

RESEARCH ARTICLE

Open Access



# Genome-wide comparison reveals divergence of cassava and rubber aquaporin family genes after the recent whole-genome duplication

Zhi Zou<sup>1\*</sup> and Jianghua Yang<sup>2</sup>

## Abstract

**Background:** Aquaporins (AQPs) are a class of integral membrane proteins that facilitate the passive transport of water and other small solutes across biological membranes. Despite their importance, little information is available in cassava (*Manihot esculenta*), a perennial shrub of the Euphorbiaceae family that serves the sixth major staple crop in the world.

**Results:** This study presents a genome-wide analysis of the AQP gene family in cassava. The family of 42 members in this species could be divided into five subfamilies based on phylogenetic analysis, i.e., 14 plasma membrane intrinsic proteins (PIPs), 13 tonoplast intrinsic proteins (TIPs), nine NOD26-like intrinsic proteins (NIPs), four X intrinsic proteins (XIPs), and two small basic intrinsic proteins (SIPs). Best-reciprocal-hit-based sequence comparison and synteny analysis revealed 34 orthologous groups (OGs) present in the Euphorbiaceae ancestor, and nearly one-to-one or two-to-one orthologous relationships were observed between cassava with rubber/physic nut, reflecting the occurrence of one so-called  $\rho$  recent whole-genome duplication (WGD) in the last common ancestor of cassava and rubber. In contrast to a predominant role of the  $\rho$  WGD on family expansion in rubber, cassava AQP duplicates were derived from the WGD as well as local duplication. Species-specific gene loss was also observed in cassava, which includes the entire NIP4 group and/or six OGs. Comparison of conserved motifs and gene expression profiles revealed divergence of paralogs in cassava as observed in rubber.

**Conclusions:** Our findings will not only improve our knowledge on family evolution in Euphorbiaceae, but also provide valuable information for further functional analysis of AQP genes in cassava and rubber.

**Keywords:** Aquaporin, AQP gene family, Gene duplication, Expansion, Evolution, Orthologous group, Phylogenetic analysis, Whole-genome duplication

## Background

Cassava (*Manihot esculenta* Crantz,  $2n = 36$ ) is a perennial shrub that belongs to Euphorbiaceae, one of the largest plant families also including rubber (*Hevea brasiliensis* Muell. Arg.,  $2n = 36$ ), castor (*Ricinus communis* L.,  $2n = 20$ ), and physic nut (*Jatropha curcas* L.,  $2n = 22$ ) [1–6]. Cassava was domesticated from its wild progenitor, *M.*

*esculenta* ssp. *flabellifolia*, along the southern border of the Amazon basin [7]. Now, cassava is widely cultivated in tropical regions and represents the sixth major staple crop in the world [8, 9]. Besides servicing as human foods and livestock feeds, the starchy-enriched storage root of cassava is ideal for bio-ethanol production [10]. The cassava genome was estimated to be 772 Mb and three assemblies have been available for three lines: W14 (*M. esculenta* ssp. *flabellifolia*), a wild subspecies with low storage root yield and low root starch content; KU50 (also known as MTAI16), a widely cultivated variety with high storage root yield, high starch content, and vigorous plant growth

\* Correspondence: [zouzhi2008@126.com](mailto:zouzhi2008@126.com); [zouzhi@itbb.org.cn](mailto:zouzhi@itbb.org.cn)

<sup>1</sup>Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, People's Republic of China

Full list of author information is available at the end of the article



with wide adaptability to unfavourable conditions; and, AM560–2, a partial inbred line of MCOL1505 [8, 9]. The most complete one (i.e. AM560–2) spans about 582 Mb and 89.0% of this assembly could be anchored to 18 chromosomes (Chrs) based on 22,403 genetic markers available [9]. In addition to the ancient so-called  $\gamma$  whole-genome triplication shared by all core eudicots, comparative genomics analysis showed that the ancestor of cassava experienced one recent whole-genome duplication (WGD, named  $\rho$  in this study) after its divergence with *Ricinus* and *Jatropha* [5, 6, 11, 12]. This WGD was estimated to occur within a window of 39–47 million years ago (Mya), which is shared by *Hevea* [2, 9, 11, 12]. Despite sharing the same recent WGD, the morphology of rubber, which is characterized as a perennial big tree, is obviously distinct from cassava [5, 6, 11–13]. Moreover, rubber was shown to harbor a considerably bigger genome size, i.e. approximate 2.15 or 1.5 Gb inferred from Feulgen microdensitometry and sequencing-based K-mer analysis, respectively [14, 15]. According to the most complete assembly (i.e. Reyan7–33-97) that spans about 1.37 Gb, the number of protein-coding genes in rubber was shown to be 43,792 which is relatively more than 33,033 in cassava [9, 15], suggesting different fates of duplicated genes after the  $\rho$  WGD. Thereby, it is of particular interest to study the evolutionary fate of duplicated genes in these two special species.

Aquaporins (AQPs), a special class of integral membrane proteins in the ancient major intrinsic protein (MIP) superfamily, are distributed in all types of organisms, including microbes, animals, and plants [1, 16–18]. AQPs are characterized by six transmembrane helices (i.e. TM1–TM6) connected by five loops (i.e. LA–LE), two short helices (i.e. HB and HE), two NPA (Asn-Pro-Ala) motifs, and the ar/R (aromatic/arginine) selectivity filter (i.e. H2, H5, LE1, and LE2) [19, 20]. In addition to water, some AQP family members also transport other small solutes, e.g. glycerol, urea, boric acid, silicic acid, arsenic, ammonia (NH<sub>3</sub>), carbon dioxide (CO<sub>2</sub>), oxygen (O<sub>2</sub>), and hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), where glycerol facilitators were called aquaglyceroporins (GLPs) [16, 19, 21]. Compared with few members present in microbes and animals, the AQP family was shown to have particularly expanded in high plants, which can be divided into five main subfamilies based on sequence similarity: plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), NOD26-like intrinsic proteins (NIPs), small basic intrinsic proteins (SIPs), and X intrinsic proteins (XIPs) [1, 17, 18, 22–24]. The former four subfamilies are widely distributed, whereas XIPs are absent from monocots and the Brassicaceae family in dicots [17, 25, 26]. The fast expansion of this gene family is usually associated with several WGD events, e.g. the  $\gamma$  event for core eudicots and the  $\tau$  event for monocots [27, 28]. Moreover, it is well established that arabidopsis

(*Arabidopsis thaliana*) experienced two additional doubling events, known as  $\beta$  and  $\alpha$ , respectively [29], whereas poplar (*Populus trichocarpa*) experienced one Salicaceae-specific recent WGD [30]. As a result, a high number of AQP gene pairs (i.e. paralogs) were identified in these two species [1, 23, 24]. For example, 35 arabidopsis AQP genes were shown to result from 18 parents, including eight or two AQP genes from  $\alpha$  and  $\beta$  WGDs, respectively [31]. Genome-wide comparison of rubber, castor and, physic nut AQP genes also revealed a high number of duplicates present in rubber, and the pattern is highly similar to poplar [1, 18]. Nevertheless, the origin of rubber AQP duplicates was not well resolved due to the lack of a high-density genetic map [1, 24]. The recently available chromosome-scale structure of the cassava genome allows us to address this issue. In this paper, we report a genome-wide identification and manual curation of AQP family genes in cassava by using available genome and transcriptome datasets. Moreover, we would also like to present a comprehensive comparison of cassava and rubber AQP genes based on analysis of gene structures, sequence characteristics, orthologous relationships, and expression profiles.

## Methods

### Datasets and sequence retrieval

AQP genes reported in rubber, castor, physic nut, and poplar were obtained according to related literatures, and accession numbers can be found in Additional file 1. The cassava genome sequences were downloaded from Phytozome v12 (<https://phytozome.jgi.doe.gov/pz/portal.html>), whereas other data such as nucleotides, Sanger expressed sequence tags (ESTs) and RNA sequencing (RNA-seq) reads were all accessed from NCBI (<https://www.ncbi.nlm.nih.gov/>).

### Identification and manual curation of AQP family genes in cassava

MeAQP proteins available in GenBank were used as queries to search for homologs from the cassava genome. The E-value in the tBLASTn search [32] was set to 1e-5, and positive genomic sequences were predicted as described before [1, 24]. Predicted gene models were further validated with ESTs and RNA-seq reads when available. Homology search for nucleotides or ESTs was performed using BLASTn [32]. RNA-seq data were also adopted for the expression annotation as described before [24], where read alignment was performed using Bowtie 2 [33].

### Syntenic analysis and gene expansion patterns

The all-to-all BLASTP was used to identify homolog pairs as described before [5, 6]. Syntenic blocks and gene collinearity were inferred using MCScanX [34]. WGD duplicates were defined when duplicated genes are located in

syntenic blocks of duplicated chromosomes, while tandem duplications were considered when two duplicated genes were consecutive in a genome. For duplicate pairs, *Ka* (nonsynonymous substitution rate) and *Ks* (synonymous substitution rate) were calculated by *codeml* in the *PAML* package [35].

### Sequence alignment, phylogenetic analysis, and classification

Multiple sequence alignment of full-length AQP proteins was performed using *MUSCLE* [36]. Unrooted trees were constructed using *MEGA 6.0* [37] with the maximum likelihood method, where the bootstrap was set to 1000 replicates. Classification of AQPs into subfamilies and groups was done as previously described [23]. Orthologous groups (OGs) across different species were inferred from BRH (best reciprocal hit)-based sequence comparison as described before [2, 11, 12]. As for cassava and rubber, information from results of above synteny analysis was also considered.

### Structural features of MeAQPs

Protein features such as theoretical molecular weight (MW), isoelectric point (*pI*), and grand average of hydropathicity (GRAVY) were calculated using *ProtParam* (<https://web.expasy.org/protparam/>). Functional prediction was performed based on analysis of dual NPA motifs, ar/R filter, and five Froger's positions (five conserved residues named P1–5 for discriminating GLPs from water-conducting AQPs) from alignments with the structure resolved spinach (*Spinacia oleracea*) PIP2;1 and AtTIP2;1 as well as functionally characterized AQPs [20, 38, 39]. Additionally, conserved motifs in Me/HbAQP proteins were analyzed using *MEME* [40], and optimized parameters were as follows: any number of repetitions; maximum number of motifs, 25; and, the optimum width of each motif, between 6 and 50 residues. The *MAST* program [41] was also used to search detected motifs in protein databases.

### Gene expression analysis

Global gene expression profiles of *MeAQP* genes were investigated over various tissues as described before (GEO accession number GSE82279) [42], i.e. shoot apical meristem (SAM), lateral bud, leaf blade, leaf midvein, petiole, stem, fibrous root, storage root, root apical meristem (RAM), friable embryogenic callus (FEC), and somatic organized embryogenic structure (OES): 101 paired-end reads were generated using *Illumina HiSeq 2500*, and three biological replicates were performed for most tissues except for storage root with two replicates. Raw reads were first filtered by removing adaptor sequences, adaptor-only reads, and low quality reads containing more than 50% bases with *Q*-value  $\leq 5$ . Obtained clean reads were mapped to identified *MeAQPs* and other protein-coding genes

using *Bowtie 2* [33], and the FPKM (fragments per kilobase of exon per million fragments mapped) method [43] was used for determination of transcript levels. Unless specific statements, tools used in this study were performed with default parameters.

