Phylogenetic and Expression Analysis of the Glutamate-Receptor–Like Gene Family in *Arabidopsis thaliana*

Joanna C. Chiu, Eric D. Brenner,* Rob DeSalle,† Michael N. Nitabach,* Todd C. Holmes,* and Gloria M. Coruzzi**

*Department of Biology, New York University; and †Department of Entomology, American Museum of Natural History, New York

The ionotropic glutamate receptor (iGluR) gene family has been widely studied in animals and is determined to be important in excitatory neurotransmission and other neuronal processes. We have previously identified ionotropic glutamate receptor–like genes (*GLR*s) in *Arabidopsis thaliana*, an organism that lacks a nervous system. Upon the completion of the *Arabidopsis* genome sequencing project, a large family of *GLR* genes has been uncovered. A preliminary phylogenetic analysis divides the *AtGLR* gene family into three clades and is used as the basis for the recently established nomenclature for the *AtGLR* gene family. We performed a phylogenetic analysis with extensive annotations of the iGluR gene family, which includes all 20 *Arabidopsis GLR* genes, the entire iGluR family from rat (except NR3), and two prokaryotic iGluRs, *Synechocystis GluR0* and *Anabaena* GluR. Our analysis supports the division of the *AtGLR* gene family into three clades and identifies potential functionally important amino acid residues that are conserved in both prokaryotic and eukaryotic iGluRs as well as those that are only conserved in *AtGLR*s. To begin to investigate whether the three *AtGLR* clades represent different functional classes, we performed the first comprehensive mRNA expression analysis of the entire *AtGLR* gene family. On the basis of RT-PCR, all *AtGLR*s are expressed genes. The three *AtGLR* clades do not show distinct clade-specific organ expression patterns. All 20 *AtGLR* genes are expressed in the root. Among them, five of the nine clade-II genes are root-specific in 8-week-old *Arabidopsis* plants.

Introduction

Sequence homology screening has become an increasingly popular method for cloning genes of interest, drawing from the large amount of sequence data generated from genome sequencing projects of various prokaryotic and eukaryotic organisms. Using the *Arabidopsis thaliana* genomic sequence and expressed sequence tag database, Lam et al. (1998) identified two complementary DNA clones, *GLR1.1* and *GLR3.1* (a.k.a. *GLR1* and *GLR2*), and two genomic sequences, *GLR2.1* and *GLR3.4* (a.k.a. *GLR3* and *GLR4*), that have sequence similarity with animal ionotropic glutamate receptor (iGluR) genes. In addition to primary sequence similarity, the *Arabidopsis GLR (AtGLR)* genes are also predicted to be similar to animal iGluRs, in terms of secondary structure, i.e., transmembrane topology, as shown by hydropathy plots (Lam et al. 1998). Genes that have sequence similarity with the other family of animal glutamate receptors, the metabotropic glutamate receptors, have not been found in *Arabidopsis*. However, Turano et al. (2001) suggested a possible evolutionary link between *AtGLR*s and seven transmembrane G-protein–linked receptors, which includes metabotropic glutamate receptors as well as GABA receptors.

Animal iGluR genes encode subunits for ligand-gated ion channels, which account for a major fraction of fast excitatory neurotransmission. They form a large gene family that can be divided into multiple classes (*N*-methyl-D-aspartate [NMDA], α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid [AMPA]-kainate [KA], and Delta) based on ligand selectivity and ion conductance properties (Sprengel and Seeburg 1995). NMDA iGluRs have been implicated in neurodegenerative diseases and are believed to be involved in neuronal cell death for pathological conditions such as ischemia (Rameau et al. 2000). AMPA and KA receptors are often grouped together as non-NMDA iGluRs because of ligand cross-reactivity and their similarities in sequence as well as ion-conducting properties (Sutcliffe, Wo, and Oswald 1996). We will refer to AMPA and KA iGluRs as a single class in this paper for the reasons mentioned above (AMPA-KA). Lastly, the delta iGluRs do not appear to have functional ion channel and ligand-binding activity in their wild-type form (Araki et al. 1993; Lomeli et al. 1993). However, the widely studied ''Lurcher'' mutation in the δ2 subunit results in a constitutively open ion channel with properties similar to AMPA-KA receptor channels and leads to neurodegeneration in mice (Wollmuth et al. 2000). δ2 receptors are found to express predominantly in cerebellar Purkinje cells (Kohda, Wang, and Yuzaki 2000).

Because animal iGluRs have been studied extensively in mammalian CNS (Sprengel and Seeburg 1995), the discovery of their putative homologs in *Arabidopsis*, an organism without a nervous system, was surprising (Lam et al. 1998). By performing an initial phylogenetic analysis of animal iGluRs and four *Arabidopsis GLR* genes, Chiu et al. (1999) determined that the divergence of animal iGluRs and *Arabidopsis GLR*s preceded the divergence of animal iGluR classes (AMPA-KA, NMDA, Delta). This preliminary comparative phylogenetic analysis of plant *GLR*s and animal iGluRs sug-

Abbreviations: AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid; GlnH, glutamate-binding domain; iGluR, ionotropic glutamate receptor; KA, kainate; M, transmembrane domain; NMDA, *N*-methyl-D-aspartate; P, pore.

Key words: glutamate receptor, plant, *Arabidopsis thaliana*, mRNA expression, character mapping.

Address for correspondence and reprints: Gloria Coruzzi, Department of Biology, New York University, 100 Washington Sq. East, New York 10003. E-mail: gloria.coruzzi@nyu.edu.

