

Intron and Polypeptide Evolution of Conserved NPA to NPA Motif Regions in Plant Aquaporins

Jiahua Xie,^{1,2,3} Todd C. Wehner,^{1,4} Kurt Wollenberg,^{2,3} Michael D. Purugganan,² and Mark A. Conkling^{2,3}

North Carolina State University, Raleigh NC 27695

ADDITIONAL INDEX WORDS. *Cucumis sativus*, intron position and phase, phylogenetic tree, plant aquaporin

ABSTRACT. Aquaporin proteins are part of an ancient family that functions as water transporting facilitators in all organisms. Phylogenetic and physiological analyses have revealed that plant aquaporins consist of two groups: the plasma membrane intrinsic proteins (PIPs) and the tonoplast intrinsic proteins (TIPs). Using the conserved asparagine-proline-alanine (NPA) to NPA motif regions, we studied the evolution of 35 plant aquaporins that included nine of our newly cloned cucumber aquaporins and 26 from the GenBank database. Results indicated that NPA repeated regions were effective for phylogenetically characterizing the plant aquaporin family, and to accurately localize the introns. Phylogenetic analysis showed that 35 plant aquaporins fell into two distinct groups (except for the *Arabidopsis* gene *AtMip*)—PIPs and TIPs. The nine cucumber aquaporins belong to the PIP group that were localized further into two different sub-groups. The intron analysis showed that introns of plant aquaporins mainly consist of two types. Eighteen PIPs shared identical intron positions localized in connecting loop C between amino acids 95 and 96. Nine TIPs shared the other identical intron positions localized in connecting loop D between amino acids 44 and 45. Cucumber aquaporins CRB9 and CRB10 (with no intron in the repeated NPA regions) may be the result of intron loss events, while intronless rice (*Oryza sativa*) *Os-TIP1* and *Os-TIP2* may have resulted from other intron loss events. *PIP11* and *Os-PIP* do not have the same amino acid number as major PIP members, but combined phylogenetic analysis results along with intron positions and phases showed that they belong to the PIP group. The phylogenetic tree and intron position information suggest that *AtMip* was mis-annotated as a member of aquaporin, and is a homologue of the glycerol facilitator-like protein. Introns share identical positions and phases within the PIP group (except *PIP13*) or the TIP group, but differ between the plasma and the tonoplast membrane aquaporins matching the phylogenetic analysis results. Intron positions of the repeated NPA regions of plant aquaporins that have stable inheritance can act as molecular markers for phylogenetic studies.

Aquaporins belong to the ubiquitous Major Intrinsic Protein (MIP) family that functions as water-selective channels (Chrispeels and Agre, 1994; Chrispeels and Maurel, 1994; Engel et al., 1994; Park and Saier, 1996). Aquaporins together with transporters of glycerol and other small neutral molecules make up the MIP super-family (Johansson et al., 2000; Park and Saier, 1996). Owing to functional importance, a large number of aquaporin genes have been cloned and characterized in animals, plants, and microorganisms. Members of the aquaporin family are highly conserved from bacteria to mammals, all of which have similar sequences as well as similar molecular masses, ranging from 25 to 31 kDa (Pao et al., 1991).

Structurally, the aquaporin polypeptides have similar topology, with all having six membrane-spanning α -helices, five loops, and a C- and N- terminus (Park and Saier, 1996; Reizer et al., 1993). The primary amino acid sequence of aquaporins shows that the N- and C- terminal halves of the protein have obverse symmetry and represent a tandem sequence repeat, each comprised of three bilayer-spanning domains with two connecting loops (Cheng et al., 1997; Jung et al., 1994; Walz et al., 1995). The most highly conserved sequences are two tripeptide (NPA) motifs in the two longest loops, B (cytoplasmic) and E (extracellular), which are conserved over diverse species, including bacteria and animals

(Reizer et al., 1993; Park and Saier, 1996). Site-directed mutagenesis on the loops containing the NPA motifs confirmed that two NPA segments line the path of water permeation, and NPA and adjacent residues are essential for water transporting activity (Jung et al., 1994).

In plants, aquaporins have been localized into vacuolar and plasma membranes and were phylogenetically divided into two distinct groups: the plasma membrane intrinsic proteins (PIPs) and the tonoplast intrinsic proteins (TIPs) (Chrispeels and Agre, 1994; Höfte et al., 1992; Kammerloher and Schäffner, 1993; Kjellbom et al., 1999; Weig et al., 1997; Yamada et al., 1995; Zardoya and Villalba, 2001). PIPs have extended N-termini as compared to TIPs (Johansson et al., 2000). Although much work has been done to characterize the functional localization and phylogenetic relationship of aquaporins in plants, we know of no reports of intron evolution, perhaps because of the small number of plant aquaporin genomic sequences that have been sequenced.

Tobacco aquaporin *TobRB7* was cloned by our group more than 10 years ago (Conkling et al., 1990; Yamamoto et al., 1991). Recently, we have cloned one full length aquaporin genomic sequence and eight genomic fragments which encoded the aquaporin conserved NPA to NPA regions in cucumber (Xie et al., 2002). The *Arabidopsis* and rice (*Oryza sativa*) genomic projects also provided sequences which were annotated as homologous to the aquaporins. In total, 35 plant aquaporin genes (fragments) are available for intron tracing within the repeated NPA regions. Characterization of intron information including number, position, and phase could provide additional information for investigation of molecular evolution of this gene family.

The analysis of phylogenetic relationships and intron evolution in the repeated NPA regions has some advantages over other

Received for publication 2 Jan. 2003. Accepted for publication 1 Apr. 2003. Research supported by a grant from Pickle Packers International.

¹Department of Horticultural Science.

²Department of Genetics.

³Current addresses: Vector Research Ltd, 700 West Main Street, Durham, NC 27701; Kurt Wollenberg, Tufts Center for Vision Research, NEMC, 750 Washington Street, Boston, MA 02111.