## Results

### Characterization of 43AQP-encoding loci in cassava

The search of the cassava genome resulted in 42 AQP-encoding genes (Table 1), corresponding to 43 loci reported by the genome annotation [9]. Among them, *MeSIP2;1* (see Additional file 2), spanning 17,010 bp that was supported by three ESTs and thousands of RNA-seq reads, was annotated as two loci, i.e. Manes.09G074100 and Manes.09G074000. Moreover, based on expert revision of gene structures via aligning ESTs and reads to AQP-coding genome sequences, the gene models of three other loci (i.e. Manes.16G044000, Manes.11G089200, and Manes.11G089100) were also optimized (see Additional files 3, 4 and 5).

Except for *MePIP2;7* (GenBank accession number EU599222), homology search showed that no full-length cDNA sequences of other 41 family genes have been reported in any public database (as of Dec 2017). Nevertheless, 28 members had EST hits in GenBank and *MeTIP1;2* was found to harbor the most of 80 hits (Table 1). Moreover, the expression of other family members was supported by available RNA-seq reads derived from various transcriptomes of somatic embryo, embryogenic callus, embryogenic structure, leaf blade, leaf midvein, petiole, stem, SAM, lateral bud, fibrous root, storage root, and RAM.

These *MeAQPs* were found to locate on 15 out of the 18 chromosomes, only excluding Chromosomes 6, 15, and 18 (Fig. 1). The gene distribution looks uneven: six chromosomes (counting 40.0%) harbor a single AQP gene, whereas Chromosome 11 contains the most of seven genes. As shown in Table 2, the CDS (coding sequences) of 14 gene pairs exhibit a relatively high identity at the nucleotide level, varying from 81.7 to 93.5%. *MeXIP3;1/-3;2* can be defined as tandem duplication for their adjacent organization on the same chromosome. By contrast, other gene pairs are located in syntenic blocks of duplicated chromosomes and thus were considered to result from the  $\rho$  WGD. The *Ka/Ks* ratios of these duplicates are all below one (from 0.0258 to 0.3751) (Table 2), suggesting that their divergence was driven by purifying selection.

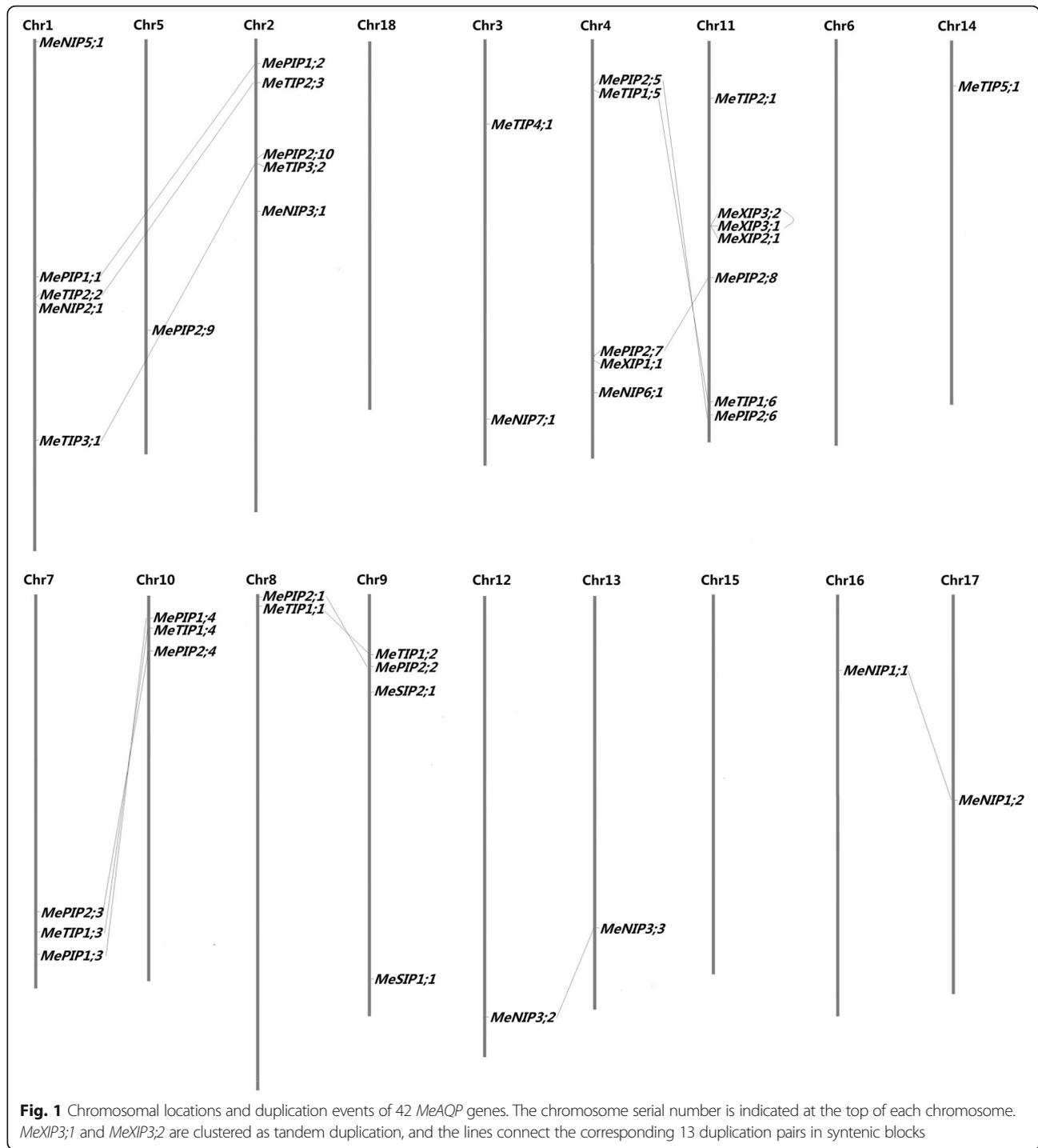
### Phylogenetic analysis and classification

To analyze the evolutionary relationships and infer putative functions, an unrooted phylogenetic tree was constructed from all *MeAQPs* together with 48 HbAQPs, 37 RcAQPs, 31 JcAQPs, and 55 PtAQPs. Poplar, a representative plant of the *Salicaceae* family also belonging to the

**Table 1** Cassava AQP family genes identified in this study

Gene name	Locus ID	Chr location	Nucleotide length (bp, from start to stop codons)		Intron no.	EST no.	Comment
			CDS	Gene			
<i>MePIP1;1</i>	Manes.01G059600	Chr1:17535987–17,539,206	864	2626	3	15	
<i>MePIP1;2</i>	Manes.02G020100	Chr2:1624224–1,626,227	864	1654	3	16	
<i>MePIP1;3</i>	Manes.07G126600	Chr7:25215682–25,217,264	864	1157	3	20	
<i>MePIP1;4</i>	Manes.10G016600	Chr10:1312629–1,314,244	864	1166	3	23	
<i>MePIP2;1</i>	Manes.08G006800	Chr8:512954–514,468	867	1127	3	21	
<i>MePIP2;2</i>	Manes.09G068800	Chr9:9307274–9,308,662	867	1157	3	4	
<i>MePIP2;3</i>	Manes.07G100500	Chr7:22842955–22,845,156	858	1852	3	3	
<i>MePIP2;4</i>	Manes.10G046200	Chr10:4375805–4,379,849	861	3655	3	5	
<i>MePIP2;5</i>	Manes.04G021200	Chr4:2304659–2,306,481	861	1396	3	2	
<i>MePIP2;6</i>	Manes.11G145300	Chr11:25677637–25,679,377	861	1300	3	10	
<i>MePIP2;7</i>	Manes.04G076500	Chr4:21375880–21,377,507	843	1265	3	21	
<i>MePIP2;8</i>	Manes.11G096600	Chr11:15805567–15,807,131	843	1236	3	2	
<i>MePIP2;9</i>	Manes.02G109200	Chr2:8173111–8,174,880	852	1330	3	17	
<i>MePIP2;10</i>	Manes.05G144100	Chr5:20526797–20,528,273	846	1122	3	3	
<i>MeTIP1;1</i>	Manes.08G012800	Chr8:924654–925,957	759	855	1	47	
<i>MeTIP1;2</i>	Manes.09G062300	Chr9:8402765–8,403,903	759	853	1	80	
<i>MeTIP1;3</i>	Manes.07G111500	Chr7:23996724–23,997,990	759	957	2	0	
<i>MeTIP1;4</i>	Manes.10G035000	Chr10:2983924–2,985,053	759	848	1	2	
<i>MeTIP1;5</i>	Manes.04G030400	Chr4:3360155–3,361,296	759	962	2	0	
<i>MeTIP1;6</i>	Manes.11G134600	Chr11:24689956–24,691,006	759	977	2	1	
<i>MeTIP2;1</i>	Manes.11G036500	Chr11:3094579–3,095,833	747	954	2	17	
<i>MeTIP2;2</i>	Manes.01G081600	Chr1:20741570–20,742,776	753	928	2	1	
<i>MeTIP2;3</i>	Manes.02G040800	Chr2:3161794–3,163,035	753	969	2	7	
<i>MeTIP3;1</i>	Manes.01G160000	Chr1:26663550–26,664,500	777	951	2	0	
<i>MeTIP3;2</i>	Manes.02G118300	Chr2:8751200–8,752,174	777	975	2	0	
<i>MeTIP4;1</i>	Manes.03G062300	Chr3:6536427–6,538,114	744	1458	2	6	
<i>MeTIP5;1</i>	Manes.14G036400	Chr14:2919392–2,920,528	759	1137	2	0	
<i>MeNIP1;1</i>	Manes.16G044000	Chr16:6177753–6,179,082	840	1600	4	0	Misannotated
<i>MeNIP1;2</i>	Manes.17G061100	Chr17:19971252–19,973,361	858	1712	4	1	
<i>MeNIP2;1</i>	Manes.01G091200	Chr1:21577225–21,580,168	867	2458	4	2	
<i>MeNIP3;1</i>	Manes.02G152300	Chr2:11346718–11,348,621	852	1213	4	0	
<i>MeNIP3;2</i>	Manes.12G133900	Chr12:28849870–28,851,480	822	1102	4	0	
<i>MeNIP3;3</i>	Manes.13G093600	Chr13:19561358–19,562,924	825	1049	4	0	
<i>MeNIP5;1</i>	Manes.01G001400	Chr1:259425–262,398	897	2448	3	4	
<i>MeNIP6;1</i>	Manes.04G104100	Chr4:24004959–24,008,297	921	2638	4	0	
<i>MeNIP7;1</i>	Manes.03G183400	Chr3:27015982–27,017,206	900	1225	4	0	
<i>MeXIP1;1</i>	Manes.04G078900	Chr4:21735846–21,737,169	885	1234	1	0	
<i>MeXIP2;1</i>	Manes.11G089300	Chr11:12980304–12,982,216	915	1615	2	2	
<i>MeXIP3;1</i>	Manes.11G089200	Chr11:12950526–12,952,826	915	2301	2	0	Misannotated
<i>MeXIP3;2</i>	Manes.11G089100	Chr11:12935690–12,936,439	924	4465	2	0	Misannotated
<i>MeSIP1;1</i>	Manes.09G144400	Chr9:26227923–26,233,566	720	5135	2	5	
<i>MeSIP2;1</i>	Manes.09G074000	Chr9:10639757–10,640,714	714	17,010	2	3	Misannotated

(bp base pair, CDS coding sequence, Chr chromosome, EST expressed sequence tag)



order Malpighiales as Euphorbiaceae, was used as an out-group of Euphorbiaceous plants. According to the tree, 42 *MeAQPs* were grouped into five subfamilies, i.e. PIP (14), TIP (13), NIP (9), SIP (2), and XIP (4). The PIP subfamily can be further divided into two phylogenetic groups (i.e. four *MePIP1s* and ten *MePIP2s*), the TIP subfamily into five groups (i.e. six *MeTIP1s*, three *MeTIP2s*, two *MeTIP3s*, one *MeTIP4*, and one *MeTIP5*), the NIP

subfamily into six groups (two *MeNIP1s*, one *MeNIP2*, three *MeNIP3s*, one *MeNIP5*, one *MeNIP6*, and one *MeNIP7*), the SIP subfamily into two groups (one *MeSIP1* and one *MeSIP2*), and the XIP subfamily into three groups (one *MeXIP1*, one *MeXIP2*, and two *MeXIP3s*) (Fig. 2). Interestingly, the widely distributed NIP4 group was not found in cassava, though genome sequences of W14 and KU50 and various transcriptome data were also mined.