Mol. Biol. Evol. 19(7):1066–1082. 2002

Table 1
GenBank Accession Numbers and Nomenclature for the *AtGLR* Gene Family

			cDNA		Genomic	
		Previous Name	Full Length	Splice Variants	BAC	Protein ID
I	*AtGLR1.1*	*AtGLR1*[a]	AF079998		AC016829	AAF26802.1
	AtGLR1.2		AY072064[b]	AY072065[b]	AB020745.1	BAA96960.1
	AtGLR1.3				AB020745.2	BAA96961.2
	AtGLR1.4		AY072066[b]	AY072067[b]	AC009853	AAF02156.1
II.	*AtGLR2.1*	*AtGLR3*[a]			AF007271	AAB61068.1
	AtGLR2.2		AY072068[b]		AC007266.1	AAD26895.1
	AtGLR2.3				AC007266.2	AAD26894.1
	AtGLR2.4				AL031004	CAA19752.1
	AtGLR2.5				AL360314.1	CAB96656.1
	AtGLR2.6				AL360314.2	CAB96653.1
	AtGLR2.7		AY072069[b]		AC005315.1	AAC33239.1
	AtGLR2.8	*Glur9*	AJ311495		AC005315.2	AAC33237.1
	AtGLR2.9				AC005315.3	AAC33236.1
III	*AtGLR3.1*	*AtGLR2*[a], *ACL1*[c]	AF079999	AF038557	AC002329	AAF63223.1
	AtGLR3.2	*AtGluR2*[d]	AF159498		AL022604	CAA18740.1
	AtGLR3.3				AC025815	AAG51316.1
	AtGLR3.4	*AtGLR4*[a], *GLUR3*[e]	AF167355	AY072070[b]	AC000098	AAB71458.1
	AtGLR3.5	*GLR6*	AF170494		AC005700.1	AAC69939.1
	AtGLR3.6				AL133452	CAB63012.1
	AtGLR3.7	*GLR5*	AF210701		AC005700.2	AAC69938.1

[a] Lam et al. 1998.
[b] Current paper.
[c] *ACL1* corresponds to the splice variant AF038557.
[d] Kim et al. 2001.
[e] *GLUR3* is deposited into GenBank as AF167355.

gested that signaling by amino acids might be a primitive mechanism that existed before the divergence of plants and animals. This hypothesis was subsequently verified with the identification of the first prokaryotic functional ionotropic glutamate receptor, *GluR0,* in *Synechocystis* (cyanobacteria) (Chen et al. 1999).

It is currently unknown whether *Arabidopsis GLR*s, like animal iGluRs and *Synechocystis* GluR0, are capable of forming functional ion channels. Glutamate-gated calcium fluxes have been observed in *Arabidopsis* roots (Dennison and Spalding 2000), but links of this activity to specific *AtGLR* genes have not yet been established. Despite the fact that the biochemical properties of GLR proteins remain an open question, the *AtGLR* genes have already been implicated in processes, such as light signal transduction (Lam et al. 1998; Brenner et al. 2000) and calcium homeostasis (Kim et al. 2001) in plants. We therefore expect that *AtGLR*s will be involved in many different signaling and physiological processes in plants, especially because a large family of 20 *AtGLR* genes has been uncovered with the completion of the *Arabidopsis* genome sequencing project (AGI 2000). A nomenclature has recently been established for the *AtGLR* gene family (Lacombe et al. 2001). The nomenclature is based on a preliminary parsimony analysis that divides the *AtGLR* gene family into three clades.

In this paper, we confirm the identification of 20 *GLR* genes in *Arabidopsis* and report the characterization of five new full-length *AtGLR* cDNA sequences. We examine the genealogical relationships of the iGluR family and identify residues that are invariant (''invariant'' carries the meaning of ''unvaried'' rather than ''invariable'' in this paper) in both prokaryotic and eukaryotic iGluRs by performing a parsimony analysis with extensive annotations using the amino acid sequences of a complete family of animal iGluRs, all 20 *Arabidopsis GLRs,* and two prokaryotic cyanobacterial iGluRs, *Synechocystis* GluR0 and *Anabaena* GluR. *Synechocystis* GluR0 encodes the first identified functional prokaryotic iGluR channel subunit (Chen et al. 1999). On the other hand, functional studies on the *Anabaena* GluR have not been published yet. This study demonstrates that phylogenetic analysis can be a powerful tool to guide functional studies of large gene families as well as to understand evolutionary relationships. We also examine mRNA expression patterns of all 20 *GLR* genes in different organs and determine cell-type–specific expression patterns for a gene from each of the three clades. This analysis has allowed us to investigate whether the phylogenetic division of the *AtGLR* gene family has functional implications by testing the hypothesis of clade-specific gene expression patterns.

Materials and Methods

Obtaining Sequences for Phylogenetic Analysis

The complete family of 20 *Arabidopsis GLR* genes identified upon the completion of the *Arabidopsis* genome sequencing project is listed in table 1. These 20 *AtGLR* genes were identified by BLAST search (Altschul et al. 1997) using *AtGLR1.1* cDNA sequence (AF079998) as the query sequence. The *AtGLR* sequences used in our parsimony analysis include a combination of full-length complementary DNA sequences that were available at the time of the analysis (*GLR1.1,*

Table 2
GenBank Accession Numbers and Abbreviations for Two Prokaryotic Glutamate Receptors, Rat Glutamate Receptors, and Bacterial Periplasmic Amino Acid–Binding Proteins Examined in this Study

Gene	Species	Accession Number	Abbreviation Used (if different from name)
Bacterial periplasmic amino acid binding protein			
glnH	*Escherichia coli*	X14180	ecoliglnh
glnP	*Archaroglobus fulgidus*	AE001090	afgglnp
glnP	*Salmonella typhimurium*	U73111	salglnp
Prokaryotic iGluR			
GluR0	*Synechocystis PCC6803*	D90909	SynGluR0
GluR	*Anabaena PCC7120*	AP003591	AnaGluR
Rat iGluR			
GluRA	*Rattus norvegicus*	*X17184*	*RatK1*
GluRB	*R. norvegicus*	*X54655*	*RatK2*
GluRC	*R. norvegicus*	*M36420*	*RatGluRC*
GluRD	*R. norvegicus*	*M36421*	*RatGluRD*
GluR5	*R. norvegicus*	*M83560*	*RatGluR5-1*
GluR6	*R. norvegicus*	*Z11715*	*RatGluR6*
GluR7	*R. norvegicus*	*M83552*	*RatGluR7*
KA-1	*R. norvegicus*	*U08257*	*RKA1*
KA-2	*R. norvegicus*	*Z11581*	*RKA2*
Delta 1	*R. norvegicus*	*Z17238*	*RDELTA1*
Delta 2	*R. norvegicus*	*Z17239*	*RDELTA2*
NMDAR1	*R. norvegicus*	*X63255*	*RNMDAR1*
NR2A	*R. norvegicus*	*M91561*	*RNR2A*
NR2B	*R. norvegicus*	*M91562*	*RNR2B*
NR2C	*R. norvegicus*	*M91563*	*RNR2C*
NR2D	*R. norvegicus*	*L31612*	*RNR2D*