⁴Correspondence author; e-mail todd_wehner@ncsu.edu.

approaches. First, the conserved tripeptide NPA is a common feature of aquaporins (Reizer et al., 1993), and thus serves as a signature sequence in the genome. Two NPA motifs have been faithfully inherited throughout evolution, which will be useful for tracing aquaporin evolution. Second, amino acid sequences are very well conserved, both in the transmembrane domains and two NPA-containing loops. However, variation in the N- and C-termini have been reported (Zeidel et al., 1994), because these termini may not directly participate in water transport, but may have a regulating function (Johansson et al., 2000). The analysis only in the repeated NPA regions may be more accurately to localize the intron position by alignment polypeptide of the conserved NPA. Third, the N-terminal half of the aquaporin protein is homologous to the C-terminal half (Johansson et al., 2000). The NPA repeated region contains three membrane spanning domains including the third domain of the first tandem sequence repeat, and the fourth and fifth domains of the second tandem sequence repeat. Therefore, the NPA repeated regions may represent the full length of aquaporin for phylogenetic analysis. Since Heymann and Engel (1999; 2000) suggested the term aquaporin family or AQP family to replace MIP family that has been used widely since the first report of a cloned aquaporin (Gorin et al. 1984). To avoid any confusion, aquaporins in this paper indicate only that those are water-channel-related proteins.

The objective of this study was to compare our nine cucumber aquaporins with 26 plant aquaporins retrieved from the GenBank database to analyze their phylogenetic relationships, and to trace patterns of intron evolution in the repeated NPA regions.

Materials and Methods

CUCUMBER AQUAPORIN FRAGMENTS. Nine cucumber aquaporin fragments (CRB1, CRB2, CRB3, CRB4, CRB5, CRB6, CRB8, CRB9, and CRB10) that encoded conserved NPA to NPA regions were cloned (Table 1) using polymerase chain reaction (PCR)-based single strand conformation polymorphism (SSCP) analysis with two degenerate primers. The primers, Cu1 5'-GG(A/T/C/G)CACATCAA(C/T)CC(A/T/C/G)GC(A/T/C/G)GT-3' and Cu2 5'-ACT GG(A/T/C/G) ATCAA(C/T)CCAGC(A/T/C/G)AG-3', were designed on the two well-conserved NPA motifs (Xie et al., 2002).

CLONING CRB9 GENE. The inverse PCR procedure that was used by Ochman et al. (1988) was modified to clone the full-length genomic CRB9 sequence to determine whether it had an intron in its unsequenced regions, since the CRB9 and CRB10 fragments did not have an intron in the NPA to NPA region. Two primers were designed according to the known sequence of the CRB9 fragment: CRB9D 5'-ATTCCCATGTTCTGTCTTG-3' (down to 3' end) and CRB9U: 5' CCAACGATATCTTTCTT GT-

TAA-3' (up to 5' end). They were synthesized by the Genosys Biotechnologies Inc. (The Woodlands, Texas). About 4 µg of the cucumber genomic DNA was initially digested with the restriction enzyme *NdeI* (CA↓TATG). Digested products were purified by phenol extraction and the alcohol precipitation for ligation. Then 220 µL ligation solution was set containing 22 µL 10× ligation buffer (0.5 M Tris-HCl pH 7.8, 50 mM MgCl₂, 0.1 M β-mercaptoethanol, 50 mM ATP, 5 mg/ml BSA), 5 units of T4 DNA ligase and about 1.0 µg digested DNA. Ligation reaction was carried out at 16 °C for 14 h. Then, ligated DNA solution was purified again using phenol extraction and alcohol precipitation for PCR. The Expand long template PCR system (Boehringer Mannheim, Indianapolis, Ind.) and a Perkin Elmer machine were used for the experiments.

Amplifications were in 0.5 mM dNTP, 0.5 µM each primer, 0.50 µL Expand long template enzyme mixture, 1× Expand long template PCR buffer 1 to a final concentration of 50 mM Tris-HCl, 16 mM (NH₄)₂SO₄, 1.75 mM MgCl₂, pH 9.2, and ≈130 ng digested and circulated DNA was used as template DNA for PCR. The PCR profile was at 92 °C for 2 min for initial denaturation, then the reaction was subjected to 35 cycles of 95 °C for 15 s, 58 °C for 30 s, and 68 °C for 3 min, with a final elongation step of 10 min at 68 °C. A 2.4-kb PCR-amplified fragment was cloned into the pCR2.1 vector with the original TA cloning kit (Invitrogen Corp., Carlsbad, Calif.) and the manufacturer's recommended protocol. DNA fragments were sequenced as dsDNA using a sequencing system (Prism ABI model 377; Perkin-Elmer) with

Table 2. List of aquaporin genes and one glycerol-like protein from GenBank database found in *Arabidopsis*, *Nicotiana*, *Oryza*, *Saccharomyces*, and *Escherichia* species.

Species	Accession no.	Abbreviation (used in paper)	No. of amino acids
<i>A. thaliana</i> (At)	AAA32748	α-TIP	117
<i>A. thaliana</i> (At)	AAA32806	γ-TIP	117
<i>A. thaliana</i> (At)	AAC28529	PIP6	124
<i>A. thaliana</i> (At)	AAC42249	TIP3	117
<i>A. thaliana</i> (At)	AAC64216	PIP7	124
<i>A. thaliana</i> (At)	AAC79629	PIP8	124
<i>A. thaliana</i> (At)	AAD18142	PIP9	124
<i>A. thaliana</i> (At)	AAD31569	TIP4	117
<i>A. thaliana</i> (At)	BAB09071	TIP5	117
<i>A. thaliana</i> (At)	CAB37860	PIP1b	124
<i>A. thaliana</i> (At)	CAB41102	PIP10	124
<i>A. thaliana</i> (At)	CAB51216	TIP6	116
<i>A. thaliana</i> (At)	CAB80801	PIP11	132
<i>A. thaliana</i> (At)	NP_171668	PIP1c	124
<i>A. thaliana</i> (At)	NP_180986	AtMip	122
<i>A. thaliana</i> (At)	NP_181254	PIP2B	124
<i>A. thaliana</i> (At)	NP_181255	PIP2C	124
<i>A. thaliana</i> (At)	NP_193465	TIP7	117
<i>A. thaliana</i> (At)	NP_194071	PIP12	124
<i>A. thaliana</i> (At)	NP_195236	SIMIP	124
<i>A. thaliana</i> (At)	NP_200874	PIP13	124
<i>N. tabacum</i> (Nt)	AAB23597	TobRB7	117
<i>O. sativa</i> (Os)	AAK21347	Os-PIP	120
<i>O. sativa</i> (Os)	AAK98737	Os-TIP1	116
<i>O. sativa</i> (Os)	BAB63833	Os-TIP2	117
<i>O. sativa</i> (Os)	CAC39073	Os-TIP3	116
<i>O. sativa</i> (Os)	BAA04257	Os-Mip	115
<i>S. cerevisiae</i> (Sc)	AAD25168	AQYZ	115
<i>E. coli</i> (Ec)	AAC43518	AQPZ	126