**Table 2** Cassava AQP duplicates identified in this study

Duplicated gene pair	Identity (%)	Ks	Ka/Ks
<i>MePIP1;1/MePIP1;2</i>	91.0	0.3169	0.0807
<i>MePIP1;3/MePIP1;4</i>	93.5	0.2442	0.0546
<i>MePIP2;1/MePIP2;2</i>	91.0	0.3260	0.0876
<i>MePIP2;3/MePIP2;4</i>	88.2	0.4224	0.1089
<i>MePIP2;5/MePIP2;6</i>	89.3	0.5591	0.0258
<i>MePIP2;7/MePIP2;8</i>	91.9	0.3106	0.0703
<i>MeTIP1;1/MeTIP1;2</i>	89.7	0.4090	0.0699
<i>MeTIP1;3/MeTIP1;4</i>	90.3	0.2952	0.1555
<i>MeTIP1;5/MeTIP1;6</i>	86.7	0.5756	0.0646
<i>MeTIP2;2/MeTIP2;3</i>	88.6	0.4055	0.0898
<i>MeTIP3;1/MeTIP3;2</i>	90.5	0.3015	0.1214
<i>MeNIP1;1/MeNIP1;2</i>	83.3	0.3958	0.2888
<i>MeNIP3;2/MeNIP3;3</i>	81.7	0.4731	0.2702
<i>MeXIP3;1/MeXIP3;2</i>	89.8	0.1877	0.3751

Ks and Ka were calculated using PAML. (Ka nonsynonymous substitution rate, Ks synonymous substitution rate)

The BRH-based sequence comparison as well as synteny analysis were also adopted to identify orthologous groups across cassava, rubber, castor, physic nut, and poplar. As shown in Table 3, a total of 34 OGs were identified and each phylogenetic group was shown to contain one to six OGs. It is worth noting that, species-specific gene expansion or loss was obviously observed, where only 28 OGs have retained in cassava (Table 3).

#### Analysis of exon-intron structure

The exon-intron structures of *MeAQP* genes were analyzed based on revised gene models. Compared with the ORF (open reading frame), the gene length (from start to stop codons) is considerably more variable, i.e. 848–17,010 bp vs 714–924 bp. The intron number of *MeAQP* genes varies from one to four, and the majority of them (accounting for 76.2%) contain two or three introns. The average intron length is about 423 bp, with the minimum of 71 bp for the second intron of *MeNIP3;2* and the maximum of 16,179 bp for the first intron of *MeSIP2;1*. The exon-intron structure is usually highly conserved in the same subfamily but distinct between different subfamilies: the PIP subfamily features three introns; the TIP subfamily features two introns except for three members (i.e. *MeTIP1;1*, *MeTIP1;2*, and *MeTIP1;4*) that harbor a single one; the NIP subfamily features four introns except for *MeNIP5;1* that harbors three introns; the SIP subfamily features two introns; and, the XIP subfamily features one (XIP1 group) or two (XIP2 and XIP3 groups) (Table 1).

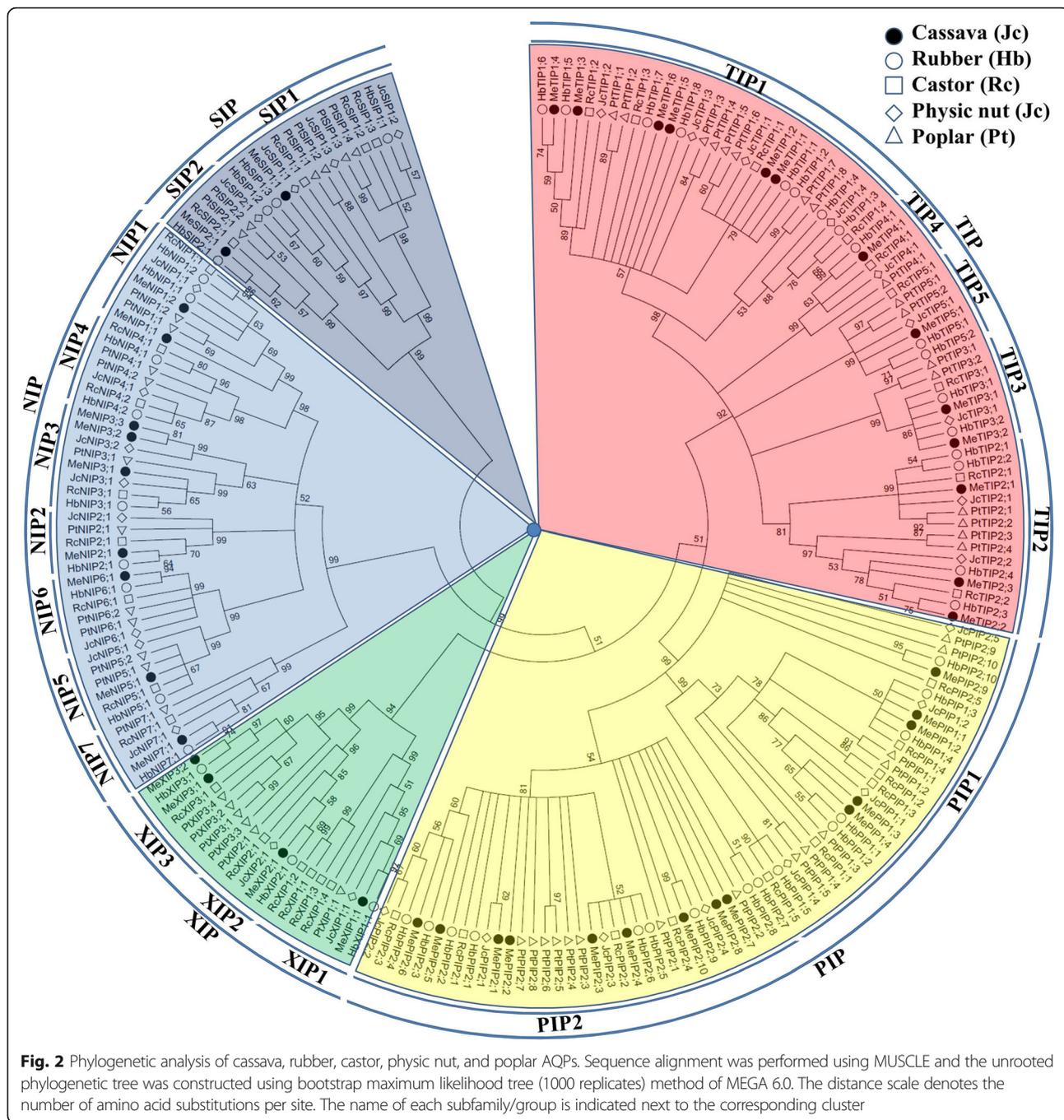
#### Structural features of MeAQPs

Sequence analysis showed that 42 MeAQPs consist of 237–307 amino acids (AA), with a theoretical molecular

weight of 25.14–32.64 kDa and a *pI* value of 4.79–9.67 (Table 4). The GRAVY value was all shown to be more than 0 (varying from 0.330 to 0.925), indicating their hydrophobic feature. In fact, multiple alignments showed that all MeAQPs harbor six TMs (Additional file 6). The average *pI* value is about 8.35, 5.88, 7.98, 7.38, or 9.42 for subfamilies PIP, TIP, NIP, XIP, and SIP, respectively (Table 4).

Conserved residues, typical of dual NPA motifs, ar/R filter, and five Froger's positions, were also identified as shown in Table 4. Two NPA motifs are usually conserved, though several variants were also observed: NPS and NPV for two NPA motifs of *MeNIP5;1* and *MeNIP6;1*; NPT or NPL for the first NPA motif of *MeSIP1;1* and *MeSIP2;1* respectively; and, SPV, NPV or NPL for the first NPA motif of *MeXIP1;1*, *MeXIP2;1*, and *MeXIP3;1/–3;2*, respectively (Table 4). Most MeAQPs exhibit AqpZ-like Froger's residues that favor the permeability of water [38]. By contrast, NIP subfamily members as well as *MeSIP2;1* feature mixed key residues of GlpF for P1/P5 and AqpZ for P2–P4 (Table 4). Actually, the glycerol permeability of NIPs has been well established [44, 45]. Interestingly, *NtAQP1*, a PIP1 group member from *Nicotiana tabacum*, was also shown to transport glycerol [46]. All MePIPs represent the F-H-T-R ar/R filter as observed in the pure water channel AqpZ, suggesting their putatively water permeability (Table 4). The high water permeability of plant PIP2s has widely described, however, PIP1s exhibit no or extremely low water permeability when expressed in *Xenopus laevis* oocytes [47–52]. Moreover, PIPs were also shown to transport urea, boric acid, CO<sub>2</sub>, and H<sub>2</sub>O<sub>2</sub> [45, 53–57]. In addition to water, TIPs were also proven to transport glycerol, urea, boric acid, NH<sub>3</sub>, and H<sub>2</sub>O<sub>2</sub> [45, 54, 55, 58–60]. NIPs have reported to transport water, glycerol, urea, boric acid, silicic acid, NH<sub>3</sub>, and H<sub>2</sub>O<sub>2</sub> [54, 55, 61–63], whereas XIPs have reported to transport water, glycerol, urea, boric acid, and H<sub>2</sub>O<sub>2</sub> [64–67].