GLR1.2, *GLR1.4*, *GLR2.2*, *GLR2.7*, *GLR3.1*, *GLR3.2*, *GLR3.4*, *GLR3.5*, and *GLR3.7*) and predicted cDNA sequences from genomic sequence analysis for the remaining genes. Whereas the full-length cDNA sequences for *GLR1.1*, *GLR3.1*, *GLR3.2*, *GLR3.5*, and *GLR3.7* were obtained from GenBank, the full-length cDNA clones for *GLR1.2*, *GLR1.4*, *GLR2.2*, and *GLR2.7* were newly constructed for this analysis by RT-PCR (see *Materials and Methods*—RT-PCR section), and the full-length cDNA clone for *GLR3.4* was obtained through cDNA screening. The remaining 10 *AtGLR* sequences used in our analysis (*GLR1.3*, *GLR2.1*, *GLR2.3*, *GLR2.4*, *GLR2.5*, *GLR2.6*, *GLR2.8*, *GLR2.9*, *GLR3.3*, and *GLR3.6*) were predicted from genomic sequences. In our preliminary alignment, cDNA sequences predicted by GenBank were used for all *AtGLR* genes for which cDNA clones were not available. Upon inspection of the preliminary alignment, conserved iGluR exon sequences in some of the predicted *AtGLR* genes were regarded as intron sequences by the gene prediction program used by GenBank and were spliced out. We therefore edited the predicted cDNA sequence from GenBank for the *AtGLR* genes in question using alignment and existing and newly constructed *AtGLR* cDNAs as guidelines.

The sequences for the two prokaryotic iGluRs (*Synechocystis* GluR0 and *Anabaena* GluR) and the rat iGluR genes used in this study were obtained from GenBank (table 2). Bacterial periplasmic amino acid–binding protein sequences from *Escherichia*, *Salmonella*, and an archaebacteria were included in this analysis as outgroups (table 2), as in our previous analysis (Chiu et al. 1999).

Phylogenetic Analysis

Full-length amino acid sequences were aligned by CLUSTAL X using three separate sets of alignment parameters (gap-to-change = 10, 15, and 20; gap extension cost = 1; amino acid substitution matrix = Blosum 30). The three values for gap-to-change ratio were arbitrarily chosen as reasonable starting points for examining the effect of gap-to-change cost on final alignment and hence on tree topology. A gap extension cost of 1 is conservative and will produce reasonable alignments without excessive hypotheses of gap insertions. These alignments were examined for alignment-ambiguous regions, i.e., columns of amino acids that change from alignment-cost to alignment-cost (Gatesy, DeSalle, and Wheeler 1993). These regions of alignment instability were culled from the matrix to ensure validity of the analysis. Before parsimony analysis, the matrix was also gap-coded as described in Chiu et al. (1999). In brief, we chose to gap-code the regions in our alignment where long stretches of positions showed gaps, so that these regions are not weighted excessively in our phylogenetic analysis. Our gap coding was implemented as unordered. Indel events, which created gaps, oftentimes spanned several amino acid positions, and such regions were recoded as single multistate characters on the basis of their lengths (DeSalle and Brower 1997). The treatment of bacterial amino acid–binding protein sequences presented another data-coding issue. There are two regions in these bacterial genes that show similarity with eukaryotic iGluRs (Paas 1998). These two regions correspond to the glutamate-binding domains in iGluRs and

are used to root the present parsimony analysis as well as our previous analysis (Chiu et al. 1999). All other characters in the glutamate receptor sequences that are not in these two regions and are absent in the bacterial sequences were coded as missing in the outgroups.

In order to strengthen and confirm our analysis, we also performed a second parsimony analysis where we processed the data matrix using elision (Wheeler, Gatesy, and DeSalle 1995) instead of a combination of culling and gap-coding. As in the first parsimony analysis, we used CLUSTAL X to align the sequences using different sets of alignment parameters (gap-to-change = 10, 15, and 20; gap extension cost = 1; amino acid substitution matrix = Blosum 30). We then combined the three matrices to generate a larger matrix, which we used for parsimony analysis.

We used PAUP* (phylogenetic analysis using parsimony; Swofford 1998) to infer phylogeny. Ten random-addition heuristic searches with TBR branch swapping were implemented, and all characters were equally weighted in our analyses. We did not place a limit on the number of trees saved in each search. When more than one equally parsimonious tree was obtained, we generated strict consensus trees to represent our phylogenetic hypothesis. To measure the robustness of all nodes in our parsimony analysis, bootstrap analysis (100 replicates) was performed using PAUP*. Ten random-addition heuristic searches with TBR branch swapping were used for each replicate of the bootstrap analysis.

Computation of Percent Identity Values

We used CLUSTAL X to generate the percent identity values between different iGluRs when comparing the entire amino acid sequences. The default parameters (gap-to-change cost = 10; gap extension cost = 1; amino acid substitution matrix = Blosum 30) were used for the alignment, and percent identity values were generated along with the alignment. We then compiled these values to generate—(1) the percent identity values between genes within the same rat iGluR classes or *AtGLR* clades (AMPA and KA iGluRs are grouped together as a single class), (2) the percent identity values between genes that are in different rat iGluR classes or *AtGLR* clades, and (3) the percent identity values between rat and *Arabidopsis* iGluR genes. The percent identity values for the glutamate-binding domains as well as the pore (P) and surrounding transmembrane regions (M1 and M2) were generated using the same method. The sequence boundaries for the functional domains in rat iGluR genes used in this study follow the boundaries that were previously established by structural and functional studies (Hollmann and Heinemann 1994; Stern-Bach et al. 1994). The sequence boundaries of transmembrane domains in *AtGLR* genes were predicted by the software PHD version 1996.1 (Rost 1996). The computer predictions generally coincided with the *AtGLR* transmembrane domain boundaries predicted when using the alignment of the *AtGLR* genes and rat iGluR genes as guidelines. In cases where the two predictions did not match exactly, they only differed by one or two amino acids. The sequence boundaries for the glutamate-binding domains were predicted using the alignment of *AtGLR*s and rat iGluRs as guidelines because programs specializing in predicting glutamate-binding domains are not available. Amino acid residues numbers 406–527 and 628–737 were used for the alignment for the glutamate-binding domains, and amino acid residues numbers 543–626 were used for the alignment for the P and surrounding transmembrane regions. The amino acid numbering follows the sequence of *AtGLR1.1* (AF079998). The compiled percent identity values were then mapped onto a schematized version of a tree generated from our parsimony analysis.

