AQPs, and in this respect, the average *Ka*/*Ks* value between *Oleur*AQP duplicate pairs (i.e., 0.1637) is very low (except for *Oleur*NIP7;1b, which is under a diversifying selection with a ratio of 1.1560) or even smaller than the evolutionary dynamics between

*Oeu*AQP and *Oleur*AQP homologues (i.e., 0.4183) (except for AQP2;10, TIP1;6; TIP1;7, XIP1;3 and NIP7;1b, which are under diversifying selection).

All these results suggest a relatively lower rate of evolution of the AQP-type (including the duplicate members) in both *Olea* species, which has resulted in a very recent duplication with very little time to evolve, preserving their potential biological functions, a priori.

2.8. Transcriptional Profiling of Olive OleurAQPs

Olea in general and in particular cv. 'Picual' adapt well to a wide variety of soils and arid environmental conditions. Timely up- or down-regulation of certain aquaporin isoforms in normal physiological processes and under stress conditions is thought to be important for acclimation and/or tolerance to stresses. For this purpose, *Oleur*AQP transcriptional levels were monitored using RNA-seq in various organs (fruits, flowers, leaves, roots and stems) and tissue (meristems). In addition, the *Oleur*AQP expression kinetics were recorded from the early stage of growing young seedlings to the emergence of the juvenile plant and under three stress conditions, wound and *V. dahliae* inoculation on roots and cold stress on leaves (Figure 6).

As previously discussed, 21 *Oleur*AQP-types are represented by two duplicates or more. But whatever the variability level between counterparts, RNaseq analyses reveal a transcriptional expression for each, except for *Oleur*TIP1;3c, *Oleur*TIP7;1b, *Oleur*XIP1;1b and *Oleur*NIP5;1b, which show null expression. Every *Oleur*AQP-type with its full-length duplicate sequence is presented here. However, we shall develop our arguments mainly on the expression data from the *Oleur*AQP-type, and this for two reasons: (i) even if the paralogues for a *Oleur*AQP-type show significant variabilities in their transcriptional expression, the kinetic of expression (i.e., up or down profiles) are quite similar; (ii) we are analyzing the variations in *Oleur*AQP expression only with respect to the potential functional roles the related proteins would play in different physiological processes in olive trees. Concerning the pseudogenes and incomplete sequence, all show null expression in any condition (Table S4). This confirms the pseudogenization of these sequences (if misassembled, they might show some expression patterns). Thus, they were taken out of the discussion.

The olive *AQP* gene family displays complex differential expression patterns according to organs or tissues. Among the six organs tested, fruits, flowers and roots showed the highest levels of AQP abundance, thereby indicating their potential physiological function in these tissues during intense developmental or mineral and/or water absorption processes. In this regard, thirty *OleurAQP*-type genes significantly express in at least one organ. However, some exclusive or highly confined expressions of some of the AQPs in specific organs can be observed. In addition, every *OleurAQP*-type gene exhibits differential modulation during plantlet development or biotic and abiotic stresses.

The two subfamilies *Oleur*PIP and *Oleur*TIP show the highest expression in all organs compared to other subfamilies. *OleurPIP1s* ubiquitously express but with strong variation between members, whereas *OleurPIP1;1* is the most expressed member in all organs (RPKM ≈ 403), except in roots where *OleurPIP1:3* shows the highest expression (RPKM ≈ 266). *OleurPIP1;6* shows no expression, except for the first month of plantlet development where it expresses slightly (RPKM = 31). Five *Oleur*PIP2s, *OleurPIP2;1-2-5-8* and *-12*, present the strongest transcript accumulations in all organs (RPKM ≈ 75), but with strong differential expression levels according to the organs. *OleurPIP;12* is the most common PIP2 member expressed and it accumulates in all bioassays (RPKM ≈ 155). In terms of genome evolution, the case of this isoform is interesting because the loss of *Oleur*PIP;12 in the wild variety suggests that its functions were shared with other AQPs. *OleurPIP2;6* is expressed in most organs (RPKM ≈ 70), except in roots where it expresses slightly (RPKM = 14), while *OleurPIP2;8* mainly expresses in flowers and roots. *OleurPIP2;1* seems to be mainly expressed in roots.



Figure 6. Expression profile of *Oleur*AQP genes from the cultivated *Olea europaea* cv. Picual in different physiological conditions. *Oleur*AQsP are grouped by subfamily. Fr, fruit; Fl, flower; Le, leaf; Me, meristem; Ro, root; St, stem. Control assays are marked C. h, hours post treatment; d, days post treatment. The horizontal color scales at the bottom of the figure represent the normalized expression of the aquaporins in terms of reads per kilobase of transcript per million mapped reads (RPKM). Expression levels are displayed by the depth of red, and the related values are detailed in Tables S4 and S5.

These contrasted transcript accumulations can reveal very interesting plausible organ-specific compensations between protein isoforms, for example *Oleur*PIP1;3 with *Oleur*PIP1;6 and *Oleur*PIP2;6 with *Oleur*PIP2;8. In addition, members of PIP1 and PIP2 subgroups can associate together in heterodimers and tetramers. This association enhances water transport, in particular for PIP1 members which originally had no water permeability [74,75]. PIP heteromerization events can be then viewed throughout the expressions of some *OleurPIP* counterparts that are well-correlated, i.e., *OleurPIP1;3* with *OleurPIP2;8* and *OleurPIP1;6* with *OleurPIP2;6*. The fact that these *Oleur*PIPs (as their paralogous *Oeu*PIPs; Figure S5A) display the conserved cysteine residue in loop A, which is essential for the formation of the disulphide bonds between PIP monomers, could hypothetically underpin, in part, the reason of such compensatory organ-dependent co-expressions observed for a multigenic family. Three-dimension simulating prediction coupled to P_f calculation and oocyte experiments would bring relevant biochemical information to these protein heteromerizations.

During early plantlet development stages or stress conditions on roots and leaves, *OleurPIP1;1* is significantly up-regulated in any biological conditions, while *OleurPIP1;2*, *OleurPIP1;4*, *OleurPIP2;1* and *OleurPIP2;12* are strongly repressed. *OleurPIP1;3* is up-regulated during plantlet development and wounding, up-regulated at 48hpi and then repressed during *V. dahliae* inoculation and by cold stress. *OleurPIP2;5* is up-regulated during plantlet development and *V. dahlia* inoculation but repressed by wound and cold stress. *OleurPIP2;8* is up-regulated by wounding and by *V. dahliae* and repressed at four months of plantlet development and cold. *OleurPIP2;6* seems to be only up-regulated by cold. These contrasted profiles demonstrate, once again, the complex and plural fine-tuned regulations of AQP in plants.