Table 1. Information of nine fragments of cucumber aquaporins.

Name	Fragment length (bp)	Intron length (bp)	Putative amino acids
CRB1	482	110	124
CRB2	467	95	124
CRB3	468	96	124
CRB4	455	83	124
CRB5	482	110	124
CRB6	482	110	124
CRB8	482	110	124
CRB9	372	0	124
CRB10	372	0	124

analysis version 3.4.1. M13 universal forward and reverse primers were used to determine end sequences using the dye deoxy terminator cycle sequencing method.

OTHER AQUAPORIN GENE SEQUENCES. The other 26 plant aquaporin genomic sequences from three species used in these analysis were retrieved from the GenBank database. One yeast aquaporin, one bacterial aquaporin, and one rice glycerol facilitator-like protein were added to construct a phylogenetic tree of the plant aquaporins (Table 2).

SEQUENCE ALIGNMENT AND PHYLOGENETIC TREE CONSTRUCTION. Predicted NPA repeated protein sequences of 37 aquaporin and one glycerol facilitator-like protein genes were aligned and used to construct the phylogenetic tree, which included the nine newly cloned cucumber aquaporins. The amino acid sequences were aligned using the Clustalw algorithm (Thompson et al., 1994). Neighbor-joining (NJ) trees (Saitou and Nei, 1987) were constructed based upon both the *p*-distance (fraction of sites that are different) and Poisson-corrected distances in the MEGA package (Kumar et al., 1994).

COMPARISON OF INTRONS. Together with our nine newly cloned cucumber genomic fragments, there were 35 plant genomic DNA fragments encoding two NPA regions available for analysis of intron evolution. Intron number, position, and phase in the repeated NPA regions of these 35 plant aquaporins have been characterized. The intron position of each gene was calculated by designating the asparagines (N) of the first conserved NPA motif as the first amino acid instead of starting at the encoded codon.

Results and Discussion

CHARACTERIZATION OF INTRON/EXON OF CUCUMBER AQUAPORIN GENES. Nine fragments that encoded the conserved NPA to NPA motifs of cucumber aquaporin were cloned using PCR-based SSCP analysis (Xie et al., 2002). When these nine genomic

	10	20	30	40	50
CRB1	NPAVTFGLFLARKLSL	TRAVFYIVMQCLG	ICGAGVVKGFQ	PKPYDTLGG	
CRB2	NPAVTFGLFLARKVSLVRALLY	IIAQCAGAI	CGCGLVKAFQ	SAYVRYNG	
CRB3	NPAVTFGLFLARKLSL	TRAFIFYIVMQCLG	ICGAGVVKGFQ	PKPYERLGG	
CRB4	NPAVTSGLRLALKISL	VRAFSYIIAQCLG	ICGCLAKSLQKTYVQYNG		
CRB5	NPAVTFGLFLARKLSL	TRAVFYIVMQCLG	ICGAGVVKGFQ	PKPYDTLGG	
CRB6	NPAVTFGLFLARKLSL	TRAVFYIVMQCLG	ICGAGVVKGFQ	PKPYDTLGG	
CRB8	NPAVLSGCFLARKLSL	TRAVFYIVMQCLG	ICGAGVVKGFQ	PKPYDTLGG	
CRB9	NPAVTFGMFLTRKISL	VRALLYIIAQCIGAL	CGCALVKT	TLQRDHYNHYGG	
CRB10	NPAVTFGMFLTRKISL	VRALLYIIAQCIGAL	CGCALVKT	TLQRDHYNHYGG	
	****	* . *	****	* . *	****
	60	70	80	90	100
CRB1	GANMVAKGYTKGDGLGAE	IVGTFILVYTVF	SATDAKRSARD	SHVPILAPL	
CRB2	GANMLSDGYNKGTGLGAE	IIGTFVLVYTVF	SATDPKRNARD	SHVPVLAPL	
CRB3	GANVVS DGYTKGDGLGAE	IVGTFILVYTVF	SATDAKRSARD	SHVPILAPL	
CRB4	AANMVANEYSIGTGLAAE	IIGTFVLVYTVF	SATDPKRNARD	SHI PVLAPL	
CRB5	GANMVAKGYTKGDGLGAE	IVGTFILVYTVF	SATDAKRSARD	SHVPILAPL	
CRB6	GANMVAKGYTKGGGLGAE	IVGTFILVYTVF	SATDAKRSARD	SHVPILAPL	
CRB8	GANMVAKGYTKGDGLGAE	IVGTFILVYTVF	SATDAKRSARD	SHVPILAPL	
CRB9	GANQLVDGYSRGTGLAVE	IMGTFILLYTVF	SATDPKRNARD	SHVPVLAPL	
CRB10	GANQLVDGYSRGTGLAVE	IMGTFILLYTVF	SATDPKRNARD	SHVPVLAPL	
	**	.	*	*	****
	110	120			
CRB1	PIGFAVFLIHLTTPI	ITGTGINPA			
CRB2	PIGFAVFMVHLATIP	ITGTGINPA			
CRB3	PIGFAVFLVHLATIP	ITGTGINPA			
CRB4	PIGFAVTMVHLATIP	ITGTGINPA			
CRB5	PIGFAVFLVHLATIP	ITGTGINPA			
CRB6	PIGFAVFLIHLATIP	ITGTGINPA			
CRB8	PIGFAVFLVHLATIP	ITGTGINPA			
CRB9	PIGFAVFLVHLATIP	ITGTGINPA			
CRB10	PIGFAVFLVHLATIP	ITGTGINPA			
	*****	. . *	****	*****	

Fig. 1. Alignment of the amino acid sequences of nine cucumber aquaporins.