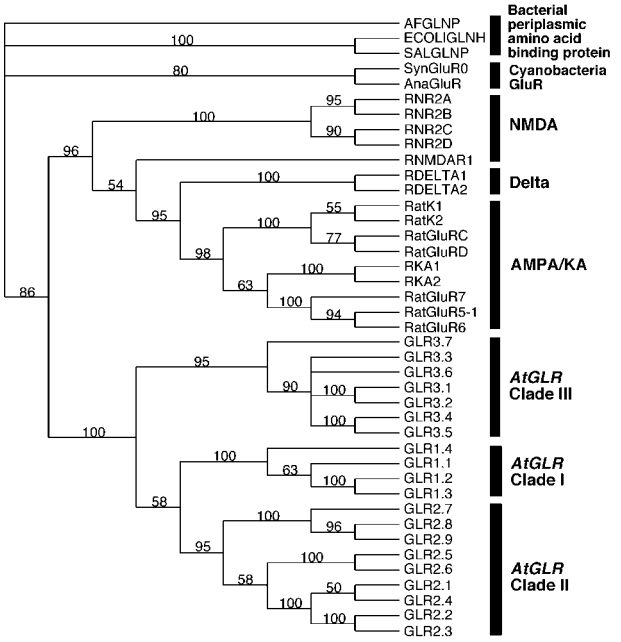

FIG. 1.—*Arabidopsis GLR* genes group into three distinct clades. Phylogenetic tree generated from parsimony analysis of amino acid sequences of rat iGluRs, *AtGLRs*, and two prokaryotic cyanobacterial iGluRs using three bacterial periplasmic amino acid–binding protein sequences as outgroups. Our search resulted in two equally parsimonious trees. The bootstrap consensus tree is shown here.

Character Mapping

We imported the processed data matrix used for parsimony analysis and the consensus tree obtained from parsimony analysis into MacClade version 4.0 (Maddison WP and Maddison DR 1992) for character-mapping analysis. We identified conserved residues at different nodes by using the trace character function in MacClade.

For mapping RT-PCR expression data on the parsimony tree, we converted *AtGLR* mRNA expression in leaves, roots, flowers, and siliques as detected by RT-PCR into binary format (0 = undetected and 1 = detected) and appended the four additional characters to the matrix used to generate the parsimony tree shown in figure 1. The corresponding characters for the rat and cyanobacteria genes are coded as missing. We used this new matrix, with four additional characters, to generate a second parsimony tree. The resulting phylogenetic tree

is identical to the one shown in figure 1. We then used the trace-character function in MacClade to map each of the four expression characters onto the phylogenetic tree.

Examining AtGLR Expression by RT-PCR

First-strand synthesis reactions using Thermoscript RT (Invitrogen) were performed on mRNA (200 ng) generated from leaves, roots, flowers, and siliques, respectively (Ambion PolyA Pure mRNA isolation kit). Leaf and root tissues were collected from 8-week-old *Arabidopsis* (Columbia Col-0 ecotype) plants grown in short days (8 h light-16 h dark) and hydroponic conditions (1 mM KH_2PO_4, 1 mM $MgSO_4$, 0.1 mM NaFe-EDTA, 50 μM KCl, 30 μM H_3BO_3, 5 μM $MnSO_4$, 1 μM $ZnSO_4$, 1 μM $CuSO_4$, 0.1 μM $(NH_4)_6Mo_7O_{24}$, 0.5 mM KNO_3, and 0.25 mM $Ca(NO_3)_2$). Flowers and siliques were collected from mature *Arabidopsis* plants (8-week-old Columbia Col-0 ecotype) grown under long-day condition (16 h light-8 h dark). To examine the expression patterns of all 20 *AtGLR* genes, 20 sets of specific primers corresponding to the distinct *AtGLR* genes (sequences for specific primers are available from authors upon request) were designed. The specificity of the primers was tested by using them to amplify *Arabidopsis* genomic DNA in which all *AtGLR* genes are present at the same copy number. The resulting PCR products were then sequenced to determine if the expected products were amplified. The 20 primer sets were used to perform PCR using the cDNA produced from the first strand synthesis reactions as templates (Takara Ex-Taq polymerase from Panvera). All 20 sets of primers were designed to span intron sequences, thus allowing us to account for signals resulting from genomic DNA contamination. Primers that amplify *TUB5* (accession number M84702) from *Arabidopsis* (5′-CTCGCAACAATA CATCTCACTCACC-3′ and 5′-AAGACACCCAAAAT AAATGGAAACTTC-3′) were used as control primers to ensure uniform amplification and were added to each PCR reaction together with the *AtGLR* gene-specific primers. All primers were used at a concentration of 40 ng/μl. Thermocycling time and temperature were as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, and a final extension period of 72°C for 10 min. Because of the difference in primer efficiencies, we cannot compare the expression levels between different *AtGLR* genes. We therefore assess the relative difference in expression levels in different organs for each *AtGLR* gene, respectively. It is important to note that the RT-PCR results presented here are not truly quantitative because of the problem of saturation (30 amplification cycles used). However, we confirmed the relative expression level of each *AtGLR* gene in the four different organ types by performing a second set of PCR reactions using 20 amplification cycles. In addition, the same experiments were repeated using 40 amplification cycles to address the problem of inefficient amplification in cases where PCR products were not detected.

We analyzed the PCR products using gel electrophoresis followed by NIH image version 1.62. The graphic representation of the RT-PCR data is generated by normalizing *AtGLR* expression against *TUB5* expression. To construct each *AtGLR* graph, we measured the integrated density values of the *TUB5* band and the *AtGLR* band for each of the four organ lanes on the digitalized picture of the agarose gel (samples are from 30 PCR cycles) using NIH image. The integrated density values of the four *AtGLR* bands were then normalized using the integrated density values of the four *TUB5* bands as control. The normalized integrated density values of the four *AtGLR* bands were then used to calculate the relative intensity values shown on the graph. The highest of the four intensity values was regarded as 1, and the others were calculated and represented as a fraction of the highest value.

The above RT-PCR protocol was also used to obtain full-length cDNA clones for *GLR1.2*, *GLR1.4*, *GLR2.2*, and *GLR2.7*. However, instead of using the gene-specific primers used to examine expression patterns, primers that were predicted from genomic sequence analysis to represent the beginning and end of each full-length cDNA sequence were used in the PCR amplification reactions.