In the *Oleur*TIP subfamily, only nine *OleurTIP*-types are expressed in all organs but in very imbalanced levels. *OleurTIP1;2* and *OleurTIP1;3* are the highly expressed members (RPKM > 500), while *OleurTIP1;1* and *OleurTIP4;1* present infinitesimal expression (RPKM < 10). However, the confined expressions of some of the TIP subfamily in specific organs are more pronounced than for the PIP subfamily. *OleurTIP1;3* expresses mainly in fruits, roots and stems, while *OleurTIP2;1* and *OleurTIP2;3* are expressed in fruits, flowers and roots. *OleurTIP1;6* expresses in fruits, flowers and leaves, and *OleurTIP4;1* mainly expresses in fruits. During plantlet development or stress conditions on roots and leaves, *OleurTIP1;3* expression is maximal in the first month of the platelet emergence (RPKM = 1318), and then decreases progressively (RPKM = 137). Concerning the stress conditions for the most significant, *OleurTIP1;3a* expression is repressed drastically ($RPKM_{essai}$ = 162/114 vs. $RPKM_{control}$ = 64,5), and ultimately, it is up-regulated significantly (RPKM=751), while the expression of *OleurTIPde* duplicates are up-regulated ($RPKM_{essai}$ > 150 vs. $RPKM_{control}$ < 100), except in leaves subjected to cold stress where they are all up-regulated. *OleurTIP1;2* shows similar kinetics in wounded roots and roots inoculated with *V. dahliae*. As for the *OleurTIP1;2b* duplicate, it highly expresses in fruits and moderately in roots and leaves. During abiotic stress in roots, this duplicate is up-regulated, while it is repressed by cold. Previous studies have shown a substantial role for TIP3 in flowering and seed development, specifically in the desiccation process [21,33,76–78]. As for the three last *Oleur*AQP subfamilies, *OleurNIP*, *OleurXIP* and *OleurSIP*, they show relatively very low expression levels (RPKM < 50 in most biological situations). In the *Oleur*NIP subfamily, the NIP transcript levels are lower than other MIPs, which is consistent with those observed in several plant species. *Oleur*NIP7;1 is strongly expressed in flowers (RPKM > 378), and *Oleur*NIP5;1 and *Oleur*NIP5;3 are expressed in roots and meristems (RPKM = 40 and 35, resp.). Interestingly, *At*NIP5;1 expression and function are linked to boric acid uptake into the roots of Arabidopsis [42]. This suggests that *Oleur*NIP5 channels play a potential role in boron transport in organs other than the root of olive trees. Concretely, although the water transport capacity of NIPs is rather limited, several NIP members are reported to be crucial for plant growth and development under boron limitation or in response to various abiotic stresses. In this regard, their biological roles in olive trees are worthy of further confirmation. Lastly, *OleurXIP1;1* is mainly expressed in flowers, roots and leaves (RPKM ≈ 35), and *OleurSIP1;1* is ubiquitously expressed (RPKM ≈ 35), with a preferential expression in flowers (RPKM ≈ 100).

In our analysis, *Oleur*TIP3s genes, which are represented by three duplicates, showed null expression in all tissues. That includes flowers, but even more during the initial phases of seed germination, for which the TIP3 subfamily is showed to be specifically and/or strongly expressed [62,76–78]. These results suggested that TIP3's gene function is not strictly conserved in plants, and this assertion was also reported in *Jatropha curcas* L. [18], banana [79] and carnation [80]. It is reasonable to hypothesize that other AQP proteins may replace this TIP3 function in flowers in several plant species, including olive trees. *OleurTIP1;3* or *OleurTIP2;3*, which showed high expression in flowers, could be the alternative counterparts. These results underpin the scheme that the functions of the same subgroup are differentiated in different species and that the aquaporin family is functionally redundant.

2.9. Functional Consideration of OleurAQP

The ubiquitous and high expression of PIP and TIP in various organs and tissues is widely reported [18,81,82]. In this regard and whatever the stress, reduced expression of these members is generally expected to prevent loss of metabolic energy in situations of severe stress and/or to prevent loss of water from challenged organs or tissues to the hypertonic surrounding environment. However, a global analysis of these subfamilies, as for *Oleur*PIPs and *Oleur*TIPs, shows concomitant up- and down-regulations between members. These differential and spatio-temporal gene responses indicate that AQP might perform diverse functions in a variety of biological situations and that AQPs operate together as a network, probably due to the multigenic diversity of members and possible compensation mechanisms between close homologs [83,84]. For example, during *V. dahliae* root inoculation, most AQP expressions are significantly down-regulated, except for *OleurPIP1;3*, *OleurPIP2;5*, *OleurPIP2;8*, *OleurTIP1;3* and some *OleurTIP2s* which are up-regulated. Plant pathogen development depends on the host for nutrition and water. Therefore, water movement at the sites of infection is crucial for disease progress, and limiting its availability could be interpreted as a potential defense mechanism. AQPs have rarely been considered as key candidates in biotic stress studies, the pathogenicity-related genes and their host counterpart resistance genes and disease response genes being the attractive targets. However, one of the efficient cell defense responses is linked to an oxidative burst with the production of reactive oxygen species (ROS). Among ROS, H_2O_2 is a signaling molecule involved in various cell responses, including plant disease immunity, and its transport across membranes is provided by specific aquaporines (also called peroxyporines in such a context) [85,86]. Despite 'Picual' cultivar being quite susceptible to *V. dahliae*, challenged roots significantly overexpressed some *Oleur*APQs (*Oleur*PIP1;3 and *Oleur*PIP2;8 at 48hpi, or *Oleur*TIP2;1 and *Oleur*TIP2;2 at 15dpi) that display H_2O_2-type SDPs (Figures 4 and 5), and for which transcriptional expressions correlate well with proteins directly involved in ROS protection [87]. Likewise, RNA-seq analyses clearly show several overexpressed *Oleur*PIP and *Oleur*TIP during cold acclimation. This biological phenomenon is the result of several biochemical processes, such as the de novo synthesis and the cell exchanges of cryoprotectant molecules, maintenance of ion homeostasis and a stress cell response mainly mediated by the scavenging and the removing of ROS [86,88,89]. Owing to their ability to transport such permeant cryoprotectant molecules, various AQPs belonging to PIP and TIP subfamilies would function in concert to regulate membrane water and solutes permeability, a key step in tolerating cold acclimation and freeze-induced dehydration stress [90,91]. Further studies are needed to elucidate the direct role of *Oleur*AQPs in biotic and abiotic stress regulations, in particular with regard to ROS.

A similar global trend is observed during the early stages of the olive-seedling development where most *Oleur*AQP are up-regulated. Cell division and expansion require the continuous uptake of water to maintain turgor pressure. This process is controlled by a gradient in water potential, which itself is generated by the accumulation of solutes. In this regard, cell and tissue hydraulic properties during expansion need to be tightly regulated, and the significance of PIP and TIP aquaporins in tissue elongation implies a high hydraulic permeability of the plasma membrane and tonoplast. In addition, seed germination, embryonic development and the emergence of the juvenile plants are eminently

intricate physiological processes [92]. The functional roles that some *Oleur*PIPs and *Oleur*TIPs could perform in these successive plant life processes in olive trees need to be exhaustively resolved.

2.10. AQP in the Evolution Course of the Olive Tree Domestication

The domestication process allows plant species to artificially change their phenotypes, aiming to obtain traits human beings find desirable. Several genome rearrangement processes are observed, but duplicate genes derived from polyploidy and their ultimate potential divergence during the evolution of the "novel" plant species can be the key to crop domestication and the evolution of various phenotype traits (development, maturation or stress tolerance) [93,94]. Divergence of duplicated gene pairs is a pervasive evolutionary process after duplication events, and the recent duplication of the cultivated olive genome gave rise to a panel of very close AQP isoforms for which the functional full-length paralogues are under purifying selection (Figure S7). It is established that the normal process after a duplication of the genome is that some paralogs become silenced and finally become pseudogenes. In the cultivated olive, that would concern 41 duplicate *OleurAQP* gene sequences, presenting deletion of a portion of the duplicate sequences and/or pseudogenization through loss-of-function mutations (Table S2). As for the rest, that is 27 full-length duplicates, the last duplication is so recent that all duplicated genes coexpressed with their paralogues.