To learn more about the diversity of motif compositions among different MeAQPs as well as HbAQPs, an unrooted phylogenetic tree was constructed and conserved motifs were predicted using MEME. Among 25 motifs identified, Motifs 1, 2, 3, 4, and 5 are widely found in PIP and TIP subfamilies; Motif 6 is widely present in TIP and NIP subfamilies and several members of the PIP subfamily; Motif 9 is widely present in TIP, NIP, and XIP subfamilies; Motif 10 is widely present in NIP, XIP, and SIP subfamilies and several members of the TIP subfamily; Motif 15 is widely present in NIP and XIP subfamilies and several members of the TIP subfamily; Motif 16 is widely present in PIP, XIP, and SIP subfamilies; Motif 20 is widely present in NIP and SIP subfamilies; Motif 21 is widely present in XIP and SIP subfamilies and several members of the NIP subfamily; Motifs 7, 8, 12, and 22 are widely found in the PIP subfamily; Motifs 11, 14, and 19 are widely found in the TIP



**Fig. 2** Phylogenetic analysis of cassava, rubber, castor, physic nut, and poplar AQPs. Sequence alignment was performed using MUSCLE and the unrooted phylogenetic tree was constructed using bootstrap maximum likelihood tree (1000 replicates) method of MEGA 6.0. The distance scale denotes the number of amino acid substitutions per site. The name of each subfamily/group is indicated next to the corresponding cluster

subfamily; Motif 23 is limited to the XIP subfamily; Motifs 13 and 17 are limited to the PIP2 or PIP1 group respectively; Motif 24 is limited to the TIP1 group; Motif 18 is present in several members of TIP and NIP subfamilies; and, Motif 25 is only present in several members of the TIP subfamily (see Fig. 3 and Additional file 7). Among them, Motif 1 spans TM2, HB, and TM3; Motif 23 spans TM2 and HB; Motifs 3 and 23 span TM1; Motifs 12 and 14 span TM2; Motifs 8, 11, and 21 span TM3; Motifs 5 and 9 span TM4; Motifs 2, 15, and 25 span TM5; Motifs 6,

16, and 18 span HE; Motifs 4, 10, and 20 span TM6; and, Motif 13 includes a putative phosphorylation site corresponding to S274 in SoPIP2;1 [20].

Gain or loss of certain motifs was observed within a high number of orthologous groups. Motif 17, which is PIP1-specific, is absent from HbPIP1;5. The PIP2 group usually features 11 motifs, i.e. Motifs 7, 3, 12, 1, 8, 22, 5, 2, 18, 4, and 13, however, Motif 1 is absent from MePIP2;2, MePIP2;5, MePIP2;9, and HbPIP2;9, whereas Motif 3 is absent from MePIP2;9 and HbPIP2;10. Moreover, Motifs 1

**Table 3** 34 identified OGs based on comparison of five examined species

Group	OG	Cassava	Rubber	Castor	Physic nut	Poplar
PIP1	1-1a	<i>MePIP1;1</i>	<i>HbPIP1;4</i>	<i>RcPIP1;4</i>	<i>JcPIP1;2</i>	<i>PtPIP1;3</i>
		<b><i>MePIP1;2</i></b>	<b><i>HbPIP1;3</i></b>			
PIP1	1-1b	<i>MePIP1;3</i>	<i>HbPIP1;2</i>	<i>RcPIP1;2</i>	<i>JcPIP1;1</i>	<i>PtPIP1;1</i>
		<b><i>MePIP1;4</i></b>	<b><i>HbPIP1;1</i></b>	<i>RcPIP1;3</i>		<i>PtPIP1;2</i>
PIP1	1-1c	–	–	<i>RcPIP1;1</i>	–	<i>PtPIP1;4</i> <i>PtPIP1;5</i>
PIP1	1-1d	–	<i>HbPIP1;5</i>	<i>RcPIP1;5</i>	<i>JcPIP1;4</i>	–
PIP2	1-2a	<i>MePIP2;1</i>	<i>HbPIP2;1</i>	<i>RcPIP2;1</i>	<i>JcPIP2;1</i>	<i>PtPIP2;5</i>
		<b><i>MePIP2;2</i></b>	<b><i>HbPIP2;2</i></b>			<i>PtPIP2;6</i> <i>PtPIP2;7</i>
PIP2	1-2b	<i>MePIP2;3</i>	<i>HbPIP2;5</i>	<i>RcPIP2;2</i>	<i>JcPIP2;2</i>	<i>PtPIP2;3</i>
		<b><i>MePIP2;4</i></b>	<b><i>HbPIP2;6</i></b>			<i>PtPIP2;4</i>
PIP2	1-2c	<i>MePIP2;5</i>	<i>HbPIP2;4</i>	<i>RcPIP2;3</i>	<i>JcPIP2;3</i>	<i>PtPIP2;8</i>
		<b><i>MePIP2;6</i></b>	<b><i>HbPIP2;3</i></b>			
PIP2	1-2d	<i>MePIP2;7</i>	<i>HbPIP2;7</i>	<i>RcPIP2;4</i>	<i>JcPIP2;4</i>	<i>PtPIP2;1</i>
		<b><i>MePIP2;8</i></b>	<b><i>HbPIP2;8</i></b>			<i>PtPIP2;2</i>
PIP2	1-2e	<i>MePIP2;9</i>	<i>HbPIP2;9</i>	–	–	–
PIP2	1-2f	<i>MePIP2;10</i>	<i>HbPIP2;10</i>	<i>RcPIP2;5</i>	<i>JcPIP2;5</i>	<i>PtPIP2;9</i> <i>PtPIP2;10</i>
TIP1	2-1a	<i>MeTIP1;1</i>	<i>HbTIP1;2</i>	<i>RcTIP1;1</i>	<i>JcTIP1;1</i>	<i>PtTIP1;5</i>
		<b><i>MeTIP1;2</i></b>	<b><i>HbTIP1;1</i></b>			<i>PtTIP1;6</i>
TIP1	2-1b	–	<i>HbTIP1;3</i> <i>HbTIP1;4</i>	<i>RcTIP1;4</i>	<i>JcTIP1;4</i>	<i>PtTIP1;7</i> <i>PtTIP1;8</i>
TIP1	2-1c	<i>MeTIP1;3</i>	<i>HbTIP1;6</i>	<i>RcTIP1;2</i>	<i>JcTIP1;2</i>	<i>PtTIP1;4</i>
		<b><i>MeTIP1;4</i></b>	<b><i>HbTIP1;5</i></b>			<i>PtTIP1;3</i>
TIP1	2-1d	<i>MeTIP1;5</i>	<i>HbTIP1;8</i>	<i>RcTIP1;3</i>	<i>JcTIP1;3</i>	<i>PtTIP1;1</i>
		<b><i>MeTIP1;6</i></b>	<b><i>HbTIP1;7</i></b>			<i>PtTIP1;2</i>
TIP2	2-2a	<i>MeTIP2;1</i>	<i>HbTIP2;1</i> <i>HbTIP2;2</i>	<i>RcTIP2;1</i>	<i>JcTIP2;1</i>	<i>PtTIP2;1</i> <i>PtTIP2;2</i>
TIP2	2-2b	<i>MeTIP2;2</i>	<i>HbTIP2;3</i>	<i>RcTIP2;2</i>	<i>JcTIP2;2</i>	<i>PtTIP2;3</i>
		<b><i>MeTIP2;3</i></b>	<b><i>HbTIP2;4</i></b>			<i>PtTIP2;4</i>
TIP3	2-3	<i>MeTIP3;1</i>	<i>HbTIP3;1</i>	<i>RcTIP3;1</i>	<i>JcTIP3;1</i>	<i>PtTIP3;2</i>
		<b><i>MeTIP3;2</i></b>	<b><i>HbTIP3;2</i></b>			<i>PtTIP3;1</i>
TIP4	2-4	<i>MeTIP4;1</i>	<i>HbTIP4;1</i>	<i>RcTIP4;1</i>	<i>JcTIP4;1</i>	<i>PtTIP4;1</i>
TIP5	2-5	<i>MeTIP5;1</i>	<i>HbTIP5;1</i> <i>HbTIP5;2</i>	<i>RcTIP5;1</i>	<i>JcTIP5;1</i>	<i>PtTIP5;1</i> <i>PtTIP5;2</i>
NIP1	3-1	<i>MeNIP1;1</i>	<i>HbNIP1;1</i>	<i>RcNIP1;1</i>	<i>JcNIP1;1</i>	<i>PtNIP1;1</i>
		<b><i>MeNIP1;2</i></b>	<b><i>HbNIP1;2</i></b>			<i>PtNIP1;2</i>
NIP2	3-2	<i>MeNIP2;1</i>	<i>HbNIP2;1</i>	<i>RcNIP2;1</i>	<i>JcNIP2;1</i>	<i>PtNIP2;1</i>
NIP3	3-3a	<i>MeNIP3;1</i>	<i>HbNIP3;1</i>	<i>RcNIP3;1</i>	<i>JcNIP3;1</i>	<i>PtNIP3;1</i>
NIP3	3-3b	<i>MeNIP3;2</i> <i>MeNIP3;3</i>	–	–	<i>JcNIP3;2</i>	–
NIP4	3-4a	–	<i>HbNIP4;1</i>	<i>RcNIP4;1</i>	–	<i>PtNIP4;1</i>
NIP4	3-4b	–	<i>HbNIP4;2</i>	<i>RcNIP4;2</i>	<i>JcNIP4;1</i>	<i>PtNIP4;2</i>
NIP5	3-5	<i>MeNIP5;1</i>	<i>HbNIP5;1</i>	<i>RcNIP5;1</i>	<i>JcNIP5;1</i>	<i>PtNIP5;1</i> <i>PtNIP5;2</i>

**Table 3** 34 identified OGs based on comparison of five examined species (Continued)

Group	OG	Cassava	Rubber	Castor	Physic nut	Poplar
NIP6	3-6	<i>MeNIP6;1</i>	<i>HbNIP6;1</i>	<i>RcNIP6;1</i>	<i>JcNIP6;1</i>	<i>PtNIP6;1</i> <i>PtNIP6;2</i>
NIP7	3-7	<i>MeNIP7;1</i>	<i>HbNIP7;1</i>	<i>RcNIP7;1</i>	<i>JcNIP7;1</i>	<i>PtNIP7;1</i>
XIP1	4-1	<i>MeXIP1;1</i>	<i>HbXIP1;1</i>	<i>RcXIP1;1</i>	<i>JcXIP1;1</i>	<i>PtXIP1;1</i>
				<i>RcXIP1;2</i> <i>RcXIP1;3</i> <i>RcXIP1;4</i>		
XIP2	4-2	<i>MeXIP2;1</i>	<i>HbXIP2;1</i>	<i>RcXIP2;1</i>	<i>JcXIP2;1</i>	<i>PtXIP2;1</i>
XIP3	4-3	<i>MeXIP3;1</i>	<i>HbXIP3;1</i>	<i>RcXIP3;1</i>	–	<i>PtXIP3;1</i>
				<i>MeXIP3;2</i>		<i>PtXIP3;2</i> <i>PtXIP3;3</i>
SIP1	5-1a	<i>MeSIP1;1</i>	<i>HbSIP1;2</i> <i>HbSIP1;3</i>	<i>RcSIP1;1</i>	<i>JcSIP1;1</i>	<i>PtSIP1;1</i> <i>PtSIP1;2</i>
				<i>RcSIP1;2</i> <i>RcSIP1;3</i>	<i>JcSIP1;2</i> <i>JcSIP1;3</i>	<i>PtSIP1;3</i> <i>PtSIP1;4</i>
SIP1	5-1b	–	<i>HbSIP1;1</i>	<i>RcSIP1;2</i> <i>RcSIP1;3</i>	<i>JcSIP1;2</i> <i>JcSIP1;3</i>	<i>PtSIP1;3</i> <i>PtSIP1;4</i>
SIP2	5-2	<i>MeSIP2;1</i>	<i>HbSIP2;1</i>	<i>RcSIP2;1</i>	<i>JcSIP2;1</i>	<i>PtSIP2;1</i> <i>PtSIP2;2</i>