Examining AtGLR Expression in Transgenic Plants

Representative *AtGLR* genes from each of the three *GLR* clades were examined for cell-type expression patterns using a reporter gene system. Transcriptional fusions were made between the promoters of *GLR1.1*, *GLR2.1* or *GLR3.1*, and the reporter gene β-glucuronidase (GUS) (Jefferson 1989). *AtGLR* promoter regions were amplified by PCR from Columbia (Col-0) genomic DNA using Ex-Taq polymerase (Panvera). A 1,713-bp region upstream of the start codon of *GLR1.1* was amplified at the 5′ end with the primer JC99 (5′-CGTC GAAGCTTATAAGAAACG-3′) and at the 3′ end with the primer JC101 (5′-CATATCTACTTGTGCCATGG-3′). The amplified fragment has a *Hin*dIII site at the 5′ end and an *Nco*I site at the 3′ end so that GUS is translated at the presumptive ATG start site for *GLR1.1*. A 1,553-bp region upstream of the start codon of *GLR2.1* was amplified at the 5′ end with the primer EBP027 (5′-GTCGACGATCAAAGGTTATGTCGCTAAAGGAG-3′) and at the 3′ end with the primer EBP017 (5′-GT GGATCCTACTTAGCCGAAAAAGAATGAAACTTG-3′). This introduced a *Sal*I restriction site at the 5′ end and an *Nco*I site at the 3′ end. The *Nco*I site facilitated the translation of GUS at the presumptive start ATG site for *GLR2.1*. A 1,780-bp upstream region from *GLR3.1* was amplified at the 5′ end with the primer EBP4 (5′-TGAAGCTTCGTTCACTAATTGGAGTGCAT-3′) and at the 3′ end with EBP1 (5′-ATGACCATGGAGCTTA ACATTGAACAACAAAAAGAG-3′). This introduced a *Hin*dIII restriction site at the 5′ end and an *Nco*I site at the 3′ end so that GUS is translated at the presumptive start ATG site for *GLR3.1*. All fragments were cut at the introduced restriction sites and cloned into pTZGUS (Ngai, Tsai, and Coruzzi 1997). The promoter::GUS fu-

sions were excised with *Eco*RI and *Hin*dIII for *GLR1.1* and *GLR3.1* and with *Eco*RI and *Sal*I for *GLR2.1* and independently subcloned into the binary Agrobacterium vector pBI101 (Jefferson, Kavanaugh, and Bevan 1987) at those respective sites. Binary constructs were transformed into Agrobacterium strain GV3101, and *Arabidopsis* plants (Columbia Col-0 ecotype) were transformed (Clough and Bent 1998). At least 20 independent transgenic lines for each construct were selected for analysis. Transgenic *Arabidopsis* seedlings were germinated on MS media containing 1% sucrose and grown in 16 h light-8 h dark day-night cycles. GUS staining was performed according to the method of Jefferson, Kavanaugh, and Bevan (1987).

Results

Arabidopsis GLR Gene Family: Genomic and cDNA Clone Identification

We have identified a total of 20 distinct open reading frames that represent glutamate-receptor–like genes in *Arabidopsis* (also listed in Lacombe et al. 2001) by using *AtGLR1.1* (AF079998) as the query sequence to perform a BLAST search (Altschul et al. 1997) against the completed *A. thaliana* genome (AGI 2000). Table 1 lists the 20 *AtGLR* genes using the recently established nomenclature for glutamate receptor genes in *Arabidopsis* (Lacombe et al. 2001). Here we report the characterization of the corresponding full-length complementary DNA clones for five new *GLR* cDNAs, *GLR1.2*, *GLR1.4*, *GLR2.2*, *GLR2.7*, and *GLR3.4*, isolated from cDNA libraries or by RT-PCR. Together with the full-length cDNA clones that we have previously isolated, *GLR1.1* and *GLR3.1* (Lam et al. 1998), and the full-length cDNA clones that were isolated by other groups, *GLR2.8*, *GLR3.2*, *GLR3.5*, and *GLR3.7*, a total of 11 *AtGLR* genes are currently found to have corresponding cDNAs (table 1). Although cDNAs have not been isolated for the other nine members, Northern and RT-PCR analyses presented herein indicate that all *AtGLR* genes are expressed (see RT-PCR section).

Parsimony Analysis of iGluRs from Prokaryotes and Eukaryotes Defines Three AtGLR Clades

Previously, a phylogenetic analysis of the iGluR gene family was performed using animal iGluRs and the four *AtGLR* genes (*GLR1.1*, *GLR2.1*, *GLR3.1*, and *GLR3.4*) that were available at the time (Chiu et al. 1999). This initial analysis showed that the divergence of animal iGluRs and *AtGLR*s preceded the divergence of animal iGluR classes (AMPA-KA, NMDA, and Delta). The completion of the *Arabidopsis* genome sequencing project led to the identification of 16 additional *AtGLR* genes. We therefore performed a new, more comprehensive phylogenetic analysis using all the 20 *AtGLR* gene family members of the model plant *Arabidopsis* (table 1) and all known iGluR genes of a single animal, rat (except NR3; table 2). In addition, two prokaryotic iGluRs, *Synechocystis GluR0* (Chen et al. 1999) and *Anabaena* GluR (table 2), were also included in this phylogenetic analysis to examine where they fit into the evolutionary history, relative to the complete plant and animal glutamate receptor gene families.

Visual inspection of the alignment of all *Arabidopsis*, rat, and two prokaryotic glutamate receptor genes reveals a high degree of similarity in most of the iGluR functional domains. These include the two glutamate-binding domains (GlnH1 and 2), transmembrane domain M1, the P, and transmembrane domain M2. On the other hand, most of the sequences before GlnH1 and the regions after GlnH2, which includes the last transmembrane domain M3 (absent in *Synechocystis GluR0* and *Anabaena* GluR), show limited similarity. These regions of low conservation and ambiguous alignment were culled (Gatesy, DeSalle, and Wheeler 1993) accordingly during matrix processing before the first parsimony analysis. Although these highly variable regions may not be absolutely required for protein activity, they may contain domains with modulatory functions, e.g., regulatory kinase-binding sites (Nitabach et al. 2001).