Nonetheless, some nuances in the transcriptional profiles can be highlighted. Several *Oleur*AQP duplicated gene pairs displayed quite similar expression patterns (*Oleur*PIP1;4; *Oleur*PIP1;6; *Oleur*PIP2;2; *Oleur*PIP2;8; *Oleur*TIP1;2; *Oleur*SIP1;1). This would indicate that their biological functions might not be diverged during plant development and stress responses, but determining whether these genes had similar functions in different tested tissues would require further analysis. In contrast, functional differentiations seem to address five duplicated gene pairs which show different expression patterns in a tissue-specific manner or during stress: *OleurPIP2;1ab*, *OleurPI2;6ab*, *OleurPI2;9ab*, *OleurTIP1;3ade* and *OleurTIP2;3ab*. For these paralogues, the putative recent duplication event might have been a way to control specific expression of particular genes according to developmental and environmental conditions. Most probably, the coding regions have gained new regulatory ways by insertion or deletion of tissue-specific enhancers or repressors, causing spatial and temporal variations in expression patterns of duplicated genes [95]. In addition to specific cis-acting elements that would modulate expression of some *Oleur*AQP genes, the different expression patterns between two homologous genes can also be explained by difference in intron lengths. Implication of introns in the regulation gene expressions in different biological contexts is still under-estimated, and yet, although the intron regions are spliced out during transcription, introns have increasingly been shown to serve important biological functions; notably, they can elevate gene expression without functioning as a binding site for transcription factors [96,97]. This phenomenon was termed 'intron-mediated enhancement'. Hence, it can be speculated that the intron-exon organization of some *Oleur*PIP and *Oleur*TIP can play regulatory roles in olive tree responses to various abiotic stresses and developmental processes. Further studies of the expression patterns of *Oleur*AQP gene family members by including these intron divergences would facilitate a more comprehensive understanding of the specific functions of these genes, and ultimately, it will aid in selecting candidate genes for functional analysis of their role in specific tissues.

However, overall, the expression divergence between some *Oleur*AQP duplicated gene pairs can be considered as an important first step in functional divergence, increasing the likelihood of their retention in the genome after the duplication event. Incidentally, it would also indicate that the *Oleur*AQP duplicates experienced functional diversity during the evolutionary process tied to the domestication of the olive trees, possibly combined with a gain of specific physiological roles during plant development and stress tolerance.

3. Materials and Methods

3.1. Genome-Wide Identification of Aquaporin Genes in Wild and Cultivated Olive Trees

Aquaporin genes and deduced protein sequences of the wild olive tree *O. europaea* subsp. *europaea* var. *sylvestris* (v1.0) were retrieved from the phytozome database, a Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Oeuropaea_er). The investigations were conducted using keyword queries ("Major Intrinsic Protein" and "Aquaporin"), complemented by *t*BLASTn searches with conservative criteria requiring a cut-off of *E*-value of 1.0^{-5} by using as queries the amino acid sequences of aquaporin sequences from *A. thaliana* (*At*AQPs) [35], *Populus trichocarpa* (*Poptr*AQPs) [30] and *Solanum lycopersicum* (*Sl*AQPs) [19], obtained from previous reports. Every putative aquaporin sequence (*Oeu*AQPs) was carefully scrutinized, including expected AQP motifs (NPA and ar/R selectivity filter) and the prediction of the transmembrane topology, with Interproscan from EMBL (http://www.ebi.ac.uk/Tools/pfa/iprscan/) and the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The identified sequences of *Oeu*AQP were subsequently employed as queries to recover, using *t*BLASTn searches, their analogous *Oleur*AQPs from the related cultivated olive tree *O. europaea* L. var. Picual available with the OliveTreeDB website (https://genomaolivar.dipujaen.es/db/index.php). This investigation was also complemented with a search by using annotation keyword queries ("Major Intrinsic Protein" and "Aquaporin"). Putative AQP sequences from the Oleaceae related species *Fraxinus excelsior* were retrieved from the Hardwood Genomics Project HWG website (DOI: https://doi.org/10.25504/FAIRsharing.srgkaf) by using *Sl*AQP, *Oeu*AQP and *Oleur*AQP sequences as queries. *F. excelsior* AQP sequences were specifically annotated for this work according to the standardized gene nomenclature which obeys the following format: genus (*three letter*)—species (*two letters*)—gene name (locus) or MIP nomenclature (subfamily x;y). All AQPs from *O.* var. *sylvestris* (*Oeu*AQPs), *O.* cv. Picual (*Oleur*AQPs) and *Fraxinus excelsior* (*Fraex*AQPs) are "roughly" named in the databases of origin. For this study, we renamed each sequence according to the standardized MIP nomenclature. In order to eliminate any ambiguity, each AQP appellation was phylogenetically established to the ones derived from *A. thaliana*, *P. trichocarpa* and *L. esculentum*. *Oeu*AQP and *Oleur*AQP protein names, accession numbers and sequences (genomic, CDS and deduced proteins) used in this work are listed in Figure 3 and Figures S1 and S2, respectively. *Fraex*AQP protein sequences are listed in Figure S3, respectively.

3.2. Multiple Sequence Alignment, Phylogenetic Analysis and Subfamily Classification

From the sixty-eight candidate AQP sequences recovered, fifty-two full-length deduced peptide sequences were aligned together with orthologous *At*AQPs, *Sl*AQPs and *Poptr*AQPs using ClustalW software (http://www.ebi.ac.uk/Tools/clustalx2/index.html). The results were viewed with Jalview (v2.10.5). A phylogenetic analysis based on the deduced peptide sequences of AQPs encoded by tomato, *A. thaliana* and poplar was used to assign the *Oeu*AQP sequences to the five established AQP subfamilies (i.e., PIPs, TIPs, NIPs, SIPs and XIPs). Functional annotations were initially attributed for most of the *Oeu*AQP loci in the databases of origin and renamed as mentioned above. The phylogenic studies were done with the *AQP*-deduced peptide sequences. A total of 323 amino acid sequences were aligned using the ClustalW progressive alignment method. Then, the phylogenetic tree of the sequence dataset was inferred using the maximum parsimony method. Maximum parsimony analysis was conducted using the subtree-pruning-regrafting (SPR) algorithm and was bootstrapped with 5000 replicates. All analyses were performed using the MEGA X program. A graphical representation and edition of the phylogenetic tree were performed with iTOL software (https://itol.embl.de/) [98].

3.3. Chromosomal Location, Gene Structure, Structural Features of OeuAQPs Values

The phytozome assembly database was used to reveal the chromosomal location for each *Oeu*AQP and to determine their related intron/exon structure. The exon–intron organization of

OeuAQP genes were compared together using Gene Structure Display Server (GSDS) software (http://gsds.cbi.pku.edu.cn) [99] based on the alignment of CDS sequences with the corresponding genomic sequences. Amino acid (AA) translations were obtained using Bioinformatics Resource Portal Expasy software (https://web.expasy.org/translate/). Protein features such as the theoretical isoelectric point (*p*I), the molecular weight (MW) and the grand average of hydropathicity (GRAVY) were calculated using the ProtParam program (https://web.expasy.org/protparam/). The classical conserved domains, i.e., NPA motifs, the ar/R filter, the Froger's positions and a panel of post-translational and regulatory positions (i.e., phosphorylation, methylation, the leucine blocking residue and the histidine residue involved in gating) were manually identified based on multiple sequence alignments with heterologous *Sl*AQPs and *At*AQPs. The transmembrane regions were predicted using the TMHMM software tool (www.cbs.dtu.dk/services/TMHMM/) [100] and SOSUI tools (http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html) [101] and, if necessary, manually corrected with heterologous *Sl*AQP and *At*AQP sequences. The subcellular localizations of *Oeu*AQPs were predicted using the online tools WoLF PSORT (http://wolf-psort.hgc.jp) and Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) [102]. The conserved motifs were detected using the multiple em for motif elicitation (MEME) software (http://meme-suite.org/tools/meme) [103], with the maximum number of motifs set as 15. Specificity-determining positions (SDP1–9) were collected from multiple sequence alignments, which were performed using ClustalW as described in [21]. Protein sequences corresponding to functional characterized AQPs transporting non-aqua substrates were obtained from Genbank according to related literature. To evaluate the evolution pressure of AQP genes, the *K*a/*K*s ratio was calculated. The non-synonymous (*K*a) and synonymous (*K*s) substitution ratio of AQP genes was calculated to test selection pressure. Transcript sequences and their deduced protein previously aligned with ClustalW (https://www.genome.jp/tools-bin/clustalw) were submitted to the online program PAL2NAL (http://www.bork.embl.de/pal2nal/) [104]. A ratio significantly greater than 1 indicates positive selective pressure. A ratio around 1 indicates either neutral evolution at the protein level or an averaging of sites under positive and negative selective pressures. A ratio less than 1 indicates pressures to conserve protein sequence.