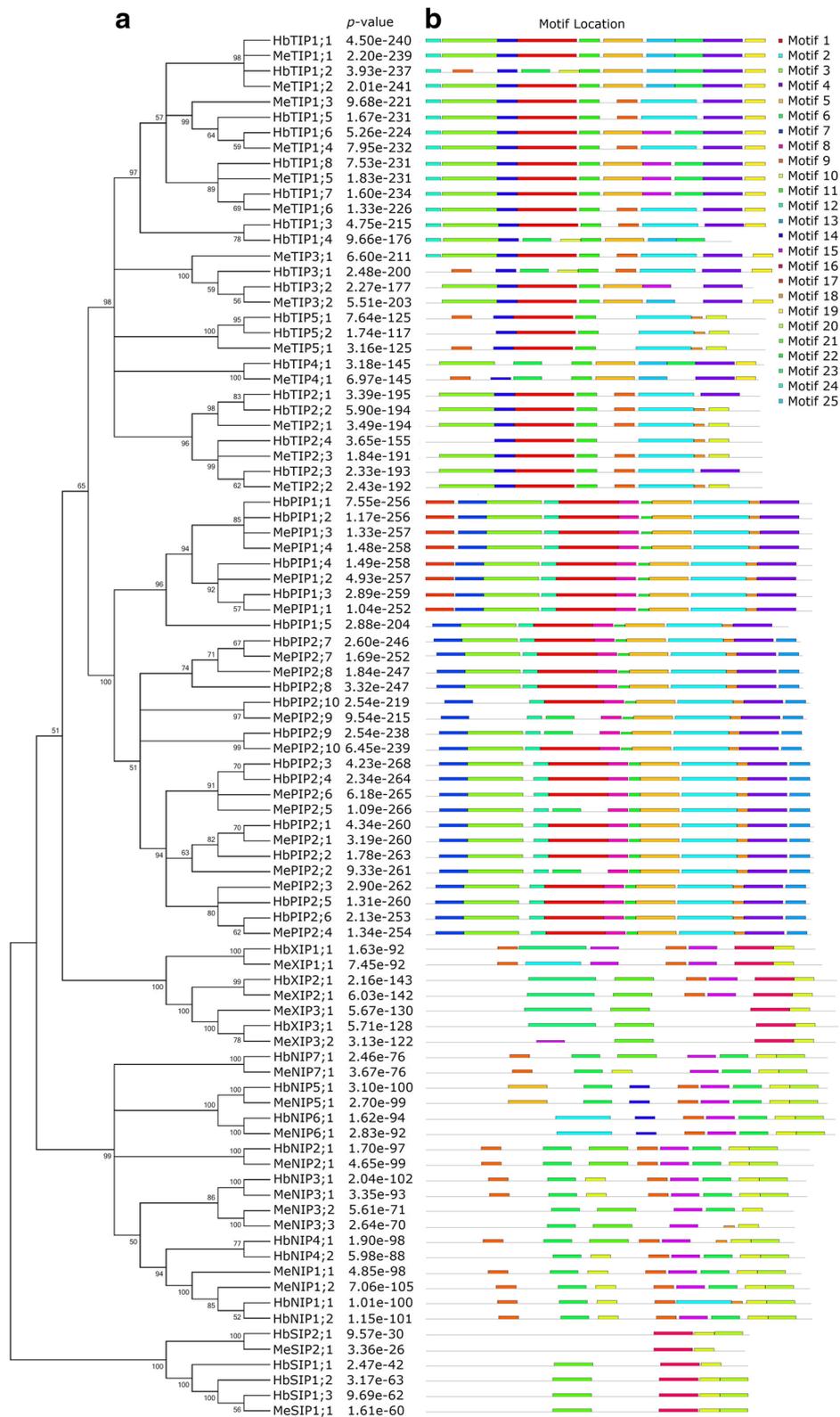
OGs that are limited to rubber and cassava were shown in black. (OG orthologous group)

and 3, which is widely distributed in OG2-1a, are placed by Motifs 6 and 10, or Motif 9 in *HbTIP1;2*, respectively. Motifs 2 and 9 found in OG2-1c are placed by Motifs 6 and 15, or Motif 5 in *HbTIP1;6*, respectively. Motifs 5 and 6/15 found in OG2-1d are placed by Motif 9, or Motif 2 in *MeTIP1;6*, respectively. In OG2-1b, the motif compositions of *HbTIP1;3* is similar to those in OG2-1c, in contrast, Motifs 4 and 9 are absent from *HbTIP1;4*; Motif 1 is placed by Motifs 6 and 10; Motif 9 is placed by Motif 5; and, Motif 2 is placed by Motifs 6 and 25. Motif 24 found in the TIP1 group is also present in *MeTIP3;1* but absent from other TIP3 members. Moreover, Motifs 1 and 3 are placed by Motifs 6 and 10, or Motif 9, respectively. Compared with *MeTIP3;2*, Motif 19 is absent from *HbTIP3;2*, and Motif 25 is placed by Motif 15. Motif 9 found in OG2-5 is absent from *HbTIP5;2*. Compared with *HbTIP4;1*, the Motif 6 behind Motif 25 is absent from *MeTIP4;1*, and Motif 3 is placed by Motifs 9 and 14. Motifs 10 and 18 found in OG2-2a are placed by Motif 4 in *HbTIP2;1*. Motifs 3 and 9 found in the TIP2 group are absent from *HbTIP2;4*, whereas Motifs 10 and 18 are placed by Motif 4 in *HbTIP2;3*. Compared with *HbNIP7;1*, Motif 21 is placed by Motif 10 in *MeNIP7;1*. Motifs 9 and 20 found in OG3-3a of the NIP3 group are absent from OG3-3b. Moreover, Motif 6 found in this group is placed by Motif 18 in *MeNIP3;3*. Compared with *HbNIP4;1*, Motif 9 at N-terminus is absent from *HbNIP4;2*, and Motifs 18 and 21 are placed by Motifs 6 and 10, respectively. Motifs 15 and 6 found in OG3-1 are placed by Motifs 2 and 18, respectively. Motif 23 found in the XIP subfamily is placed by Motif 15 in *MeXIP3;2*. Motif 20 found in the SIP subfamily

**Table 4** Structural features of MeAQPs

Name	AA	MW (kDa)	pI	GRAVY	Ar/R selectivity filter				NPA motifs		Froger's Position				
					H2	H5	LE1	LE2	Loop B	Loop E	P1	P2	P3	P4	P5
MePIP1;1	287	30.62	8.84	0.409	F	H	T	R	NPA	NPA	E	S	A	F	W
MePIP1;2	287	30.74	9.00	0.404	F	H	T	R	NPA	NPA	E	S	A	F	W
MePIP1;3	287	30.74	8.59	0.354	F	H	T	R	NPA	NPA	E	S	A	F	W
MePIP1;4	287	30.79	8.59	0.330	F	H	T	R	NPA	NPA	E	S	A	F	W
MePIP2;1	288	30.59	8.20	0.461	F	H	T	R	NPA	NPA	Q	S	A	F	W
MePIP2;2	288	30.71	6.99	0.438	F	H	T	R	NPA	NPA	Q	S	A	F	W
MePIP2;3	285	30.39	8.89	0.441	F	H	T	R	NPA	NPA	Q	S	A	F	W
MePIP2;4	286	30.42	9.08	0.480	F	H	T	R	NPA	NPA	Q	S	A	F	W
MePIP2;5	286	30.58	8.50	0.449	F	H	T	R	NPA	NPA	Q	S	A	F	W
MePIP2;6	286	30.55	6.99	0.451	F	H	T	R	NPA	NPA	Q	S	A	F	W
MePIP2;7	280	29.76	8.97	0.528	F	H	T	R	NPA	NPA	M	S	A	F	W
MePIP2;8	280	29.53	9.13	0.535	F	H	T	R	NPA	NPA	M	S	A	F	W
MePIP2;9	283	30.04	8.67	0.507	F	H	T	R	NPA	NPA	M	S	A	F	W
MePIP2;10	281	29.93	6.51	0.484	F	H	T	R	NPA	NPA	M	S	A	F	W
MeTIP1;1	252	25.97	5.55	0.730	H	I	A	V	NPA	NPA	T	S	A	Y	W
MeTIP1;2	252	25.97	6.12	0.742	H	I	A	V	NPA	NPA	T	S	A	Y	W
MeTIP1;3	252	25.85	5.19	0.825	H	I	A	V	NPA	NPA	T	S	A	Y	W
MeTIP1;4	252	25.83	5.13	0.801	H	I	A	V	NPA	NPA	T	S	A	Y	W
MeTIP1;5	252	25.69	4.79	0.846	H	I	A	V	NPA	NPA	T	S	A	Y	W
MeTIP1;6	252	25.66	4.94	0.893	H	I	A	V	NPA	NPA	T	S	A	Y	W
MeTIP2;1	248	25.14	6.15	0.925	H	I	G	R	NPA	NPA	T	S	A	Y	W
MeTIP2;2	250	25.41	5.09	0.873	H	I	G	R	NPA	NPA	T	S	A	Y	W
MeTIP2;3	250	25.26	5.66	0.916	H	I	G	R	NPA	NPA	T	S	A	Y	W
MeTIP3;1	258	27.45	7.14	0.567	H	I	A	R	NPA	NPA	T	A	A	Y	W
MeTIP3;2	258	27.32	6.75	0.640	H	I	A	R	NPA	NPA	T	A	A	Y	W
MeTIP4;1	247	25.79	6.12	0.835	H	I	A	R	NPA	NPA	T	A	A	Y	W
MeTIP5;1	252	25.92	7.79	0.749	N	V	G	C	NPA	NPA	T	A	A	Y	W
MeNIP1;1	279	29.94	8.87	0.427	W	V	A	R	NPA	NPA	F	S	A	Y	I
MeNIP1;2	285	30.52	8.90	0.448	W	V	A	R	NPA	NPA	F	S	A	Y	I
MeNIP2;1	288	30.40	9.37	0.430	G	S	G	R	NPA	NPA	L	T	A	Y	L
MeNIP3;1	283	30.46	8.43	0.478	W	A	A	R	NPA	NPA	F	S	A	Y	I
MeNIP3;2	273	28.87	4.96	0.639	W	M	A	R	NPA	NPA	F	S	A	Y	V
MeNIP3;3	274	29.31	6.51	0.659	W	I	A	R	NPA	NPA	L	S	A	Y	I
MeNIP5;1	298	31.03	8.64	0.388	A	I	G	R	NPS	NPV	F	T	A	Y	L
MeNIP6;1	306	31.59	7.71	0.344	T	I	A	R	NPS	NPV	F	T	A	Y	L
MeNIP7;1	299	31.86	8.45	0.551	A	V	G	R	NPA	NPA	Y	S	A	Y	I
MeXIP1;1	294	31.83	6.29	0.707	V	I	V	R	SPV	NPA	M	C	A	F	W
MeXIP2;1	304	32.04	8.60	0.634	I	T	V	R	NPV	NPA	V	C	A	F	W
MeXIP3;1	304	32.30	8.25	0.741	V	T	A	R	NPL	NPA	V	C	A	F	W
MeXIP3;2	307	32.64	6.37	0.788	V	T	A	R	NPL	NPA	V	C	A	F	W
MeSIP1;1	239	26.07	9.17	0.819	V	V	P	N	NPT	NPA	I	A	A	Y	W
MeSIP2;1	237	26.10	9.67	0.479	S	Q	G	S	NPL	NPA	F	V	A	Y	W

(AA amino acid, ar/R aromatic/arginine, GRAVY grand average of hydropathicity, pI isoelectric point, kDa kilodalton, MW molecular weight, NPA Asn-Pro-Ala)



**Fig. 3** Structural and phylogenetic analyses of cassava and rubber AQPs. **a** Shown is the unrooted phylogenetic tree resulting from full-length AQPs with MEGA 6.0. **b** Shown is the distribution of conserved motifs among AQPs, where different motifs are represented by different color blocks as indicated at the bottom of the figure and the same color block in different proteins indicates a certain motif

is absent from *HbSIP1;1* and *MeSIP2;1*, which belong to the SIP1 or SIP2 group, respectively (Fig. 3).