Parsimony analysis of the iGluR gene family using bacterial sequences as outgroups gave two equally parsimonious trees of 4,235 steps, with a consistency index of 0.616 and a retention index of 0.730. Figure 1 shows the consensus parsimony tree with bootstrap values. As shown in our previous analysis (Chiu et al. 1999), the divergence of animal iGluR and *AtGLR* precedes the divergence of animal iGluR classes (AMPA-KA, NMDA, and Delta). This phylogenetic inference is supported by in vivo data showing that glutamate-gated calcium fluxes in *Arabidopsis* roots, which may be linked to *AtGLR* gene products, cannot be induced when L-glutamate is replaced by animal iGluR class-specific ligands, such as AMPA or NMDA (Dennison and Spalding 2000). The two prokaryotic iGluRs, *Synechocystis GluR0* and *Anabaena* GluR, are sister taxa, and they fall outside the multicellular clade. The well-characterized rat iGluR genes belonging to different animal iGluR classes all reside within their appropriate clades, with the exception of RNMDAR1. Instead of forming a monophyletic clade with the other four NMDA iGluRs (RNR2A to 2D), it groups with the rest of the animal iGluRs with a weak bootstrap value of 54% (fig. 1). It is important to note that although RNMDAR1 has indeed diverged greatly from the other four NMDA iGluRs, its placement outside the clade consisting of the other four NMDA genes is only weakly supported. In fact, in a similar analysis, where all but the *Anabaena* GluR sequence is included, RNMDAR1 forms a monophyletic clade with the other four NMDA genes (bootstrap value = 62%, data not shown). In general, the animal iGluR clades resulting from this parsimony analysis coincide with the classes established in functional studies. Like animal iGluRs, the *AtGLR* gene family also separates into clades (Clades I, II, and III) with strong node support (Bootstrap values = 95%–100%).

The division of the *AtGLR* gene family into three clades is also demonstrated by the comparison of amino acid percent identity values between genes that are in the same *AtGLR* clade (40%–57%) and genes that are from different *AtGLR* clades (21%–33%; fig. 2*a*). As

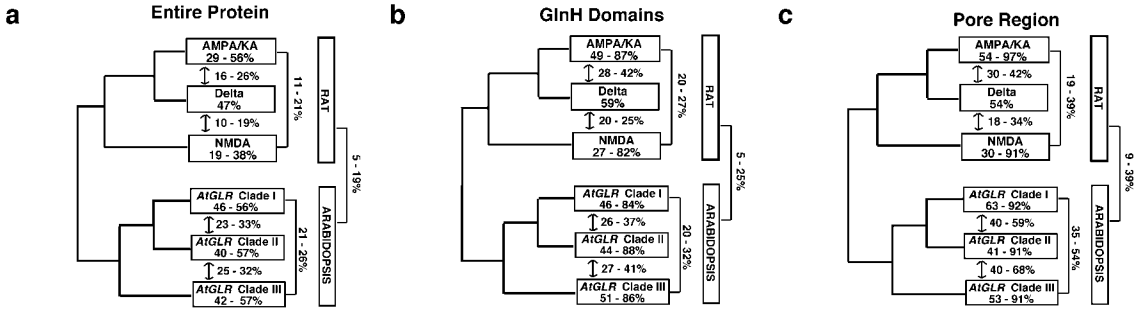

FIG. 2.—Amino acid percent identity divides *AtGLR* genes into three clades. Amino acid percent identity values between different glutamate receptor genes generated using: *a,* full-length amino acid sequences; *b,* sequences for the two glutamate-binding domains (amino acid residue numbers 406–527 and 628–737); *c,* sequences for the P and the surrounding M1 and M2 transmembrane domains (amino acid residue numbers 543–626). Amino acid numbering follows sequence of *AtGLR1.1* (GenBank accession number AF079998).

shown by the amino acid percent identity values generated by comparing entire proteins, there is a clear boundary between *AtGLR* clades based on sequence identity. As a test case, amino acid percent identity values between different rat iGluRs also support the tree topology established in our parsimony analysis, which agrees with the established animal iGluR classes. There is a clear separation between animal iGluR classes based on amino acid percent identity values, with the exception of the NMDA class. The amino acid percent identity is relatively low (19%–38%) within the NMDA class because NMDAR1 has diverged greatly from the other four NMDA genes (NR2A-D). When the percent identity values generated for the genes within the NMDA class are excluded, the percent identity values between genes that are within the same animal iGluR class are relatively high (29%–56%), whereas that between genes from different animal iGluR classes are relatively low (10%–26%).

We then repeated the analysis using amino acid residues for important functional domains (see *Materials and Methods* for boundaries of functional domains). When the amino acid percent identities were generated using only the ligand-binding domains (GlnH1 and 2) instead of the entire protein, the percent identity values increase overall, but the boundary between *AtGLR* clades remains clear-cut (fig. 2*b*). The same holds true for the animal iGluR classes as long as the percent identity values generated for the genes within the NMDA class are excluded. On the other hand, when the P and surrounding regions (M1, P, M2) were used to generate amino acid percent identities, the clear boundary between *AtGLR* clades disappeared (fig. 2*c*). Using only the P and surrounding regions, the amino acid percent identities between genes that are in the same *AtGLR* clade (41%–92%) now overlap with the amino acid percent identities between *AtGLR* genes from different clades (35%–68%), which might suggest an overlap in ion selectivity, if AtGLR proteins are indeed ion channel subunits. Interestingly, for rat iGluRs, the amino acid percent identities for the P and surrounding regions provide an observable boundary between distinct iGluR classes with different ion selectivity (as long as the percent identity values generated for the genes within the NMDA class are excluded).

In addition to establishing the three clades of *AtGLR* genes, our phylogenetic analysis also showed that clades I and II are sisters to each other (bootstrap value = 58%). To confirm the results of our analysis, we performed a second parsimony analysis using elision (Wheeler, Gatesy, and DeSalle 1995) as an alternative method to process the matrix, instead of a combination of culling and gap-coding. The topology of the resulting parsimony tree is essentially identical to the one discussed above (data not shown).

Defining Invariant Amino Acid Residues for the iGluR Gene Family

With the inclusion of all 20 *AtGLR* gene family members and the two prokaryotic iGluRs (*Synechocystis GluR0* and *Anabaena* GluR) in our analysis, we can now identify potential functionally important iGluR amino acid residues that are absolutely conserved before the divergence of plants and animals and those that are conserved even before the divergence of prokaryotes and eukaryotes. These iGluR invariant residues are presented in figure 3*a*. The present analysis is more comprehensive than a previous one (Chiu et al. 1999) because the sequences for the entire *AtGLR* gene family (16 more genes) and two prokaryotic iGluRs are now included. Another difference between this and our previous analysis is that animal iGluR sequences included in this analysis are limited to the sequences of a single animal, rat. In our previous analysis, animal iGluR sequences from other species, such as *Drosophila melanogaster*, *Caenorhabditis elegans*, pigeon, and fish were included. The exclusion of non-rat animal iGluR sequences has allowed us to directly compare the entire iGluR gene family of a single plant (*Arabidopsis*) and a single animal (rat). Such an analysis could potentially increase the number of conserved residues, but this does not appear to be the case. The only invariant residue identified in this analysis as a result of excluding non-rat animal iGluR sequences is G701 in GlnH2 (fig. 3*b;* amino acid residue numbering is based on sequence of *AtGLR1.1*), which is found in all animal iGluRs, except *Drosophila* GluR2 (M73271, data not shown). G701 is also replaced by proline and asparagine in *Synechocystis GluR0* and *Anabaena* GluR, respectively, and is there-