DNA fragments were compared with the cDNA derived from RT-PCR amplification (data not shown), seven out of nine fragments (CRB1, CRB2, CRB3, CRB4, CRB5, CRB6, and CRB8) were found to have an intron inside the NPA repeated region. All introns of these seven fragments had canonical 5' and 3' splice sites (GT--AG) (Lee et al. 1991) with 83 to 110 bp nucleotides (Table 1). CRB9 and CRB10 did not have an intron in that region. However, the length of the exon region of seven fragments with

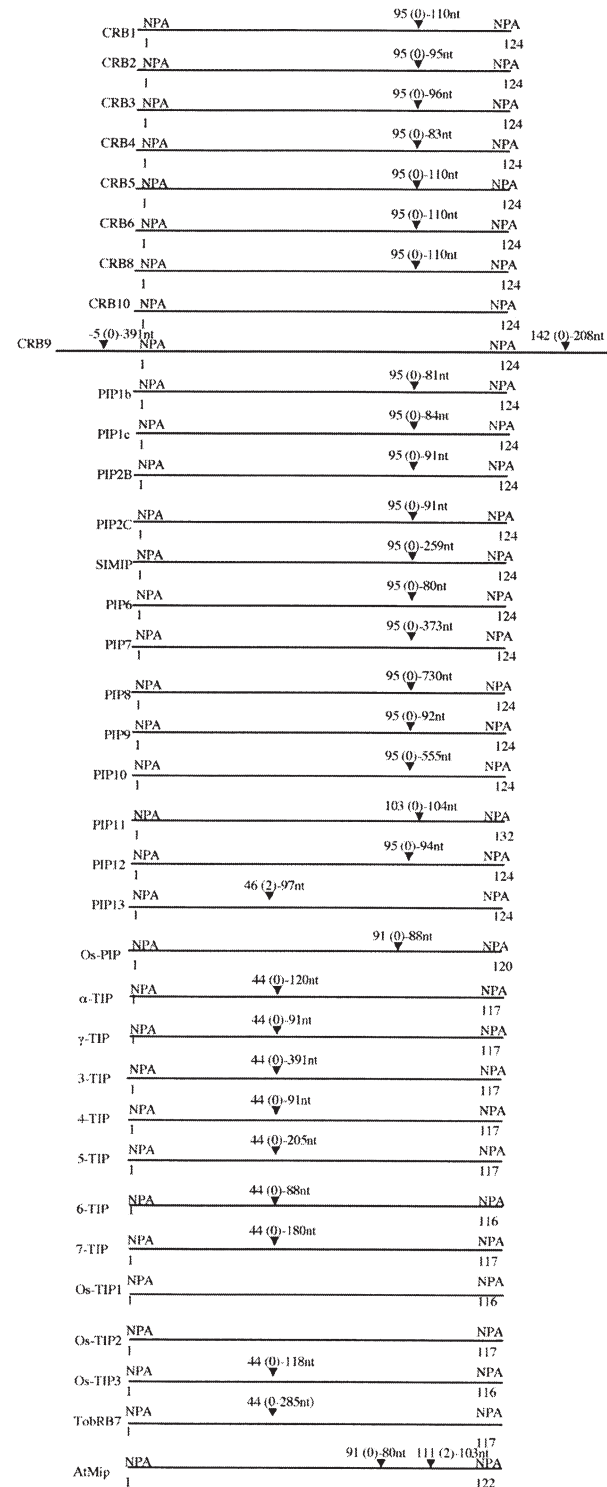


Fig. 2. Introns in the conserved NPA to NPA motif regions of 34 plant aquaporins and one glycerol facilitator-like protein *AtMip*. Intron length, intron position and phase of each gene are indicated (nt = nucleotides).

an intron and the two fragments without an intron was the same: 372 nucleotides encoding 124 putative amino acids (Table 1). The amino acid sequences of nine cucumber aquaporins were aligned to compare their homologies (Fig. 1). Alignment results showed that CRB9 and CRB10 have an identical polypeptide sequence even though they have a 6 base pair nucleotide difference in the NPA conserved region (Xie et al., 2002). CRB1, CRB5, CRB6, and CRB8 are highly conserved. There were only 2, 2, 4, 2, 2, and 4 amino acid differences, respectively in the NPA repeated regions between the pairs of CRB1/CRB5, CRB1/CRB6, CRB1/CRB8, CRB5/CRB6, CRB5/CRB8, and CRB6/CRB8. On the other hand, CRB2, CRB3, and CRB4 had much variation when compared with each other and with the other members of the group.

Genomic fragments CRB9 and CRB10 do not have an intron in the NPA to NPA regions, and no hybridization signal has been captured using CRB9 and CRB10 fragments as probes in the Northern blot analysis with root and leaf RNAs, while the other seven fragments with introns were expressed either on leaves or roots, or both (data not shown). In order to determine whether it was a real gene with introns in the unsequenced regions, we cloned the full length of CRB9 using inverse PCR and the known sequence.