#### Tissue-specific transcriptional profiling of *MeAQP* genes

To reveal the expression evolution of *MeAQP* genes, their expression profiles were investigated based on Illumina RNA-seq data representing 11 tissue types, i.e. SAM, lateral bud, leaf blade, leaf midvein, petiole, stem, fibrous root, storage root, RAM, FEC, and OES. Except for *MeXIP3;2*, the expression of other *MeAQP* genes was all detected in at least one of the examined tissues, though the transcript level is diverse. Based on the FPKM value, the transcript of the total gene family was shown to be most abundant in storage root (defined as Class I); moderate in fibrous root, petiole, stem, RAM, and SAM (Class II, accounting for 38.4–66.0% of Class I); and, relatively low in leaf midvein, lateral bud, leaf blade, OES, and FEC (Class III, accounting for 5.8–28.9% of Class I). Subfamilies PIP and TIP contribute the major transcripts in most examined tissues, varying from 83.2% in leaf midvein to 99.1% in storage root. However, the transcript level of the SIP subfamily is comparative to that of the PIP subfamily in FEC, and the XIP subfamily contributes more than the TIP subfamily in leaf blade. Several key members were identified in a certain tissue: *MeTIP1;2* represents the most expressed gene in storage root, lateral bud, and SAM; *MePIPI;2* represents the most expressed gene in leaf midvein, and the second most expressed gene in fibrous root, petiole, lateral bud, and OES; *MePIP2;4* represents the most expressed gene in fibrous root; *MeTIP1;1* represents the most expressed gene in petiole, RAM, and OES, and the second most expressed gene in stem, leaf midvein, and FEC; *MeTIP2;1*, *MeXIP2;1* or *MeTIP3;1* represents the most expressed gene in stem, leaf blade and FEC, respectively. According to their expression patterns over various tissues, 41 *MeAQP* genes were grouped into seven main clusters: Clusters I, II, III, IV and VI are predominantly expressed in FEC, lateral bud, leaf blade, RAM or storage root, respectively; Cluster V is preferentially expressed in petiole and stem, including *MeNIP3;3*, *MePIP2;7*, *MeTIP2;1*, *MePIPI;1*, *MePIP2;8*, *MeNIP5;1*, *MeNIP6;1*, and *MeNIP7;1*, where the latter five were also highly expressed in leaf midvein; and, Cluster VII is typically expressed in fibrous root, including *MePIPI;3*, *MePIP2;1*, *MeTIP1;4*, *MePIP2;4*, *MeTIP1;5*, *MeTIP2;2*, *MeTIP1;3*, *MeTIP2;3*, *MeSIP1;1*, *MePIPI;4*, and *MePIP2;2*, where the latter three were also highly expressed in storage root (Fig. 4).

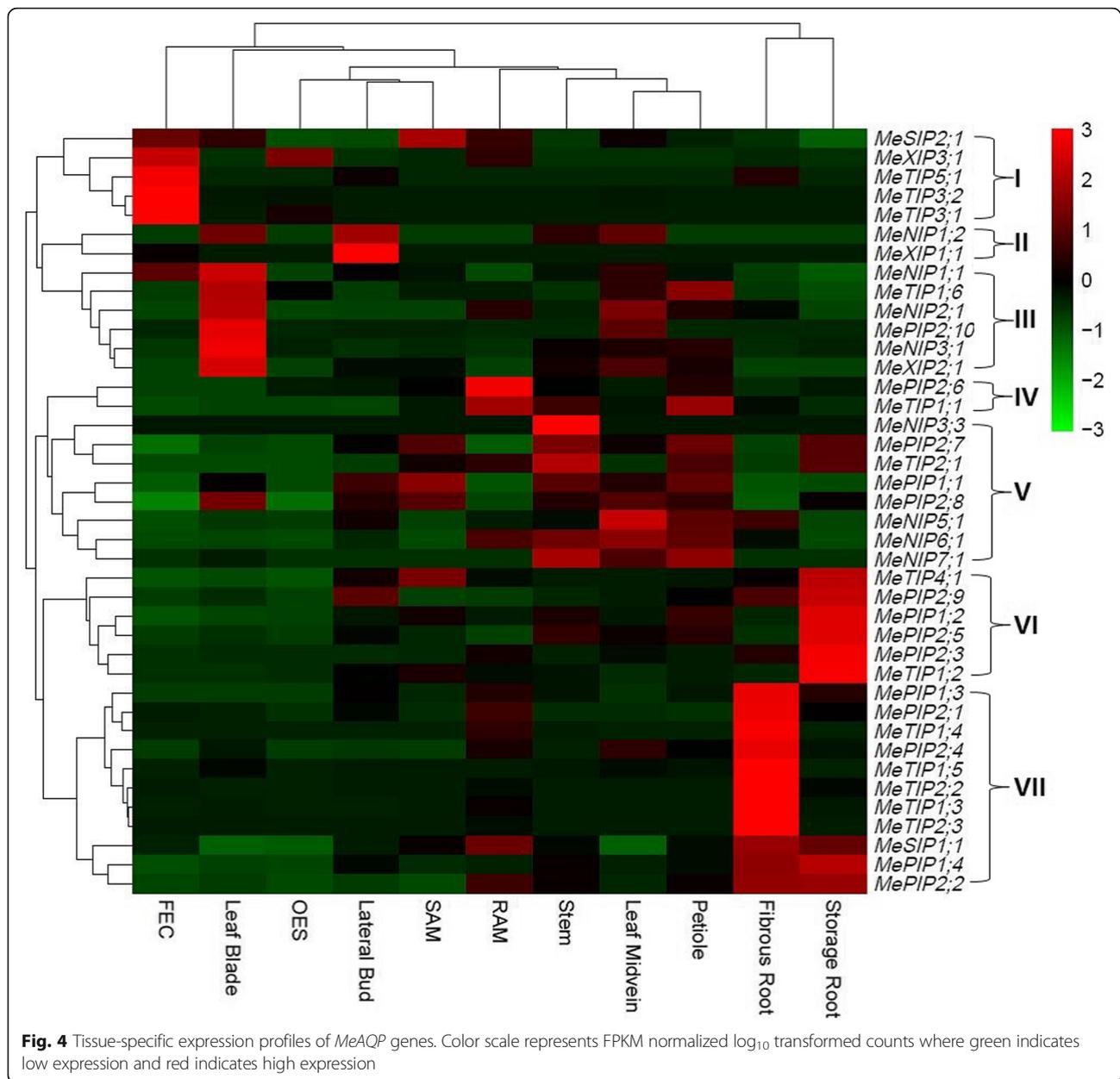
Expression divergence of paralogs was also observed. For example, the expression of *MeXIP3;2* was not detected in all tissues examined, whereas *MeXIP3;1* was expressed in FEC, OES, SAM, RAM, and fibrous root. *MeNIP3;2* and *MeNIP3;3* were only lowly expressed in stem or OES, respectively. Three genes (i.e. *MePIPI;2*, *MeTIP1;5*, and *MeNIP1;1*) were shown to express more than their paralogs

(i.e. *MePIPI;1*, *MeTIP1;6*, and *MeNIP1;2*, respectively) in all examined tissues, whereas *MePIPI;4* was expressed more than *MePIPI;3* in most tissues with the exception of FEC. By contrast, other duplicate pairs seem to complementally express in different tissues (Fig. 4).

#### Discussion

Polyploidy or WGD, multiplication of the whole genome content, is an important evolutionary force that acts as a major mechanism for acquiring new genes. Increasing sequenced genomes showed that WGD is widespread and more than 30 events have been described in plant lineages [68, 69]. At approximately 117 Mya, all core eudicots, including arabidopsis, poplar, physic nut, castor, rubber, and cassava, experienced the  $\gamma$  event [27]. After that, physic nut and castor didn't undergo any additional WGD [3, 4, 17, 70–72], whereas arabidopsis, poplar, rubber, and cassava experienced one or two recent WGDs [1, 2, 5, 6, 9, 11, 12, 29, 30]. Up to date, genome-wide analysis of the *AQP* gene family has been reported in most of these species with the exception of cassava. The physic nut genome was shown to encode 32 *AQP* genes that include one pseudogene (i.e. *JcPIP1;3*), accounting for about 0.12% of total protein-coding genes [1, 18]. Compared with physic nut, castor encodes relatively more *AQP* genes (i.e. 37, also accounting for 0.12% of total predicted genes), reflecting more protein-coding genes present in this species (i.e. 31,221 vs 27,172). Few recent duplicates were identified in physic nut and castor, i.e. *JcSIP1;2/-1;3*, *RcPIP1;2/-1;3*, *RcSIP1;1/-1;2*, *RcXIP1;1/-1;4/-1;2/-1;3* (Additional file 8). They all resulted from local duplication, which is consistent with no recent WGD occurred in these two species [1, 17, 23, 72]. By contrast, except for arabidopsis that experienced chromosomal rearrangement and massive gene loss after WGDs [73, 74], considerably more *AQP* genes were found in poplar and rubber [22–24]. In arabidopsis, 35 *AQP* genes were found and duplicates were shown to result from different modes of gene duplication, i.e.,  $\gamma$  WGD (1),  $\beta$  WGD (2),  $\alpha$  WGD (8), tandem (2), and transposed (4) [21, 29]. There are 55 *AQP* genes in poplar and duplicates were derived from tandem duplication (4) and the recent WGD (20) (Additional file 8). In rubber, 51 *AQP* genes were previously described, however, three genes (i.e. *HbXIP1;2*, *HbXIP1;3*, and *HbXIP1;4*) were shown to be pseudogenes. The remaining 48 *AQP* genes are distributed across 46 scaffolds (Additional file 1). Although conserved synteny can be observed between *HbAQP* duplicate pairs, whether they were derived from the  $\rho$  WGD or segmental duplications still need to be resolved since 7,453 scaffolds available have not been anchored to 18 chromosomes yet [15].

In this study, we present a genome-wide survey and characterization of *AQP* family genes in cassava. The family of 42 members is relatively more than that in physic nut (31) and castor (37), but relatively less than



that in rubber (48) and poplar (55) [1, 18, 22–24]. Phylogenetic analysis of 213 AQPs from cassava, rubber, castor, physic nut, and poplar revealed five main clades representing five subfamilies, i.e. PIP with two groups, TIP with five groups, NIP with seven groups, XIP with three groups, and SIP with two groups. The classification was supported by exon-intron structures, conserved motifs, and BRH-based sequence comparison.