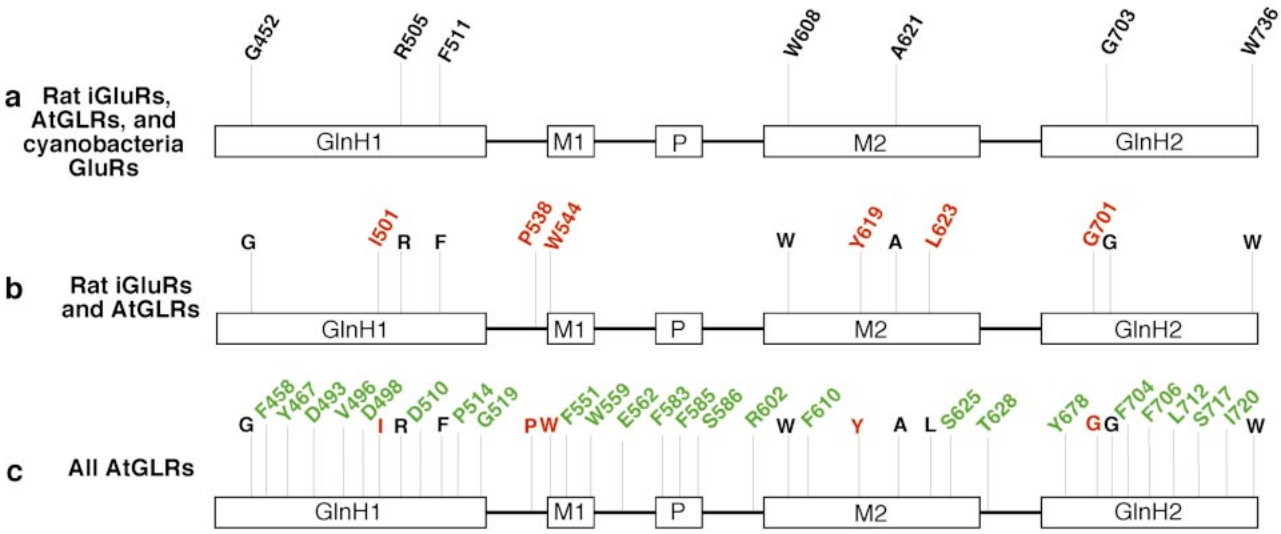

FIG. 3.—Identification of residues invariant to animal, plant, and cyanobacterial iGluRs. Diagrams showing the functional domains of the glutamate receptor protein and the locations of invariant amino acid residues between (*a*) rat iGluRs, *AtGLRs*, and cyanobacterial iGluRs; (*b*) rat iGluRs and *AtGLRs;* (*c*) all *AtGLRs.* The boxes representing the functional domains are not drawn to scale and do not reflect sequence length. Invariant residues between rat iGluRs, *AtGLRs*, and cyanobacterial iGluRs are shown in black in all three panels. Invariant residues between rat iGluRs and *AtGLR*s that are absent in cyanobacterial iGluRs are shown in red in panels b and c. Invariant residues that are specific to *AtGLR*s are shown in green in panel c. All amino acid numbers correspond to the sequence of *AtGLR1.1* (accession number AF079998).

fore conserved only in rat iGluRs and *Arabidopsis GLRs.* By comparing figure 3*a* and *b,* we identified other amino acid residues that are invariant in eukaryotes but are altered in at least one of the two prokaryotic cyanobacterial iGluRs. These residues include I501, P538, W544, Y619, and L623. Among these five residues, L623 is the only one that is not altered in both cyanobacterial genes. Whereas L623 is substituted with an alanine in *Anabaena* GluR, it is conserved in *Synechocystis GluR0.* The other four residues that are not conserved between *Synechocystis GluR0* and animal iGluRs may be directly linked to the difference in ligand specificity and ion selectivity between *Synechocystis GluR0* and animal iGluRs.

In addition to identifying invariant amino acid residues for the iGluR gene family, we have identified invariant residues specifically for the *Arabidopsis GLR* gene family (fig. 3*c*). As expected, the conserved residues unique to *AtGLR* genes span all the important functional domains defined for animal iGluRs. Future mutagenesis experiments on these residues may be useful for examining comparative differences between functional properties of animal iGluRs and *AtGLRs.*

AtGLR Genes from Different Clades Overlap in Their Organ Expression Profiles

Our parsimony analysis divides the animal iGluR gene family into three distinct clades (AMPA-KA, NMDA, and delta), with the exception of RNMDAR1 (see parsimony analysis results section). These three clades defined by parsimony analysis agree with the established classification of these distinct animal iGluR classes based on biochemical properties, ligand specificity, and ion selectivity. The *Arabidopsis GLR* gene family is also divided into three clades based on our parsimony analysis. By analogy to animal iGluRs, we hypothesize that the three *AtGLR* clades may represent biochemically or functionally distinct glutamate receptor protein classes. Whereas glutamate-gated calcium fluxes have been observed in *Arabidopsis* (Dennison and Spalding 2000), no data are currently available for the biochemical activity encoded by specific *AtGLR* genes. Thus, we have begun to address whether the three *GLR* clades in *Arabidopsis* represent functionally distinct *AtGLR*s by first looking at the mRNA expression of *AtGLR* genes from each clade.

We first examined the expression of all 20 *AtGLR* genes using organ-specific RT-PCR (in situ hybridization experiments have been attempted for *AtGLR* genes, but the level of expression is too low to be detected). Because iGluR channels in animals are homotetramers or heterotetramers assembled from proteins encoded by genes in the same functional class (Rosenmund, Stern-Bach, and Stevens 1998), it is possible that *AtGLR* genes from the same functional class (not defined yet) may encode proteins that have the ability to form heteromers in vivo. In order for different proteins to form multimers, the colocalization of their mRNAs is likely and that of their proteins is necessary. Therefore, we hypothesize that mRNAs of *AtGLR* genes that are in the same clade may be present in the same organs. Distinct clade-specific expression patterns may be the first indication that phylogenetically defined *AtGLR* clades represent functional *AtGLR* classes. This hypothesis is

based on the assumption that class-specific expression is an important mechanism that regulates *AtGLR* heteromeric protein coassembly.