3.4. RNA-Seq Analysis

*Oleur*AQP transcriptional steady state levels on plant organ/tissues were recorded from the plant samples that were obtained and studied by [105]. Briefly, samples were collected from roots, stems, meristems, leaves, flowers and fruits (epicarp and mesocarp) of three healthy 10-year-old cv. Picual olive trees under field conditions. Two biological replicates per sample were sequenced, and each biological replicate consisted of an equilibrated pool of three plant RNAs.

Early development plant samples were obtained and studied by [92]. Briefly, seeds of cv. Arbequina were induced to geminate. Seedlings were grown in vitro under chamber conditions with a 16 h photoperiod of fluorescent light and at 25 °C constant temperature until they were two months old and then potted and grown in a conditioned greenhouse (25 °C). The aerial parts of 10 plants were collected at 1, 2, 3, 4, 5 and 6 months after seed activation. Two biological replicates of ten pooled plants per sample were sequenced.

Verticillium dahliae infected plant samples were obtained and studied by [87,106]. Briefly, 40 plants were infected by root-dip inoculation in a conidia suspension (10^7 conidia mL^{-1}) of defoliating *V. dahliae* isolate V937I. Forty non-inoculated plants manipulated in the same way were used as a control group of plants in the absence of the pathogen. Three plants were pooled for each biological replicate.

The cold stress was induced and the RNA-seq performed by [88]. Briefly, 35 four-month-old potted olive plants of the cv. Picual acclimated at 24 °C, and then incubated with a 14 h photoperiod of fluorescent light at 65 µmoles m^2 s (10 °C day/4 °C night) for 10 days and constant 76%–78% relative humidity (RH). An additional group of 15 plants were used as a control treatment. Aerial tissues were then harvested at 0, 24 h and 10 days (three plants/time point). Three plants were pooled for each biological replicate.

Each sample was collected and frozen in liquid nitrogen immediately and stored at −80 °C prior to the molecular assays. The total RNA was extracted using a Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's instructions. Two technical replicates of each sample were then sequenced on different lanes in the flow cell using paired-end sequencing (101 × 2) in an Illumina® HiSeq sequencer (Illumina, San Diego, CA, USA).

4. Conclusions

To the best of our knowledge, this is the first comprehensive study and systematic genome-wide analysis of AQP gene families in *O. europaea* L. This study highlighted several novel findings explaining the structural conservation and possible functional diversity of AQPs in wild and cultivated olive tree varieties and their involvement in cell responses to various challenged environments. Overall, 68 and 127 non-redundant AQP genes were identified in the wild variety (*Oeu*AQPs) and the cultivated cultivar (*Oleur*AQPs), including 52 and 79 putative functional AQPs, respectively. All *Olea* AQP belong to five distinct subfamilies: PIP, TIP, XIP, NIP and SIP. This doubling observed in the cultivated cultivar could be related to classical polyploidization and a recent gene duplication phenomenon. Despite this expansion, every paralogous full-length AQP pair encodes very similar proteins, and the overall gene structure is highly conserved. This conservation results from a quite recent expansion of the sequences, for which strong purifying selections act on the duplicated gene pairs identified in the domesticated species. Expression profiling of *Oleur*AQP genes revealed high expression of various PIP1s, PIP2s and TIPs, in almost every organ and tissue, suggesting the probable important roles of these AQPs in the early developmental processes and abiotic and biotic stress responses. Because aquaporins play prominent roles in a plethora biological processes, further studies are required to ascertain the functions of the individual selected *OleurAQP* genes identified in this study and to reveal more functional mechanisms for these genes. All these results increase our knowledge of the molecular mechanisms behind the actions of AQPs in olive domestication, and the integration of bioinformatics analysis with biological experiment validations will provide further insight into the key roles that some AQPs play in development processes and stress tolerance in domesticated olive trees.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/11/4183/s1. Figure S1. Detailed *Oeu*MIP sequences (genomic, CDS and protein sequences) from the wild olive species *O. europaea* var. *sylvestris*. Figure S2. Detailed *Oleur*AQP sequences (genomic, CDS and protein sequences) from the domesticated olive species *Olea europaea* cv. Picual. Figure S3. Deduced protein sequences of *Fraex*AQPs from *Fraxinus excelsior*. Sequences are listed according to their subfamilies. Pseudogenes and/or incomplete sequences are listed at the end of the document. Figure S4. Phylogenetic analysis of protein sequences of the 52 *Oeu*AQPs and the 79 *Oleur*AQPs with *Fraxinus excelsior*, *Arabidopsis thaliana* and *Populus trichocarpa* hortologs. This supplemental file is similar to Figure 1 and it is shown in high definition for ease of reading. Figure S5. Amino acid sequence alignments per subfamily of the *Oeu*AQPs members from *Olea europaea* var. *sylvestris*. 5A, PIP; 5B, TIP; 5C, XIP; 5D, NIP; 5E, SIP. Legends are detailed at the end of each subfamily alignment. Figure S6. Structuration of the 15 motives identified in the 52 full-length *Oeu*AQP protein sequences (Figure 2). Motif analysis was performed with the MEME web server. Details of the amino acid constituting these motifs are listed in the legend, and residue occurrences are provided on page 2 of the file. Figure S7. Comparison of protein structural features, evolution pressure and exon-intron structures of full-length *Oeu*AQP and *Oleur*AQP genes from the wild *Olea europaea* var. *sylvestris* and the cultivated *Olea europaea* cv. Picual, respectively. Figure S8. SDP analysis of *Oeu*AQPs from the wild olive species *Olea europaea* var. *sylvestris* from alignments with putative amino acid sequences of AQPs transporting non-aqua substrates. Multiple alignments were performed using ClustalW. Shown are the sequences containing the nine SD positions which are highlighted in yellow and the representative sequences are marked in blue. The Genbank accession numbers: *At*PIP1;2 (Q06611), *At*PIP2;1 (P43286), *At*PIP2;4 (Q9FF53), *At*TIP1;1 (P25818), *At*TIP1;2 (Q41963), *At*TIP1;3 (NP;192056), *At*TIP2;1 (Q41951), *At*TIP2;3 (Q9FGL2), *At*TIP4;1 (O82316), *At*TIP5;1 (NP;190328), *At*NIP1;1 (CAA16760), *At*NIP1;2 (Q8LFP7), *At*NIP5;1 (NP;192776), *At*NIP6;1 (NP;178191), *Cm*NIP2;1 (BAK09176), *Cp*NIP1 (CAD67694), *Gm*NOD26 (P08995), *Gm*NIP2;1 (XP;003534451), *Gm*NIP2;2 (NP;001240190), *Hv*PIP1;3 (BAA23745), *Hv*PIP1;4 (BAF33068), *Hv*PIP2;1 (BAA23744), *Hv*NIP2;1 (BAH24163), *Nt*AQP1 (O24662), *Nt*TIPa (Q9XG70), *Nt*XIP1;1α (ADO66667), *Nt*XIP1;1β (ADO66666), *Os*NIP2;1 (Q6Z2T3), *Os*NIP2;2 (Q67WJ8), *Os*NIP3;2 (Q7EYH7), *Sl*XIP1;1α (ADO66672), *Sl*XIP1;1β (ADO66671), *St*XIP1;1α (ADO66669), *St*XIP1;1β (ADO66670), *Ta*Lsi1 (ADM47602), *Ta*TIP2;1 (AAS19468), *Ta*TIP2;2 (AAS19469), *Tg*TIP1;1 (BAL41683), *Tg*TIP1;2 (BAL41684), *Zm*PIP1;1 (Q41870), *Zm*PIP1;5 (Q9AR14), *Zm*NIP2;1 (Q19KC1), *Zm*NIP2;2 (Q9ATN2). Table S1. Nomenclatures of *Oeu*APQ sequences from the wild olive species *Olea europaea* var. *sylvestris*. *Oeu*AQP sequences (genomic, CDS and protein) are detailed in Supplemental Figure S1. Table S2. Correspondence between aquaporin gene sequences from *Olea europaea* var. *sylvestris* (*Oeu*APQ) and *O. europaea*

cv. Picual (*Oleur*AQP). Table S3. Percent identity matrix between the 52 full-length *Oeu*AQP sequences (edited with Clustal2.1). Table S4. Expression profile of *Oleur*AQP genes from the cultivated *Olea europaea* cv. Picual in different physiological conditions. Table S5. MIP expression raw data for each biological sample used to calculate RNA-seq of Figure 6.