Compared with other examined species, species-specific gene expansion or loss was observed in cassava. Except for *MeXIP3;1/–3;2* that resulted from tandem duplication, synteny analysis revealed that other 13 recent duplicates were derived from the  $\rho$  WGD shared by

rubber. Conserved synteny was also observed between cassava and rubber *AQP* genes, and 39 *HbAQP* genes could be anchored to 13 cassava chromosomes on the basis of synteny analysis (Additional file 9). These results also supported that all 16 *HbAQP* duplicates were derived from the  $\rho$  WGD. Among them, the orthologs of 11 duplicated *HbAQP* genes were also preserved in cassava (Table 3 and Additional file 9). Nevertheless, the average *Ks* value of *AQP* duplicate pairs in rubber (i.e. 0.2609) is relatively smaller than that in cassava (i.e. 0.3730), suggesting a relatively lower rate of evolution of *HbAQP* genes (Table 2). The result is consistent with a slow genome evolution in long-lived woody perennials as well as a considerably

longer generation of rubber than cassava [2, 11, 12, 75, 76]. In fact, similar Ks value (i.e. 0.2713) was also observed in poplar, another big tree species (Additional file 8). The Ks value of *AQP* duplicates resulted from recent WGD varies from 0.2442 to 0.5756 in cassava, from 0.1405 to 0.3491 in rubber, or from 0.2011 to 0.5224 in poplar (Table 2 and Additional file 8), suggesting that the evolution rate is distinct between different duplicate pairs.

Orthology defines genes in different organisms that evolved from a common ancestral gene via speciation. Normally, orthologs retain the same function in the course of evolution [77]. Thereby, identification of orthologs or orthologous groups is useful for functional inference, comparative genomics, and studies on gene/protein evolution [78]. When using poplar as an out-group, the BRH-based sequence comparison revealed 34 OGs present in the common ancestor of Euphorbiaceae plants and each group contains one to six OGs, i.e. PIP1 (4), PIP2 (6), TIP1 (4), TIP2 (2), TIP3 (1), TIP4 (1), TIP5 (1), NIP1 (1), NIP2 (1), NIP3 (2), NIP4 (2), NIP5 (1), NIP6 (1), NIP7 (1), XIP1 (1), XIP2 (1), XIP3 (1), SIP1 (2), and SIP2 (1). Interestingly, six OGs are absent from cassava, i.e. OG1-1c, OG1-1d, OG2-1b, OG3-4a, OG3-4b, and OG5-1b (Table 3). Among them, OG3-4a and OG3-4b belong to the NIP4 group which is widely distributed in most examined species, however, species-specific gene loss of this whole group was observed in cassava. In fact, OG3-4a has also been lost in physic nut [1, 18]. OG1-1c and OG1-1d belong to the PIP1 group which includes four OGs. OG1-1c, which is present in both castor and poplar, has expanded in poplar via WGD, but has been lost in cassava as well as rubber and physic nut. The widely distributed OG1-1d has been lost in cassava as well as poplar. OG2-1b belongs to the TIP1 group which also includes other three OGs. In contrast to most OGs in this group have expanded along with recent WGD, species-specific loss of OG2-1b occurred in cassava after its divergence with rubber. OG5-1b, which belongs to the SIP1 group with two OGs, has expanded in most examined species via WGD or tandem duplication, but has been lost in cassava. Moreover, orthologs of *HbTIP2;2*, *HbTIP5;2*, and *HbSIP1;1* have also been lost in cassava after its divergence with rubber. Species-specific gene loss was also observed in rubber, which includes OG1-1c and OG3-3b. OG3-3b, which belongs to the NIP3 group and has expanded in cassava via WGD, has been lost in rubber as well as castor and poplar (Table 3). The loss of OG3-3b in rubber is more likely to occur after its divergence with cassava. By contrast, it's not easy to determine when the loss of OG1-1c occurred, since it is absent from both rubber and cassava. The exon-intron structure was shown to be highly conserved within orthologous groups and even within phylogenetic groups, though *RcPIP2;5* (a member of OG1-2f) has gain one small intron close to the 5'-terminal [23]. Based on

analysis of conserved protein motifs among different *Me/HbAQP*s, gain or loss of certain motifs was observed within orthologous groups (Fig. 3), suggesting possible functional divergence of cassava and rubber duplicates.

In addition to structural divergence, expression divergence also plays a role in the evolution of duplicates [11, 12, 79, 80]. In our previous studies, tissue-specific expression profiles of castor, physic nut, and rubber *AQP* genes were investigated based on paired-end RNA-seq data generated via the Illumina platform, which revealed similar expression pattern of orthologs in a certain tissue [1, 23, 24]. This case is prevailing between castor and physic nut which usually have no paralog, by contrast, expression divergence of rubber paralogs was frequently observed [1, 18]. As shown in Fig. 4, similar results were also observed in cassava. For example, the transcript level of *MePIP1;4* (the ortholog of *HbPIP1;1*) was relatively higher than *MePIP1;3* (the ortholog of *HbPIP1;2*) in most examined tissues. Nevertheless, different evolutionary patterns were also observed. The transcript level of *MePIP1;2* was relatively higher than *MePIP1;1* in all examined tissues, in contrast, their orthologs in rubber (i.e. *HbPIP1;3* or *HbPIP1;4*, respectively) were shown to exhibit similar expression profiles in bark. Compared with tissue-specific expression of *MePIP2;3* and *MePIP2;4*, *HbPIP2;5*, and *HbPIP2;6* exhibited similar expression patterns in leaf and bark [1, 18, 24]. Thereby, further functional analysis of species-specific isoforms in cassava and rubber is of particular interest.

## Conclusions

This study presents a genome-wide analysis of the *AQP* gene family in cassava, an Euphorbiaceae plant of economic importance. Despite sharing the  $\rho$  WGD, 42 *AQP* family genes in cassava is relatively less than 48 in rubber. These *MeAQP* genes are distributed across 15 chromosomes and conserved synteny can be observed between cassava and rubber *AQP* genes. Phylogenetic and BRH-based sequence analyses further assigned *MeAQP* genes into five subfamilies or 28 out of 34 identified OGs: each subfamily contains two to six phylogenetic groups, and each group includes one to six OGs. In contrast to a predominant role of the  $\rho$  WGD on family expansion in rubber, cassava *AQP* duplicates were derived from the  $\rho$  WGD as well as local duplication. Compared with rubber and other Euphorbiaceae plants, species-specific gene expansion or loss was observed in cassava, which includes the loss of the entire NIP4 group. Furthermore, gene structures, sequence characteristics, and expression profiles of *MeAQP* genes were also investigated, which provides insights into the evolution of *Me/HbAQP* genes, especially functional divergence of recent duplicates. These findings will not only improve our knowledge on family evolution in Euphorbiaceae, but also provide valuable information for future functional analysis of *AQP* genes in cassava and rubber.

## Additional files

**Additional file 1:** Accession numbers of AQPs identified in rubber, castor, physic nut, and poplar. (XLSX 130 kb)

**Additional file 2:** The gene model for *MeSIP2;1*. (PDF 177 kb)

**Additional file 3:** The gene model for *MeNIP1;1*. (PDF 78 kb)

**Additional file 4:** The gene model for *MeXIP3;1*. (PDF 96 kb)

**Additional file 5:** The gene model for *MeXIP3;2*. (PDF 100 kb)

**Additional file 6:** Alignment of cassava AQPs with structure determined Spinach PIP2;1. (PDF 237 kb)

**Additional file 7:** Detailed information of 25 motifs identified in this study. (JPG 409 kb)

**Additional file 8:** List of recent AQP duplicates identified in rubber, castor, physic nut, and poplar. Ks and Ka were calculated using PAML. (XLSX 12 kb)

**Additional file 9:** Matched positions of 39 *HbAQP* genes on cassava chromosomes. The positions were based on synteny analysis, where *HbAQP* genes were marked in orange just following their syntenic genes in cassava. (JPG 1561 kb)

## Abbreviations

AQP: Aquaporin; ar/R: Aromatic/arginine; BRH: Best reciprocal hits; Chr: Chromosome; EST: Expressed sequence tag; FPKM: Fragments per kilobase of exon per million fragments mapped; GLP: Aquaglyceroporphin; GRAVY: Grand average of hydropathicity; Ka: Nonsynonymous substitution rate; Ks: Synonymous substitution rate; MIP: Major intrinsic protein; MW: Molecular weight; Mya: Million years ago; NIP: Nod26-like intrinsic protein; NPA: Asn-Pro-Ala; OG: Orthologous group; ORF: Open reading frame; P1-P5: Residues at P1 to P5 positions; pI: Isoelectric point; PIP: Plasma membrane intrinsic protein; RNA-seq: RNA sequencing; SIP: Small basic intrinsic protein; TIP: Tonoplast intrinsic protein; TM: Transmembrane helix; WGD: Whole-genome duplication; XIP: Uncategorized X intrinsic protein

## Acknowledgements

The authors appreciate those contributors who make related genome and transcriptome datasets accessible in public databases. They would also like to thank reviewers for their careful reading and valuable suggestions.

## Funding

This work was supported by the National Natural Science Foundation of China (31771883 and 31700580), the Natural Science Foundation of Hainan province (319MS093), the Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630052017011 and 1630022019017), and the Research Fund of Guangdong University of Petrochemical Technology (2018rc55). The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

All supporting data can be found in additional files.

## Authors' contributions

The study was conceived and directed by ZZ. All the experiments and analysis were directed by ZZ and carried out by ZZ and JY. ZZ wrote the paper. All the authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interest.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, People's Republic of China. <sup>2</sup>Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, People's Republic of China.