A 2.4 kb fragment 1.6 kb upstream of the first NPA motif and 0.8 kb downstream of the second NPA motif was amplified using inverse PCR. The full length CRB9 sequence was characterized with 1364 nucleotides. Comparing the genomic sequence of CRB9 with its cDNA (data not shown), there are two introns in CRB9 gene which are located outside of the NPA repeated region. One intron is 391 bp in length located upstream of the first NPA, and the second is 208 bp in length located downstream of the second

NPA (Fig. 2). The encoded region of CRB9 has 765 nucleotides with 255 putative amino acids. That is similar to other plant aquaporins (Park and Saier, 1996; Froger et al., 1998). Therefore, CRB9 (having 2 introns) can be considered a real gene. CRB10 may also be a real gene and have introns in the other regions, owing to its having identical amino acid sequence with CRB9 in the conserved NPA repeated region. The two could be used for evolutionary analysis with other aquaporins.

PHYLOGENETIC ANALYSIS OF REPEATED NPA REGIONS. It has been reported widely in the literature that the plant aquaporins function as water-selective channels in the plasma and tonoplast membranes of cells, and that aquaporins can be classified phylogenetically into plasma membrane and tonoplast intrinsic proteins (Chrispeels and Agre, 1994; Park and Saier, 1996; Weig et al., 1997; Zardoya and Villalba, 2001). In the present study, phylogenetic analysis using the NPA repeated motif regions of 35 plant aquaporins (including 26 from the GenBank database and nine newly cloned cucumber aquaporins) also yields two distinct phylogenetic groups: the PIPs and the TIPs, with the exception of *Arabidopsis* gene *AtMip* (Accession: NP_180986) (Fig. 3).

Our results are similar to previous reports that plant aquaporins consist of two groups (Chrispeels and Agre, 1994; Park and Saier, 1996). The NPA repeated motif regions can be used to represent whole protein sequences for phylogenetic analysis, perhaps because the primary sequence of aquaporins was an obverse symmetry with internal homology (Reizer et al., 1993). The repeated NPA to NPA region contains the third domain of the first tandem sequence repeat and the fourth and fifth domains of the second tandem sequence repeat, and the third domain is homologous to the sixth. Thus, the repeated NPA regions have the full information of one repeated tandem sequence. However, the phylogenetic study using only the repeated NPA region polypeptides of 73 plant aquaporins instead of the whole gene failed to recognize clear subgroups within PIPs and TIPs (data not shown). The sub-groups within PIPs and TIPs could be identified (using high bootstrap support) when the whole gene was analyzed (Zardoya and Villalba, 2001).

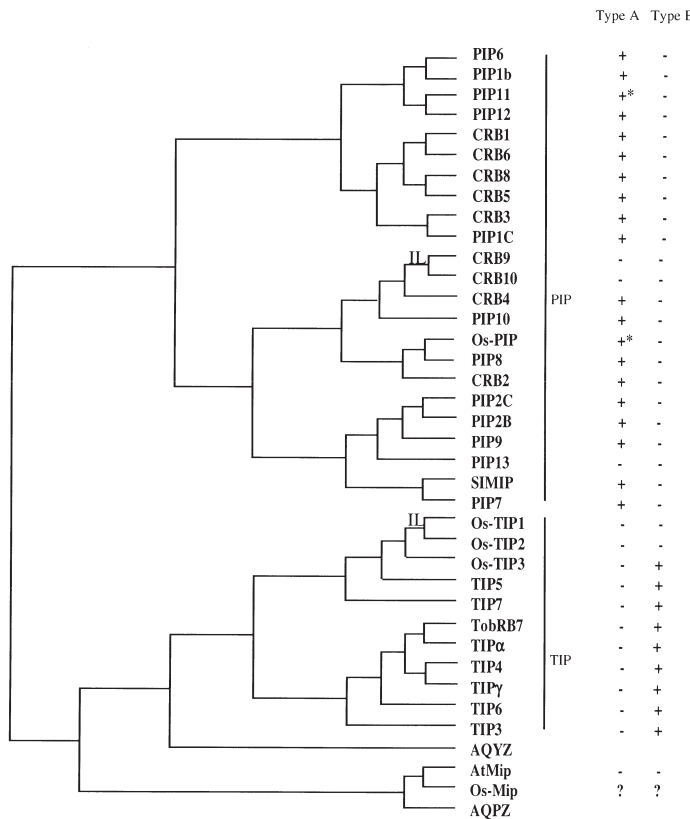


Fig. 3. Phylogeny of plant aquaporins. One yeast aquaporin (*AQYZ*), one bacterial aquaporin (*AQPZ*), and one rice glycerol facilitator-like protein were added to root and construct a phylogenetic tree for the plant aquaporins. Plant aquaporins are divided into two groups: PIPs and TIPs. *Introns of *PIP11* and *Os-PIP* may belong to type A intron. IL = Intron loss.

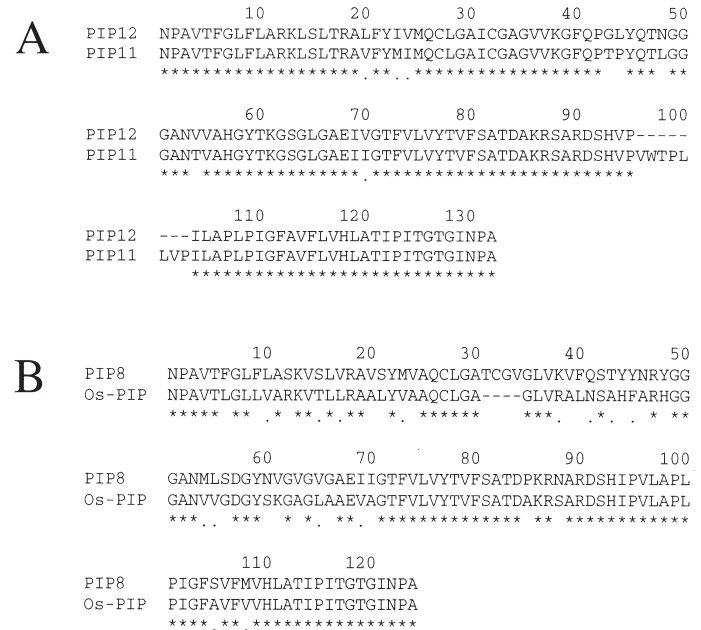


Fig. 4. Amino acid sequence alignment of four plant aquaporins. A: *PIP11* and *PIP12*; B: *PIP8* and *Os-PIP*. Four amino acids of *Os-PIP* were deleted while 8 amino acids of *PIP11* were inserted.