Author Contributions: Conceived and designed the experiments, J.-S.V. and M.F.; performed the experiments and analyzed the data, M.F., B.F., J.A.R.-T., X.Q., L.F. and A.G.-D.; validation, M.F., F.L., Z.Z., P.R.-D, P.L. and J.-S.V.; formal analysis, B.F., F.L., Z.Z., and J.-S.V; resources, F.L., J.A.R.-T., Z.Z. and X.Q.; data curation, B.F., F.L., M.F. and J.-S.V.; writing—original draft preparation, J.-S.V. and M.F.; review and editing, F.L., Z.Z., P.R.-D., P.L. and J.-S.V.; supervision, M.F. and J.-S.V.; project administration and funding acquisition, J.S.-V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was granted access to the French HPC resources of IDRIS under the allocation 2019-A0060710751 made by GENCI, and I-SITE CAP 20-25 (ANR grant 16-IDEX-0001) Emergence 2017 from the University of Clermont-Auvergne.

Acknowledgments: The authors thank those contributors who make the *Olea europaea* var. *sylvestris* genome data accessible in public databases of Phytozome. We are grateful to Céline Sac and Caroline Savel for their technical assistance in molecular biology. We are indebted to the two anonymous reviewers for their helpful suggestions.

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

Abbreviations

AQP	Aquaporin
AQGP	Aquaglyceroporin
ar/R	"Aromatic/arginine" selectivity filter
GIP	GlpF-like intrinsic protein
GRAVY	Grand average of hydropathicity
HIP	Hybrid Intrinsic protein
H_2O_2	Hydrogen peroxide
LIP	Large intrinsic Protein
MIP	Major Intrinsic Protein
MW	Molecular weight
NH_3	Ammonia
NIP	Nodulin 26-like intrinsic protein
NPA	"Asparagine–proline–alanine" motif
*p*I	Theoretical isoelectric point
PIP	Plasma membrane intrinsic protein
RPKM	Reads per kilo base per million mapped reads
TMH	trans-alpha helical transmembranes region
TIP	Tonoplast intrinsic protein
Si	Silicic acid
SIP	Small basic intrinsic protein
XIP	X-intrinsic protein

References

1. Therios, I.O. *Olives. Crop Production Science in Horticulture Series*; CABI: Wallingford, UK, 2009; p. 425.
2. Maurel, C.; Boursiac, Y.; Luu, D.T.; Santoni, V.; Shahzad, Z.; Verdoucq, L. Aquaporins in plants. *Physiol. Rev.* **2015**, *95*, 1321–1358. [CrossRef] [PubMed]
3. Singh, R.K.; Deshmukh, R.; Muthamilarasan, M.; Rani, R.; Prasad, M. Versatile roles of aquaporin in physiological processes and stress tolerance in plants. *Plant Physiol. Biochem.* **2020**, *149*, 178–189. [CrossRef]
4. Murata, K.; Mitsuoka, K.; Hirai, T.; Walz, T.; Agre, P.; Heymann, J.B.; Engel, A.; Fujiyoshi, Y. Structural determinants of water permeation through aquaporin-1. *Nature* **2000**, *407*, 599–605. [CrossRef] [PubMed]
5. Savage, D.F.; O'Connell, J.D., 3rd; Miercke, L.J.; Finer-Moore, J.; Stroud, R.M. Structural context shapes the aquaporin selectivity filter. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 17164–17169. [CrossRef] [PubMed]
6. Hove, R.M.; Bhave, M. Plant aquaporins with non-aqua functions: Deciphering the signature sequences. *Plant. Mol. Biol.* **2011**, *75*, 413–430. [CrossRef]

7. Azad, A.K.; Ahmed, J.; Alum, M.A.; Hasan, M.M.; Ishikawa, T.; Sawa, Y.; Katsuhara, M. Genome-wide characterization of major intrinsic proteins in four grass plants and their non-aqua transport selectivity profiles with comparative perspective. *PLoS ONE* **2016**, *11*, e0157735. [CrossRef]
8. Fetter, K.; Van Wilder, V.; Moshelion, M.; Chaumont, F. Interactions between plasma membrane aquaporins modulate their water channel activity. *Plant Cell* **2004**, *16*, 215–228. [CrossRef]
9. Kreida, S.; Tornroth-Horsefield, S. Structural insights into aquaporin selectivity and regulation. *Curr. Opin. Struct. Biol.* **2015**, *33*, 126–134. [CrossRef] [PubMed]
10. Yasui, M.; Hazama, A.; Kwon, T.H.; Nielsen, S.; Guggino, W.B.; Agre, P. Rapid gating and anion permeability of an intracellular aquaporin. *Nature* **1999**, *402*, 184–187. [CrossRef]
11. Ivanov, I.I.; Loktyushkin, A.V.; Guskova, R.A.; Vasilev, N.S.; Fedorov, G.E.; Rubin, A.B. Oxygen channels of erythrocyte membrane. *Dokl. Biochem. Biophys.* **2007**, *414*, 137–140. [CrossRef]
12. Hub, J.S.; de Groot, B.L. Mechanism of selectivity in aquaporins and aquaglyceroporins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1198–1203. [CrossRef] [PubMed]
13. Chaumont, F.; Tyerman, S.D. Aquaporins: Highly regulated channels controlling plant water relations. *Plant Physiol.* **2014**, *164*, 1600–1618. [CrossRef] [PubMed]
14. Khabudaev, K.V.; Petrova, D.P.; Grachev, M.A.; Likhoshway, Y.V. A new subfamily LIP of the major intrinsic proteins. *BMC Genom.* **2014**, *15*, 173. [CrossRef]
15. Ishibashi, K.; Morishita, Y.; Tanaka, Y. *The Evolutionary Aspects of Aquaporin Family Aquaporins*; Springer: New York, NY, USA, 2017; pp. 35–50.
16. Secchi, F.; Lovisolo, C.; Uehlein, N.; Kaldenho, V.R.; Schubert, A. Isolation and functional characterization of three aquaporins from olive (Olea europaea L.). *Planta* **2007**, *225*, 381–392. [CrossRef] [PubMed]
17. López-Escudero, F.J.; Mercado-Blanco, J. Verticillium wilt of olive: A case study to implement an integrated strategy to control a soil-borne pathogen. *Plant Soil* **2011**, *344*, 1–50. [CrossRef]
18. Zou, Z.; Yang, L.; Gong, J.; Mo, Y.; Wang, J.; Cao, J.; An, F.; Xie, G. Genome-Wide Identification of Jatropha curcas Aquaporin Genes and the Comparative Analysis Provides Insights into the Gene Family Expansion and Evolution in Hevea brasiliensis. *Front. Plant Sci.* **2016**, *7*, 395. [CrossRef]
19. Reuscher, S.; Akiyama, M.; Mori, C.; Aoki, K.; Shibata, D.; Shiratake, K. Genome-Wide Identification and Expression Analysis of Aquaporins in Tomato. *PLoS ONE* **2013**, *8*, e79052. [CrossRef]
20. Gupta, A.B.; Sankararamakrishnan, R. Genome-wide analysis of major intrinsic proteins in the tree plant Populus trichocarpa: Characterization of XIP subfamily of aquaporins from evolutionary perspective. *BMC Plant. Biol.* **2009**, *20*, 134. [CrossRef]
21. Zou, Z.; Gong, J.; Huang, Q.; Mo, Y.; Yang, L.; Xie, G. Gene Structures, Evolution, Classification and Expression Profiles of the Aquaporin Gene Family in Castor Bean (Ricinus communis L.). *PLoS ONE* **2015**, *10*, e0141022. [CrossRef]
22. Deokar, A.A.; Tar'an, B. Genome-Wide Analysis of the Aquaporin Gene Family in Chickpea (Cicer arietinum L.). *Front. Plant Sci.* **2016**, *7*, 1802. [CrossRef]
23. Shivaraj, S.M.; Deshmukh, R.; Humira, S.; Bélanger, R.R. Identification and characterization of aquaporin genes in Arachis duranensis and Arachis ipaensis genomes, the diploid progenitors of peanut. *BMC Genom.* **2019**, *20*, 222. [CrossRef] [PubMed]
24. Deshmukh, R.K.; Vivancos, J.; Ramakrishnan, G.; Guérin, V.; Carpentier, G.; Sonah, H.; Labbé, C.; Isenring, P.; Belzile, F.J.; Bélanger, R.R. A precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *Plant J.* **2015**, *83*, 489–500. [CrossRef]
25. Danielson, J.A.H.; Johanson, U. Unexpected complexity of the aquaporin gene family in the moss Physcomitrella patens. *BMC Plant Biol.* **2008**, *8*, 45. [CrossRef] [PubMed]
26. Venkatesh, J.; Yu, J.W.; Gaston, D.; Park, S.W. Molecular evolution and functional divergence of X-intrinsic protein genes in plants. *Mol. Genet. Genom.* **2015**, *290*, 443–460. [CrossRef]
27. Diehn, T.A.; Pommerrenig, B.; Bernhardt, N.; Hartmann, A.; Bienert, G.P. Genome-wide identification of aquaporin encoding genes in Brassica oleracea and their phylogenetic sequence comparison to Brassica crops and Arabidopsis. *Front. Plant Sci.* **2015**, *6*, 166. [CrossRef]
28. Deshmukh, R.; Bélanger, R.R. Molecular evolution of aquaporins and silicon influx in plants. *Funct. Ecol.* **2016**, *30*, 1277–1285. [CrossRef]