Received: 15 January 2019 Accepted: 7 May 2019

Published online: 15 May 2019

## References

- Zou Z, Yang LF, Gong J, Mo YY, Wang JK, Cao JH, et al. 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.
- Zou Z, Xie GS, Yang LF. Papain-like cysteine protease encoding genes in rubber (*Hevea brasiliensis*): comparative genomics, phylogenetic and transcriptional profiling analysis. *Planta.* 2017;246(5):999–1018.
- Zou Z, Huang QX, Xie GS, Yang LF. Genome-wide comparative analysis of papain-like cysteine protease family genes in castor bean and physic nut. *Sci Rep.* 2018;8(1):331.
- Zou Z, Zhang XC. Genome-wide identification and comparative evolutionary analysis of the Dof transcription factor family in physic nut and castor bean. *PeerJ.* 2019;7:e6354.
- Zou Z, Zhu JL, Zhang XC. Genome-wide identification and characterization of the Dof gene family in cassava (*Manihot esculenta*). *Gene.* 2019;687:298–307.
- Zou Z, Yang JH. Genomics analysis of the light-harvesting chlorophyll a/b-binding (Lhc) superfamily in cassava (*Manihot esculenta* Crantz). *Gene.* 2019;702:171–81.
- Olsen K, Schaal B. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. *Am J Bot.* 2001;88(1):131–42.
- Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, et al. Cassava genome from a wild ancestor to cultivated varieties. *Nat Commun.* 2014;5:5110.
- Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol.* 2016;34(5):562–70.
- Schmitz PM, Kavallari A. Crop plants versus energy plants—on the international food crisis. *Bioorg Med Chem.* 2009;17(12):4020–1.
- Zou Z, Yang JH. Genomic analysis of Dof transcription factors in *Hevea brasiliensis*, a rubber-producing tree. *Ind Crop Prod.* 2019;134:271–83.
- Zou Z, Yang JH, Zhang XC. Insights into genes encoding respiratory burst oxidase homologs (RBOHs) in rubber tree (*Hevea brasiliensis* Muell. Arg.). *Ind Crop Prod.* 2019;128:126–39.
- Zou Z, Liu JT, Yang LF, Xie GS. Survey of the rubber tree genome reveals a high number of cysteine protease-encoding genes homologous to *Arabidopsis* SAG12. *PLoS One.* 2017;12(2):e0171725.
- Bennett MD, Leitch IJ. Nuclear DNA amounts in angiosperms-583 new estimates. *Ann Bot.* 1997;80:169–96.
- Tang C, Yang M, Fang Y, Luo Y, Gao S, Xiao X, et al. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat Plants.* 2016;2(6):16073.
- Chaumont F, Tyerman SD. Plant aquaporins from transport to signaling. *Cham: Springer;* 2017.
- Zou Z. Mining gene families in the castor bean genome. In: Rabinowicz P, Kole C, editors. *The castor bean genome*. Cham: Springer; 2018. p. 135–73.
- Zou Z, Zhang XC. Genomics analysis of physic nut (*Jatropha curcas* L.) aquaporin genes and the comparison with castor bean and rubber tree. *Top 10 contributions on plant biology: 2nd edition*. Cham: Avid science; 2018. p. 2–59.
- Fu D, Libson A, Miercke LJW, Weitzman C, Nollert P, Krucinski J, et al. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science.* 2000;290(5491):481–6.
- Törnroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, et al. Structural mechanism of plant aquaporin gating. *Nature.* 2006;439(7077):688–94.
- Zwiazek JJ, Xu H, Tan X, Navarro-Ródenas A, Morte A. Significance of oxygen transport through aquaporins. *Sci Rep.* 2017;7:40411.

22. Gupta AB, Sankaramakrishnan 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;9:134.
23. 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(10):e0141022.
24. Zou Z, Gong J, An F, Xie GS, Wang JK, Mo YY, et al. Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. *BMC Genomics.* 2015;16:1001.
25. Azad AK, Ahmed J, Alum MA, Hasan MM, Ishikawa T, Sawa Y, et al. 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(6):e0157735.
26. Sonah H, Deshmukh RK, Labbé C, Bélanger RR. 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(1):2771.
27. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* 2012;13(1):R3.
28. Jiao Y, Li J, Tang H, Paterson AH. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell.* 2014;26(7):2792–802.
29. Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature.* 2003;422(6930):433–8.
30. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science.* 2006;313(5793):1596–604.
31. Wang Y, Tan X, Paterson AH. Different patterns of gene structure divergence following gene duplication in *Arabidopsis*. *BMC Genomics.* 2013;14:652.
32. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
33. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9.
34. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40(7):e49.
35. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
36. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
37. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9.
38. Froger A, Tallur B, Thomas D, Delamarche C. Prediction of functional residues in water channels and related proteins. *Protein Sci.* 1998;7(6):1458–68.
39. Kirscht A, Kaptan SS, Bienert GP, Chaumont F, Nissen P, de Groot BL, et al. Crystal structure of an ammonia-permeable aquaporin. *PLoS Biol.* 2016;14(3):e1002411.
40. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(Web Server issue):W202–8.
41. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics.* 1998;14(1):48–54.
42. Wilson MC, Mutka AM, Hummel AW, Berry J, Chauhan RD, Vijayaraghavan A, et al. Gene expression atlas for the food security crop cassava. *New Phytol.* 2017;213(4):1632–41.
43. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28(5):511–5.
44. Wallace IS, Choi WG, Roberts DM. The structure, function and regulation of the nodulin 26-like intrinsic protein family of plant aquaglyceroporins. *Biochim Biophys Acta.* 2006;1758(8):1165–75.
45. Bárzana G, Aroca R, Bienert GP, Chaumont F, Ruiz-Lozano JM. New insights into the regulation of aquaporins by the arbuscular mycorrhizal symbiosis in maize plants under drought stress and possible implications for plant performance. *Mol Plant-Microbe Interact.* 2014;27(4):349–63.
46. Biela A, Grote K, Otto B, Hoth S, Hedrich R, Kaldenhoff R. The *Nicotiana tabacum* plasma membrane aquaporin NtAQP1 is mercury-insensitive and permeable for glycerol. *Plant J.* 1999;18(5):565–70.
47. Chaumont F, Barrieu F, Jung R, Chrispeels MJ. Plasma membrane intrinsic proteins from maize cluster in two sequence subgroups with differential aquaporin activity. *Plant Physiol.* 2000;122(4):1025–34.
48. Eisbarth DA, Weig AR. Dynamics of aquaporins and water relations during hypocotyl elongation in *Ricinus communis* L. seedlings. *J Exp Bot.* 2005;56(417):1831–42.
49. Wang J, An F, Cai XQ, Zou Z, Zhang W, Lin WF. Function characterization and expression analysis of aquaporins (HbPIP1 and HbPIP2) in *Hevea brasiliensis*. *Scientia Silvae Sinicae.* 2014;50:69–75.
50. An F, Zou Z, Cai XQ, Wang J, Rookes J, Lin W, Cahill D, Kong L. Regulation of HbPIP2,3, a latex-abundant water transporter, is associated to the latex dilution and latex yield in rubber trees (*Hevea brasiliensis* Muell. Arg.). *PLoS One.* 2015;10(4):e0125595.
51. Khan K, Agarwal P, Shanware A, Sane VA. Heterologous expression of two *Jatropha* aquaporins imparts drought and salt tolerance and improves seed viability in transgenic *Arabidopsis thaliana*. *PLoS One.* 2015;10(6):e0128866.
52. Yaneff A, Sigaut L, Marquez M, Alleva K, Pietrasanta LI, Amodeo G. Heteromerization of PIP aquaporins affects their intrinsic permeability. *Proc Natl Acad Sci U S A.* 2014;111(1):231–6.
53. Uehlein N, Lovisolo C, Siefritz F, Kaldenhoff R. The tobacco aquaporin NtAQP1 is a membrane CO<sub>2</sub> pore with physiological functions. *Nature.* 2003;425(6959):734–7.
54. Dynowski M, Mayer M, Moran O, Ludewig U. Molecular determinants of ammonia and urea conductance in plant aquaporin homologs. *FEBS Lett.* 2008;582(16):2458–62.
55. Dynowski M, Schaaf G, Loqué D, Moran O, Ludewig U. Plant plasma membrane water channels conduct the signalling molecule H<sub>2</sub>O<sub>2</sub>. *Biochem J.* 2008;414(1):53–61.
56. Wang C, Hu H, Qin X, Zeise B, Xu D, Rappel WJ, et al. Reconstitution of CO<sub>2</sub> regulation of SLAC1 anion channel and function of CO<sub>2</sub>-permeable PIP2,1 aquaporin as CARBONIC ANHYDRASE4 interactor. *Plant Cell.* 2016;28(2):568–82.
57. Bienert GP, Chaumont F. Aquaporin-facilitated transmembrane diffusion of hydrogen peroxide. *Biochim Biophys Acta.* 2014;1840(5):1596–604.
58. Loqué D, Ludewig U, Yuan L, von Wirén N. Tonoplast intrinsic proteins AtTIP2,1 and AtTIP2,3 facilitate NH<sub>3</sub> transport into the vacuole. *Plant Physiol.* 2005;137(2):671–80.
59. Soto G, Alleva K, Mazzella MA, Amodeo G, Muschietti JP. AtTIP1,3 and AtTIP5,1, the only highly expressed *Arabidopsis* pollen-specific aquaporins, transport water and urea. *FEBS Lett.* 2008;582(29):4077–82.
60. Porcel R, Bustamante A, Ros R, Serrano R, Mulet Salort JM. BYCOLD1: a novel aquaporin from sugar beet (*Beta vulgaris* L.) involved in boron homeostasis and abiotic stress. *Plant Cell Environ.* 2018;41(12):2844–57.
61. Niemietz CM, Tyerman SD. Channel-mediated permeation of ammonia gas through the peribacteroid membrane of soybean nodules. *FEBS Lett.* 2000;465(2–3):110–4.
62. Ma JF, Tamai K, Yamaji N, Mitani N, Konishi S, Katsuhara M, et al. A silicon transporter in rice. *Nature.* 2006;440(7084):688–91.
63. Deshmukh RK, Vivancos J, Ramakrishnan G, Guérin V, Carpentier G, Sonah H, et al. A precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *Plant J.* 2015;83(3):489–500.
64. Bienert GP, Bienert MD, Jahn TP, Boutry M, Chaumont F. Solanaceae XIPs are plasma membrane aquaporins that facilitate the transport of many uncharged substrates. *Plant J.* 2011;66(2):306–17.
65. Lopez D, Bronner G, Brunel N, Auguin D, Bourgerie S, Brignolas F, et al. Insights into *Populus* XIP aquaporins: evolutionary expansion, protein functionality, and environmental regulation. *J Exp Bot.* 2012;63(5):2217–30.
66. Ampah-Korsah H, Anderberg HI, Engfors A, Kirscht A, Norden K, Kjellstrom S, et al. The aquaporin splice variant NbXIP1;1a is permeable to boric acid and is phosphorylated in the N-terminal domain. *Front Plant Sci.* 2016;7:862.
67. Noronha H, Araújo D, Conde C, Martins AP, Soveral G, Chaumont F, et al. The grapevine uncharacterized intrinsic protein 1 (VvXIP1) is regulated by drought stress and transports glycerol, hydrogen peroxide, heavy metals but not water. *PLoS One.* 2016;11(8):e0160976.
68. Vanneste K, Baele G, Maere S, Van de Peer Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the cretaceous-Paleogene boundary. *Genome Res.* 2014;24(8):1334–47.
69. Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, et al. Wide-spread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant.* 2018;11(3):414–28.

70. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, et al. Draft genome sequence of the oilseed species *Ricinus communis*. *Nat Biotechnol.* 2010;28(9):951–6.
71. Wu P, Zhou C, Cheng S, Wu Z, Lu W, Han J, et al. Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. *Plant J.* 2015;81(5):810–21.
72. Tsuchimoto S. The *Jatropha* genome. Cham: Springer; 2017.
73. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet.* 2011;43(5):476–81.
74. Proost S, Pattyn P, Gerats T, Van de Peer Y. Journey through the past: 150 million years of plant genome evolution. *Plant J.* 2011;66(1):58–65.
75. Luo MC, You FM, Li P, Wang JR, Zhu T, Dandekar AM, Leslie CA, Aradhya M, McGuire PE, Dvorak J. Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. *BMC Genomics.* 2015;6:707.
76. Xie Z, Wang L, Wang L, Wang Z, Lu Z, Tian D, Yang S, Hurst LD. Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. a low rate in woody perennials and a higher mutagenicity in hybrids. *Proc Biol Sci.* 2016;283(1841):20161016.
77. Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* 2005;39:309–38.
78. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13(9):2178–89.
79. Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell.* 2004;16(7):1679–91.
80. Yang X, Tuskan GA, Cheng MZ. Divergence of the Dof gene families in poplar, *Arabidopsis*, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol.* 2006;142(3):820–30.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