The nine newly cloned cucumber aquaporins belong to the PIPs, and fall into two distinct subgroups. All nine newly cloned cucumber aquaporins are PIPs that may be due to the problem of the degeneracy of primers. The two degenerate primers Cu1 forward and Cu2 reverse that we used to clone the cucumber aquaporins were based on the conserved amino acid sequences GHINPAV and TGINPAR of the plant PIPs (Xie et al., 2002). The forward and reverse degenerate primers based on the conserved region of GH(L)VNPAV and A(G)S(C)MNPARG from the TIPs (Park and Saier, 1996) may help to clone the tonoplast intrinsic protein genes.

The gene *AtMip* derived from the *Arabidopsis* genomic project and annotated as an aquaporin on the GenBank databases (Lin et al., 1999) is highly homologous to the rice glycerol facilitator-like protein *Os-Mip* (Accession: BAA04257) (Liu et al., 1994). In our analysis, these two genes were classified into a single group. Glycerol-like proteins also have high similarity with aquaporins that also possess the repeated NPA motifs (Reizer et al., 1993). The PIPs, the TIPs, and the glycerol-like protein are three members of the MIP super-family (Park and Saier, 1996). Therefore, *AtMip* is probably a glycerol facilitator-like protein that was misannotated.

Amino acid sequences of aquaporins were highly conserved from diverse bacteria to mammals (Park and Saier, 1996). In plants, the number of amino acids in the NPA regions is also highly conserved. For example, 21 out of 23 members of the PIPs have 124 amino acids in the two conserved NPA motif regions. The two exceptions are *PIP11* (Accession: CAB80801) and *Os-PIP* (Accession: AAK21347) with 132 and 120 amino acids, respectively. However, *PIP11* shares very high identity at the amino acid level to *PIP12* despite the presence of eight amino acids more, while *Os-PIP* has high homology with the *PIP8* except that it has four amino acids fewer (Fig. 4). Polypeptide alignment results implied that the amino acid number differences of *PIP11* and *Os-PIP* with major PIPs members might be the result of an insertion and a deletion event, respectively. For the 11 members of the TIP group, eight have 117 amino acids, and three have 116 amino acids.

Comparing the PIPs with the TIPs by amino acid sequence alignment, the TIP aquaporins were seven to nine amino acids shorter than those of the major PIPs. The shortening in the TIPs occurred at connecting loop C with four amino acids, and in loop D with three to five amino acids, respectively. None occurred inside the membrane-spanning domain. The highly conserved nature of the aquaporin family might indicate its important function for all organisms in moving water into the cell. Aquaporins may be some of the few early genes that have been under strong stabilizing selection and did not undergo much modification when they duplicated into homologous members. Thus, their sequences are conserved from species in groups as diverse as bacteria and mammals.

Phylogenetic studies of plant aquaporins indicate that both aquaporins of monocotyledonous and dicotyledonous plants have members that belong to the PIPs and TIPs. That suggests the diversification of the water channel protein into two different groups occurred before the divergence of monocotyledonous and dicotyledonous plants about 200 million years ago (Laroche et al., 1995). The two groups play different roles in the tonoplast membrane and plasma membrane.

INTRONS WITHIN NPA REPEATED REGIONS. Introns within the repeated NPA region of 35 plant aquaporins [including nine newly cloned aquaporins from cucumber and 26 aquaporins

from *Arabidopsis*, rice, and tobacco (*Nicotiana tabacum*)] are shown in Fig. 2. For the introns, the asparagine (N) of the first conserved NPA motif was counted as the first amino acid. Of the 35 aquaporins, four do not have an intron, one has two introns, and the remaining 30 have one intron in the NPA repeated region. Based on their intron/exon structure, the 35 plant aquaporin genes were divided into five groups. The first group had two members: CRB9 and CRB10, each with 124 amino acids but no intron in the two NPA conserved regions. The second group had two members: *Os-TIP1* and *Os-TIP2* with 116 and 117 amino acids but no intron in that region. The third group had 18 members (including seven cucumber aquaporins) with 124 amino acids and one intron in that region. The location of introns of the 18 genes are precisely conserved between the 95th and 96th amino acids with phase 0, although the sequence and size of each intron differs. The intron of this group is defined as type A. The fourth group, with nine members, has 117 amino acids and one intron. Members of the group share identical position between the 44th and 45th amino acids with phase 0. The intron of this group was defined as type B. The last group consists with the remaining four members: *PIP11*, *PIP13*, *Os-PIP* and *AtMip*. Their introns were different from each other, and different from those of the previous four groups.

Phylogenetic analysis indicated that cucumber aquaporins CRB10 and CRB9 with no intron and 18 members from three species with the type A intron fall into the same PIP group, with 124 putative amino acids in the NPA to NPA repeated region (Figs. 2 and 3). CRB9 and CRB10 are highly homologous to cucumber aquaporins CRB4 and CRB2, and *Arabidopsis* aquaporin *PIP10* (which has one intron in the NPA repeated region). This suggests that cucumber aquaporin CRB9 and CRB10 evolved through the loss of introns, since at least 18 other aquaporin sequences are now known, and all have introns in the identical location and phase. The loss of introns was precise, such that the reading frame was not altered. Intron deletion has reportedly occurred often during evolution, and may be associated with genome slimming (Petrov et al. 1996). The mechanism of intron loss was proposed to be due to gene conversion with an intronless reverse-transcribed cDNA intermediate (Geiss et al., 1994; Lewin, 1983). Analogical analysis suggests that two intronless rice aquaporins *Os-TIP1* and *Os-TIP2* also evolved through the loss of a type B intron.