29. Martins, C.d.P.; Pedrosa, A.M.; Du, D.; Gonçalves, L.P.; Yu, Q.; Gmitter, F.G., Jr.; Costa, C.G.M. Genome-wide characterization and expression analysis of major intrinsic proteins during abiotic and biotic stresses in sweet orange (Citrus sinensis L. Osb.). *PLoS ONE* **2015**, *10*, e0138786.
30. Lopez, D.; Bronner, G.; Brunel, N.; Auguin, D.; Bourgerie, S.; Brignolas, F.; Carpin, S.; Tournaire-Roux, C.; Maurel, C.; Fumanal, B.; et al. Insights into Populus XIP aquaporins: Evolutionary expansion, protein functionality, and environmental regulation. *J. Exp. Bot.* **2012**, *63*, 2217–2230. [CrossRef]
31. Turgay, U.; Wu, Z.; Sterck, L.; Turktas, M.; Lohaus, R.; Li, Z.; Yang, M.; He, L.; Deng, T.; Escalante, F.J.; et al. Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9413–CE9422.
32. Jiménez-Ruiz, J.; Ramírez-Tejero, J.A.; Fernández-Pozo, N.d.l.O.; Leyva-Pérez, M.; Yan, H.; Montes, E.; Rodríguez-Ariza, M.A.; Beuzón, C.R.; Valpuesta, V.; Luque, F. Transposon activation is a major driver in the genome evolution of cultivated olive trees (Olea europaea L.). *Plant Genome* **2020**, *13*, e20010.
33. Sonah, H.; Deshmukh, R.K.; Labbé, C.; Bélanger, R.R. Analysis of aquaporins in Brassicaceae species reveals high-level of conservation and dynamic role against biotic and abiotic stress in canola. *Sci. Rep.* **2017**, *7*, 2771. [CrossRef]
34. Li, W.; Zhang, D.; Zhu, G.; Mi, X.; Guo, W. Combining genome-wide and transcriptome-wide analyses reveal the evolutionary conservation and functional diversity of aquaporins in cotton. *BMC Genomics* **2019**, *20*, 538. [CrossRef] [PubMed]
35. Johanson, U.; Karlsson, M.; Johansson, I.; Gustavsson, S.; Sjövall, S.; Fraysse, L.; Weig, A.R.; Kjellbom, P. The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol.* **2001**, *126*, 1358–1369. [CrossRef] [PubMed]
36. Chaumont, F.; Barrieu, F.; Jung, R.; Chrispeels, M.J. Plasma membrane intrinsic proteins from maize cluster in two sequence subgroups with differential aquaporin activity. *Plant Physiol.* **2000**, *122*, 1025–1034. [CrossRef] [PubMed]
37. Katsuhara, M.; Hanba, Y.T.; Shiratake, K.; Maeshima, M. Expanding roles of plant aquaporins in plasma membranes and cell organelles. *Funct. Plant Biol.* **2008**, *35*, 1–14. [CrossRef]
38. Mitani-Ueno, N.; Yamaji, N.; Zhao, F.; Ma, J.F. The aromatic/arginine selectivity filter of NIP aquaporins plays a critical role in substrate selectivity for silicon, boron, and arsenic. *J. Exp. Bot.* **2011**, *62*, 4391–4398. [CrossRef] [PubMed]
39. Sui, H.; Han, B.G.; Lee, J.K.; Walian, P.; Jap, B.K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **2001**, *414*, 865–872. [CrossRef] [PubMed]
40. Fu, D.; Libson, A.; Miercke, L.J.; Weitzman, C.; Nollert, P.; Krucinski, J.; Stroud, R.M. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **2000**, *290*, 481–486.
41. Ma, J.F.; Yamaji, N. Silicon uptake and accumulation in higher plants. *Trends Plant Sci.* **2006**, *11*, 392–397. [CrossRef]
42. Takano, J.; Wada, M.; Ludewig, U.; Schaaf, G.; von Wirén, N.; Fujiwara, T. The Arabidopsis major intrinsic protein NIP5;1 is essential for efficient boron uptake and plant development under boron limitation. *Plant Cell* **2006**, *18*, 1498–1509. [CrossRef]
43. Lopez, D.; Ben Amira, M.; Brown, D.; Muries, B.; Brunel-Michac, N.; Bourgerie, S.; Porcheron, B.; Lemoine, R.; Chrestin, H.; Mollison, E.; et al. The Hevea brasiliensis XIP aquaporin subfamily: Genomic, structural and functional characterizations with relevance to intensive latex harvesting. *Plant Mol. Biol.* **2016**, *91*, 375–396. [CrossRef] [PubMed]
44. Bienert, G.P.; Bienert, M.D.; Jahn, T.P.; Boutry, M.; Chaumont, F. *Solanaceae* XIPs are plasma membrane aquaporins that facilitate the transport of many uncharged substrates. *Plant J.* **2011**, *66*, 306–317. [CrossRef] [PubMed]
45. Bienert, M.D.; Muries, B.; Crappe, D.; Chaumont, F.; Bienert, G.P. Overexpression of X Intrinsic Protein 1;1 in Nicotiana tabacum and Arabidopsis reduces boron allocation to shoot sink tissues. *Plant Direct* **2019**, *3*, e00143. [CrossRef] [PubMed]
46. Froger, A.; Tallur, B.; Thomas, D.; Delamarche, C. Prediction of functional residues in water channels and related proteins. *Protein Sci.* **1998**, *7*, 1458–1468. [CrossRef] [PubMed]
47. Gaspar, M.; Bousser, A.; Sissoëff, I.; Roche, O.; Hoarau, J.; Mahé, A. Cloning and characterization of ZmPIP1-5b, an aquaporin transporting water and urea. *Plant Sci.* **2003**, *165*, 21–31. [CrossRef]

48. Bienert, G.P.; Heinen, R.B.; Berny, M.C.; Chaumont, F. Maize plasma membrane aquaporin ZmPIP2;5, but not ZmPIP1;2, facilitates transmembrane diffusion of hydrogen peroxide. *Biochim. Biophys. Acta* **2014**, *1838*, 216–222. [CrossRef] [PubMed]
49. Heckwolf, M.; Pater, D.; Hanson, D.T.; Kaldenhoff, R. The Arabidopsis thaliana aquaporin AtPIP1;2 is a physiologically relevant CO_2 transport facilitator. *Plant J.* **2011**, *67*, 795–804. [CrossRef]
50. Flexas, J.; Ribas-Carbo, M.; Hanson, D.T.; Bota, J.; Otto, B.; Cifre, J.; McDowell, N.; Medrano, H.; Kaldenhoff, R. Tobacco aquaporin NtAQP1 is involved in mesophyll conductance to CO_2 in vivo. *Plant J.* **2006**, *48*, 427–439. [CrossRef]
51. Uehlein, N.; Lovisolo, C.; Siefritz, F.; Kaldenhoff, R. The tobacco aquaporin NtAQP1 is a membrane CO_2 pore with physiological functions. *Nature* **2003**, *425*, 734–737. [CrossRef] [PubMed]
52. Choi, W.G.; Roberts, D.M. Arabidopsis NIP2;1, a major intrinsic protein trans- porter of lactic acid induced by anoxic stress. *J. Biol. Chem.* **2007**, *282*, 24209–24218. [CrossRef] [PubMed]
53. Grégoire, C.; Rémus-Borel, W.; Vivancos, J.; Labbé, C.; Belzile, F.; Bélanger, R.R. Discovery of a multigene family of aquaporin silicon transporters in the primitive plant Equisetum arvense. *Plant J.* **2012**, *72*, 320–330. [CrossRef]
54. Loqué, D.; Ludewig, U.; Yuan, L.; Wiren, N.V. Tonoplast intrinsic proteins AtTIP2;1 and AtTIP2;3 facilitate NH3 transport into the vacuole. *Plant Physiol.* **2005**, *137*, 671–680. [CrossRef] [PubMed]
55. Suga, S.; Maeshima, M. Water channel activity of radish plasma membrane aquaporins heterologously expressed in yeast and their modification by site-directed mutagenesis. *Plant Cell Physiol.* **2004**, *45*, 823–830. [CrossRef] [PubMed]
56. Bienert, G.P.; Cavez, D.; Besserer, A.; Berny, M.C.; Gilis, D.; Rooman, M.; Chaumont, F. A conserved cysteine residue is involved in disulfide bond formation between plant plasma membrane aquaporin monomers. *Biochem. J.* **2012**, *445*, 101–111. [CrossRef]
57. Forrest, K.L.; Bhave, M. Major intrinsic proteins (MIPs) in plants: A complex gene family with major impacts on plant phenotype. *Funct. Integr. Genom.* **2007**, *7*, 263–289. [CrossRef] [PubMed]
58. Mizutani, M.; Watanabe, S.; Nakagawa, T.; Maeshima, M. Aquaporin NIP2;1 is mainly localized to the ER membrane and shows root-specific accumulation in Arabidopsis thaliana. *Plant Cell Physiol.* **2006**, *47*, 1420–1426. [CrossRef] [PubMed]
59. Ishikawa, F.; Suga, S.; Uemura, T.; Sato, M.H.; Maeshima, M. Novel type aquaporin SIPs are mainly localized to the ER membrane and show cell-specific expression in Arabidopsis thaliana. *FEBS Lett.* **2005**, *579*, 5814–5820. [CrossRef]
60. Sakurai, J.; Ishikawa, F.; Yamaguchi, T.; Uemura, M.; Maeshima, M. Identification of 33 rice aquaporin genes and analysis of their expression and function. *Plant Cell Physiol.* **2005**, *46*, 1568–1577. [CrossRef]
61. Carqueijeiro, I.; Noronha, H.; Duarte, P.; Gerôs, H.V.; Sottomayor, M. Vacuolar transport of the medicinal alkaloids from Catharanthus roseus is mediated by a proton driven antiport. *Plant Physiol.* **2013**, *162*, 1486–1496. [CrossRef]
62. Gattolin, S.; Sorieul, M.; Frigerio, L. Mapping of tonoplast intrinsic proteins in maturing and germinating Arabidopsis seeds reveals dual localization of embryonic TIPs to the tonoplast and plasma membrane. *Mol. Plant* **2011**, *4*, 180–189. [CrossRef]
63. Boursiac, Y.; Prak, S.; Boudet, J.; Postaire, O.; Luu, D.T.; Tournaire-Roux, C.; Santoni, V.; Maurel, C. The response of Arabidopsis root water transport to a challenging environment implicates reactive oxygen species- and phosphorylation- dependent internalization of aquaporins. *Plant Signal. Behav.* **2008**, *3*, 1096–1098. [CrossRef] [PubMed]
64. Luu, D.T.; Martinière, A.; Sorieul, M.; Runions, J.; Maurel, C. Fluorescence recovery after photobleaching reveals high cycling dynamics of plasma membrane aquaporins in Arabidopsis roots under salt stress. *Plant J.* **2012**, *69*, 894–905. [CrossRef]
65. Innan, H.; Kondrashov, F. The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **2010**, *11*, 97–108. [CrossRef] [PubMed]
66. Varshney, R.K.; Song, C.; Saxena, R.K.; Azam, S.; Yu, S.; Sharpe, A.G.; Cannon, S.; Beak, J.; Rosen, B.D.; Taran, B.; et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **2013**, *31*, 240–246. [CrossRef] [PubMed]
67. Abascal, F.; Irisarri, I.; Zardoya, R. Diversity and evolution of membrane intrinsic proteins. *Biochim. Biophys. Acta (Bba) Gen. Subj.* **2014**, *1840*, 1468–1481. [CrossRef] [PubMed]

68. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [CrossRef]
69. Tanaka, K.M.; Takahasi, K.R.; Takano-Shimizu, T. Enhanced fixation and preservation of a newly arisen duplicate gene by masking deleterious loss-of-function mutations. *Genet. Res.* **2009**, *91*, 267–280. [CrossRef]
70. Deshmukh, R.K.; Vivancos, J.; Guérin, V.; Sonah, H.; Labbé, C.; Belzile, F.; Belanger, R.R. Identification and functional characterization of silicon transporters in soybean using comparative genomics of major intrinsic proteins in Arabidopsis and rice. *Plant Mol. Biol.* **2013**, *83*, 303–315. [CrossRef]
71. Zhou, Y.; Tao, J.; Ahammed, G.J.; Li, J.; Yang, Y. Genome-wide identification and expression analysis of aquaporin gene family related to abiotic stress in watermelon. *Genome* **2019**, *62*, 643–656. [CrossRef]
72. Xu, G.X.; Guo, C.C.; Shan, H.Y.; Kong, H.Z. Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1187–1192. [CrossRef]
73. Fraser, J.A.; Huang, J.C.; Pukkila-Worley, R.; Alspaugh, J.A.; Mitchell, T.G.; Heitman, J. Chromosomal translocation and segmental duplication in Cryptococcus neoformans. *Eukaryot. Cell* **2005**, *4*, 401–406. [CrossRef] [PubMed]
74. Yaneff, A.; Sigaut, L.; Marquez, M.; Alleva, K.; Pietrasanta, L.I.; Amodeo, G. Heteromerization of PIP aquaporins affects their intrinsic permeability. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 231–236. [CrossRef] [PubMed]
75. Bienert, M.D.; Diehn, T.A.; Richet, N.; Chaumont, F.; Bienert, G.P. Heterotetramerization of plant PIP1 and PIP2 aquaporins Is an evolutionary ancient feature to guide PIP1 plasma membrane localization and function. *Front. Plant Sci.* **2018**, *9*, 382. [CrossRef] [PubMed]
76. Johnson, K.D.; Herman, E.M.; Chrispeels, M.J. An abundant, highly conserved tonoplast protein in seeds. *Plant Physiol.* **1989**, *91*, 1006–1013. [CrossRef]
77. Shivaraj, S.M.; Deshmukh, R.K.; Rai, R.; Bélanger, R.; Agrawal, P.K.; Dash, P.K. Genome-wide identification, characterization, and expression profile of aquaporin gene family in flax (Linum usitatissimum). *Sci. Rep.* **2017**, *7*, 46137. [CrossRef]
78. Footitt, S.; Clewes, R.; Feeney, M.; Finch-Savage, W.E.; Frigerio, L. Aquaporins influence seed dormancy and germination in response to stress. *Plant Cell Environ.* **2019**, *42*, 2325–2339. [CrossRef]
79. Hu, W.; Hou, X.; Huang, C.; Yan, Y.; Tie, W.; Ding, Z.; Wei, Y.; Liu, J.; Miao, H.; Lu, Z. Genome-wide identification and expression analyses of aquaporin gene family during development and abiotic stress in banana. *Int. J. Mol. Sci.* **2015**, *16*, 19728–19751. [CrossRef]
80. Kong, W.; Bendahmane, M.; Fu, X. Genome-wide Identification and characterization of aquaporins and their role in the flower opening processes in Carnation (Dianthus caryophyllus). *Molecules* **2018**, *23*, 1895. [CrossRef]
81. Min, X.; Wu, H.; Zhang, Z.; Wei, X.; Jin, X.; Ndayambaza, B.; Wang, Y.; Liu, W. Genome-wide identification and characterization of the aquaporin gene family in Medicago truncatula. *J. Plant. Biochem. Biotechnol.* **2019**, *28*, 320–335. [CrossRef]
82. Zhu, Y.; Yang, L.; Liu, N.; Yang, J.; Zhou, X.; Xia, Y.; He, Y.; He, Y.; Gong, H.; Ma, D.; et al. Genome-wide identification, structure characterization, and expression pattern profiling of aquaporin gene family in cucumber. *BMC Plant Biol.* **2019**, *19*, 345. [CrossRef]
83. Prado, K.; Maurel, C. Regulation of leaf hydraulics: From molecular to whole plant levels. *Front. Plant. Sci* **2013**, *4*, 255. [CrossRef] [PubMed]
84. Muries, B.; Mom, R.; Benoit, P.; Brunel-Michac, N.; Cochard, H.; Drevet, P.; Petel, G.; Badel, E.; Fumanal, B.; Gousset-Dupont, A.; et al. Aquaporins and water control in drought-stressed poplar leaves: A glimpse into the extraxylem vascular territories. *Environ. Exp. Bot.* **2019**, *162*, 25–37. [CrossRef]
85. Tian, S.; Wang, X.; Li, P.; Wang, H.; Ji, H.; Xie, J.; Qiu, Q.; Shen, D.; Dong, H. Plant Aquaporin *At*PIP1;4 Links Apoplastic H_2O_2 Induction to Disease Immunity Pathways. *Plant Physiol.* **2016**, *171*, 1635–1650. [CrossRef]
86. Wang, H.; Schoebel, S.; Schmitz, F.; Donga, H. Characterization of aquaporin-driven hydrogen peroxide transport. *Biochim. Biophys. Acta* **2020**, *1862*, 183065. [CrossRef] [PubMed]
87. Jiménez-Ruiz, J.; Leyva-Pérez, M.d.l.O.; Schilirò, E.; Barroso, J.B.; Bombarely, A.; Mueller, L.; Mercado-Blanco, J.; Luque, F. Transcriptomic analysis of the Olea europaea L. roots during the Verticillium dahliae early infection process. *Plant Genome* **2017**, *10*, 1–15.

88. Leyva-Pérez, M.d.l.O.; Valverde-Corredor, A.; Valderrama, R.; Jiménez-Ruiz, J.; Muñoz-Merida, A.; trelles, O.; Barroso, J.B.; Mercado-Blanco, J.; Luque, F. Early and delayed long-term transcriptional changes and short-term transient responses during cold acclimation in olive leaves. *Dna Res.* **2015**, *22*, 1–11. [CrossRef]
89. Wang, L.; Yao, L.; Hao, X.; Li, N.; Wang, Y.; Ding, C.; Lei, L.; Qian, W.; Zeng, J.; Yang, Y.; et al. Transcriptional and physiological analyses reveal the association of ROS metabolism with cold tolerance in tea plant. *Environ. Exp. Bot.* **2019**, *160*, 45–68. [CrossRef]
90. Soveral, G.; Veiga, A.; Loureiro-Dias, M.C.; Tanghe, A.; Dijck, P.V.; Moura, T.F. Water channels are important for osmotic adjustments of yeast cells at low temperature. *Microbiology* **2006**, *152*, 1515–1521. [CrossRef]
91. Peng, Y.; Arora, R.; Li, G.; Wang, X.; Fessehaie, A. Rhododendron catawbiense plasma membrane intrinsic proteins are aquaporins, and their over-expression compromises constitutive freezing tolerance and cold acclimation ability of transgenic Arabidopsis plants. *Plant Cell Environ.* **2008**, *31*, 1275–1289. [CrossRef]
92. Jiménez-Ruiz, J.; Leyva-Pérez, M.d.l.O.; Vidoy-Mercado, I.; Barceló, A.; Luque, F. Transcriptomic time-series analysis of early development in olive from germinated embryos to juvenile tree. *BMC Genom.* **2018**, *19*, 824. [CrossRef]
93. Renny-Byfield, S.; Wendel, J.F. Doubling down on genomes: Polyploidy and crop plants. *Am. J. Bot.* **2014**, *101*, 1711–1725. [CrossRef] [PubMed]
94. Panchy, N.; Lehti-Shiu, M.; Shiu, S.H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [CrossRef] [PubMed]
95. Yang, Z.; Gong, Q.; Qin, W.; Cheng, Y.; Lu, L.; Ge, X.; Zhang, C.; Wu, Z.; Li, F. Genome-wide analysis of WOX genes in upland cotton and their expression pattern under different stresses. *BMC Plant Biol.* **2017**, *17*, 113. [CrossRef] [PubMed]
96. Samadder, P.; Sivamani, E.; Lu, J.; Li, X.; Qu, R. Transcriptional and post-transcriptional enhancement of gene expression by the 5′ UTR intron of rice *rubi3* gene in transgenic rice cells. *Mol. Genet. Genom.* **2008**, *279*, 429–439. [CrossRef] [PubMed]
97. Shaul, O. How introns enhance gene expression. *Int J. Biochemistry Cell Biol.* **2017**, *91*, 145–155. [CrossRef] [PubMed]
98. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256¨CW259. [CrossRef] [PubMed]
99. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297.
100. Sonnhammer, E.L.L.; von Heijne, G.; Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology* **1998**, 175–182.
101. Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **1998**, *14*, 378–379. [CrossRef]
102. Chou, K.C.; Shen, H.B. Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS ONE* **2010**, *5*, e11335. [CrossRef]
103. Timothy, L.; Bailey, J.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* **1994**, 28–36.
104. Goldman, N.; Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **1994**, *11*, 725–736. [PubMed]
105. Ramírez-Tejero, J.A.; Jiménez-Ruiz, J.; Leyva-Pérez, M.d.l.O.; Barroso, J.B.; Luque, F. Gene expression pattern in olive tree organs (Olea europaea L.). *Genes* **2020**, in press.
106. Leyva-Pérez, M.d.l.O.; Jiménez-Ruiz, J.; Gómez-Lama Cabanás, C.; Valverde-Corredor, A.; Barroso, J.B.; Luque, F.; Mercado-Blanco, J. Tolerance of olive (Olea europaea) cv Frantoio to Verticillium dahlia relies on both basal and pathogen-induced differential transcriptomic responses. *New Phytol.* **2018**, *217*, 671–686.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).