Polypeptide analysis has demonstrated that eight more amino acids of *PIP11* and four fewer amino acids of *Os-PIP* than those of the other PIPs in the NPA repeated regions might have resulted from insertion and deletion events (Fig. 4). If eight inserted amino acids of *PIP11* were taken away and four deleted amino acids of *Os-PIP* were put back, the intron positions and phases of *PIP11* and *Os-PIP* would be the same as a type A intron. Therefore, introns of *PIP11* and *Os-PIP* probably belong to type A. *PIP13* has an identical amino acid number (124) with major members of the PIP group, but its intron position and phase differ from the type A intron. The intron of *PIP13* may belong to an independent group, or may have resulted from a type A intron transposition. Whether the current intron of *PIP13* resulted from an intron transposition of type A intron or not needs further study. *AtMip* has two introns whose intron positions were neither type A nor type B, and that were identical to those of nodulin 26 (Miao and Verma, 1993). This result further illustrated that *AtMip* did not belong to the PIPs or TIPs, and may be homologous to the glycerol-like protein gene.

In summary, the introns studied here share identical positions and phases within the PIPs (except *PIP13*) or the TIPs, but dif-

fer between the plasma and the tonoplast membrane aquaporins. That matches our phylogenetic analysis based on polypeptide alignment indicating that plant aquaporins contain two groups. Introns sharing identical positions within phylogenetic groups, but differing among phylogenetic groups of a family, have been reported in the T-box gene family (Wattler et al., 1998). For T-box genes, the second introns within the subgroup share identical position but differ among subgroup, and some subgroups do not even have that intron.

Constructional analysis of intron locations indicated that type A and B introns were localized in loops C and D, respectively, instead of inside the spanning domain. This is similar to the previous hypothesis that introns are more likely to locate in the module boundary region (Gilbert et al., 1997). The above polypeptide analysis also indicated that several amino acid differences between the PIPs and TIPs occurred inside loops C and D, and were close to type A and B introns. The difference of several amino acids and introns occurred in loops C and D, suggesting that these regions allow introns to be inserted or deleted. This phenomenon also implies that the construction of loops C and D is less important in the aquaporin than the six spanning domains and two long function loops B and E. Functional analysis of aquaporin polypeptides also supports the idea that loops C and D are only interlinking loops, and that loops B and E are functional (Johansson et al., 2000).

Introns from 20 genes in the PIPs and 9 genes from the TIPs have remained completely stable over time, suggesting that both type A and type B introns should exist in plant aquaporins at the same time. That would be before the time of divergence of the tonoplast and plasma membrane proteins, because of the unlikely event of the alternative explanation: an intron moving identically in 20 PIP and 9 TIP genes, followed by two PIP genes losing type A introns and two TIP genes losing type B introns. Since phylogenetic analysis showed that the divergence of the PIPs and TIPs occurred before the divergence of monocotyledonous and dicotyledonous plants about 200 million years ago (Laroche et al., 1995), these two intron positions have been inherited and faithfully preserved for at least 200 million years. However, the introns themselves may have mutated beyond recognition.

Comparing only the introns of the NPA region of the two groups of plant aquaporins is not sufficient to determine their origin, but identical intron positions and phases within functionally similar groups does provide a stable marker. It suggests that each group has a common evolutionary origin, since independent insertions into the same position are unlikely to occur (Wattler et al., 1998). Introns with identical positions have been suggested to serve as molecular markers for phylogenetic analysis (Venkatesh et al., 1999), although introns are evidenced to have been both lost and gained during evolution.

Our results confirmed that intron position and phase in the repeated NPA region can be used as evolutionary markers to discriminate between the PIPs and the TIPs. In the aquaporin family, more variations at both the C- and N-terminus variations were observed by comparing the central part of the sequence (Johansson et al., 2000; Zeidel et al., 1994). Using only the NPA to NPA motif region to trace the intron position and phase of aquaporin family may provide the most accurate information for classification. If a newly cloned plant aquaporin has a type A intron, it may belong to plasma membrane intrinsic protein group.

Literature Cited

- Cheng, A., A.N. van Hoek, M. Yeager, A.S. Verkman, and A.K. Mitra. 1997. Three-dimensional organization of a human water channel. *Nature* 387:627–630.
- Chrispeels, M.J. and P. Agre. 1994. Aquaporins: Water channel proteins of plant and animal cells. *Trends Biochem. Sci.* 19:421–425.
- Chrispeels, M.J. and C. Maurel. 1994. Aquaporins: The molecular basis of facilitated water movement through living plant cells? *Plant Physiol.* 105:9–13.
- Conkling, M.A., C.L. Cheng, Y.T. Yamamoto, and H.M. Goodman. 1990. Isolation of transcriptionally regulated root-specific genes from tobacco. *Plant Physiol.* 93:1203–1211.
- Engel, A., T. Walz, and P. Agre. 1994. The aquaporin family of membrane water channels. *Curr. Opin. Struct. Biol.* 4:545–553.
- Froger, A., B. Tallur, D. Thomas, and C. Delamarque. 1998. Prediction of functional residues in water channels and related proteins. *Protein Sci.* 7:1458–1468.
- Geiss, K.T., G.M. Abbas, and M.A. Makaroff. 1994. Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. *Mol. Gen. Genet.* 243:97–105.
- Gilbert, W., S.J. de Souza, and M.Y. Long. 1997. Origin of genes. *Proc. Natl. Acad. Sci. USA* 94:7698–7703.
- Gorin, M.B., S.B. Yancey, J. Cline, J.P. Revel, and J. Horwitz. 1984. The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning. *Cell* 39:49–59.
- Heymann, J.B. and A. Engel. 1999. Aquaporins: phylogeny, structure and physiology of water channels. *News Physiol. Sci.* 14:187–194.
- Heymann, J.B. and A. Engel. 2000. Structural clues in the sequences of the aquaporins. *J. Mol. Biol.* 295:1039–1053.
- Höfte, H., L. Hubbard, J. Reizer, D. Ludevid, E.M. Herman, and M.J. Chrispeels. 1992. Vegetative and seed-specific isoforms of a putative solute transporter in the tonoplast of *Arabidopsis thaliana*. *Plant Physiol.* 99:561–570.
- Johansson, I., M. Karlsson, U. Johanson, C. Larsson, and P. Kjellbom. 2000. The role of aquaporins in cellular and whole plant water balance. *Biochim. Biophys. Acta* 1465:324–342.
- Jung, J.S., G.M. Preston, B.L. Smith, W.B. Guggino, and P. Agre. 1994. Molecular structure of the water channel through aquaporin CHIP. *J. Biol. Chem.* 269:14648–14654.
- Kammerloher, W. and A.R. Schäffner. 1993. PIP—an *A. thaliana* plasma membrane MIP homologue cloned by expression in mammalian cells, p. 185. In: 5th Intl. Conf. *Arabidopsis* Research, Ohio State Univ.
- Kjellbom, P., C. Larsson, I. Johansson, M. Karlsson, and U. Johanson. 1999. Aquaporins and water homeostasis in plants. *Trends Plant Sci.* 8:308–314.
- Kumar, S., K. Tamura, and M. Nei. 1994. MEGA: Molecular evolutionary genetics analysis software for microcomputers. *Comput. Appl. Biosci.* 10:189–191.
- Laroche, J., P. Li, and J. Bousquet. 1995. Mitochondrial DNA and monocot-dicot divergence time. *Mol. Biol. Evol.* 12:1151–1156.
- Lee, V.D., M. Stapleton, and B. Huang. 1991. Genomic structure of chlamydomonas caltractin: Evidence for intron insertion suggests a probable genealogy for the EF-hand super-family of proteins. *J. Mol. Biol.* 221:175–191.
- Lewin, R. 1983. How mammalian RNA returns to its genome. *Science* 219:052–1054.
- Lin, X.Y., S.S. Kaul, S. Rounsley, T.P. Shea, M.I. Benito, C.D. Town, C.Y. Fujii, T. Mason, C.L. Bowman, M. Barnstead, T.V. Feldblyum, C.R. Buell, K.A. Ketchum, J. Lee, C.M. Ronning, H.L. Koo, K.S. Moffat, L.A. Cronin, M. Shen, G. Pai, S. Van Aken, L. Umayam, L.J. Tallon, J.E. Gill, M.D. Adams, A.J. Carrera, T.H. Creasy, H.M. Goodman, C.R. Somerville, G.P. Copenhaver, D. Preuss, W.C. Nierman, O. White, J.A. Eisen, S.L. Salzberg, C.M. Fraser, and J.C. Venter. 1999.

- Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402:761–768.
- Liu, Q., M. Umeda, and H. Uchimiya. 1994. Isolation and expression analysis of two rice genes encoding the major intrinsic protein. *Plant Mol. Biol.* 26:2003–2007.
- Miao, G.H. and D.P.S. Verma. 1993. Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterogonous plants. *Plant Cell* 5:781–794.
- Ochman, H., A.S. Gerber, and D.L. Hartl. 1988. Genetic applications of an inverse polymer chain reaction. *Genetics* 120:621–625.
- Pao, G.M., L.F. Wu, K.O. Johnson, H. Höfte, M.J. Chrispeels, G. Sweet, N.N. Sandal, and M.H. Saier, Jr. 1991. Evolution of the MIP family of integral membrane transport proteins. *Mol. Microbiol.* 5:33–37.
- Park, J.H. and M.H. Saier, Jr. 1996. Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J. Membrane Biol.* 153:171–180.
- Petrov, D.A., E.R. Lozovskaya, and D.L. Hartl. 1996. High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384:346–349.
- Reizer, J., A. Reizer, and M.H. Saier Jr. 1993. The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway evolution, and proposed functional differentiation of the two repeated halves of the proteins. *Cri. Rev. Biochem. Mol. Bio.* 28:235–257.
- Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.
- Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Venkatesh, B., Y. Ning, and S. Brenner. 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc. Natl. Acad. Sci. USA* 96:10267–10271.
- Walz, T., D. Typke, B.L. Smith, P. Agre, and A. Engel. 1995. Projection map of aquaporin-1 determined by electron crystallography. *Nat. Struct. Bio.* 2:730–732.
- Wattler, S., A. Russ, M. Evans, and M. Nehls. 1998. A combined analysis of genomic and primary protein structure defines the phylogenetic relationship of new members of the T-Box family. *Genomics* 48:24–33.
- Weig, A., C. Deswarte, and M.J. Chrispeels. 1997. The major intrinsic protein family of *Arabidopsis* has 23 members that form three distinct groups with functional aquaporins in each group. *Plant Physiol.* 114:1347–1357.
- Xie, J., T.C. Wehner, and M.A. Conkling. 2002. PCR-based single-strand conformation polymorphism (SSCP) analysis to clone nine aquaporin genes in cucumber. *J. Amer. Soc. Hort. Sci.* 127: 925–930.
- Yamada, S., M. Katsuhara, W. Kelly, C. Michalowski, and H. Bohmert. 1995. A family of transcripts encoding the water channel proteins: tissue-specific expression in the common ice plant. *Plant Cell* 7:1129–1142.
- Yamamoto, Y.T., C.G. Taylor, G.N. Acedo, C.L. Cheng, and M.A. Conkling. 1991. Characterization of cis-acting sequences regulating root-specific gene expression in Tobacco. *Plant Cell* 3:371–382.
- Zardoya, R. and S. Villaba. 2001. A phylogenetic framework for the aquaporin family in eukaryotes. *J. Mol. Evol.* 52:391–404.
- Zeidel, M.L., S. Nielsen, B.L. Smith, S.L. Ambudkar, A.B. Maunsbach, and P. Agre. 1994. Ultrastructure, pharmacologic inhibition, and transport selectivity of aquaporin channel-forming integral protein in proteoliposomes. *Biochemistry* 33:1606–1615.