Our RT-PCR analysis established that all 20 members of the *AtGLR* gene family are expressed genes (fig. 4*a* and *b*). This finding addresses a previous concern that several *AtGLR*s might be pseudogenes because many of the *AtGLR* gene family members are duplicated genes, existing in tandem on the chromosomes (note same BAC clone numbers in table 1). Not only are they all expressed, potential mRNA splice variants, represented by additional bands on the gel, can be observed for two *AtGLR* genes, namely *GLR 2.5* and *GLR 3.4*. Whereas cDNA clones for *GLR2.5* are not available at present, splice variants for *GLR3.4* have already been isolated (GenBank accession numbers AF167355 and AY072070) and thus confirmed our observation.

More importantly, our RT-PCR analysis of all 20 *AtGLR* genes show that although *AtGLR* genes from the same clade show similar organ expression profiles, there is an overlap in expression patterns between genes that are in different clades. Therefore, we conclude that *AtGLR* genes from different clades do not show distinct organ expression profiles. *AtGLR* clade I genes are detected in all organs examined (leaves, roots, flowers, and siliques), with highest expression in roots (fig. 4*a* and *b*). For *AtGLR* clade-II genes, the majority is uniquely expressed in roots (*GLR2.1, GLR2.2, GLR2.3, GLR2.6,* and *GLR2.9*). Exceptions include *GLR2.4,* which is detected in siliques in addition to roots; *GLR 2.5,* which is detected in all organs tested; and *GLR2.7* and *GLR2.8* whose transcripts are detectable in all organs, except flowers. *AtGLR* clade-III genes, like clade-I genes, are detected in all organs tested, with roots and siliques showing highest expression for the majority (except *GLR3.2* and *GLR3.7*).

The RT-PCR mRNA expression data of *AtGLR* genes was mapped onto the *AtGLR* portion of the iGluR parsimony tree to illustrate the relationship between *AtGLR* gene evolution and organ-specific expression (fig. 5). As shown in figure 5*b,* root expression is detected in all *AtGLR* genes and is inferred to be an expression character state ancestral to the entire *AtGLR* gene family. Similarly, mRNA expression in leaves appears to be the ancestral character state of the *AtGLR* gene family (fig. 5*a*). However, multiple independent changes occurred only in *AtGLR* clade II, resulting in the loss of leaf expression in six of the nine clade-II genes (*GLR2.1, GLR2.2, GLR2.3, GLR2.4, GLR2.6,* and *GLR2.9*). Lastly, mRNA expression in reproductive organs (flowers and siliques in *Arabidopsis*) also appears to be ancestral to the *AtGLR* gene family (fig. 5*c* and *d*). However, once again, multiple independent changes occurred specifically in *AtGLR* clade II resulting in (1) a loss of flower expression in all but one of the clade-II genes (*GLR2.5*), and (2) a loss of silique expression in five clade-II genes (*GLR2.1, GLR2.2, GLR2.3, GLR2.6,* and *GLR2.9*).

Although our RT-PCR results show that *AtGLR* genes from different clades overlap in their organ expression profiles, we cannot negate the possibility that *AtGLR* genes that are expressed in the same organs may be present in different cell types. To determine the specific cell types in the organs in which the different *AtGLR* genes are expressed, we next examined transgenic plants that have been transformed with constructs in which the putative promoter of a representative for each of the three *AtGLR* clades was fused with the reporter gene GUS (Jefferson 1989). By this method, we examined the cell-type–specific expression pattern of one gene from each *AtGLR* clade (*GLR1.1, GLR2.1, GLR3.1*). GUS expression in *GLR1.1* promoter::GUS transgenic plants first appears in 7-day-old plants in stipules and the collette region (root-shoot junction; fig. 6*a* and *d*). Expression is later detected (3- to 4-week-old plants) in all cell types of lateral roots and at the margin of mature leaves, in addition to all the organs mentioned above (fig. 7*a, d,* and *g*). *GLR1.1* expression in reproductive organs, flowers, and siliques, is not readily detectable (fig. 8*a* and *e*). According to promoter::GUS expression, *GLR2.1* (a clade-II gene) has a shoot expression pattern in young seedlings (starting from 5 days old) that is similar to *GLR1.1* in that they are both expressed in stipules (fig. 6*e* and *h*). This overlap in expression pattern supports the sister relationship of clades I and II established in our parsimony tree (fig. 1). In addition, *GLR2.1* expression is detected in the radical immediately after emergence and is clearly observable in the root starting when the seedlings are 3 days old (fig. 6*b*). *GLR2.1* is expressed in all cell types of the root (including the root hairs), except at the root apex. Root expression of *GLR2.1* is also present in more mature seedlings (fig. 7*e, h,* and *k*). Although *GLR2.1* expression is not readily detectable in flowers and siliques (fig. 8*b* and *f*), slight and transient expression has been observed in anthers and young ovules (data not shown). Whereas the shoot expression of *GLR1.1* and *GLR2.1* overlap significantly, the shoot expression pattern of *GLR3.1,* a clade-III gene, is very different (figs. 6*f,* 6*i,* and 7*c*). *GLR3.1* expression is first visible in the vasculature of the cotyledons in 5-day-old young seedlings (fig. 6*f*), and its expression is maintained in the leaf and cotyledon vascular tissues as the seedlings mature (fig. 7*c*). *GLR3.1* expression is later found in all organs, including flowers (fig. 8*c* and *d*), siliques (fig. 8*g* and *h*), and roots (fig. 7*f, i,* and *l*), specifically in the vasculature and in cell types associated with the vasculature and in particular at contact sites or sites of vascular transfer into plant organs, such as the funiculus and the developing seed. The expression pattern of a second clade-III gene, *GLR3.2,* is identical to that of *GLR3.1* based on promoter::GUS assays (data not shown).

Discussion

Arabidopsis expresses a large family of *GLR* genes that may be involved in multiple functions, such as light signal transduction (Lam et al. 1998; Brenner et al. 2000) and calcium homeostasis (Kim et al. 2001). Animal iGluR genes are traditionally divided into functional classes (AMPA-KA, NMDA, and Delta) based on ligand specificity as well as electrophysiological prop-

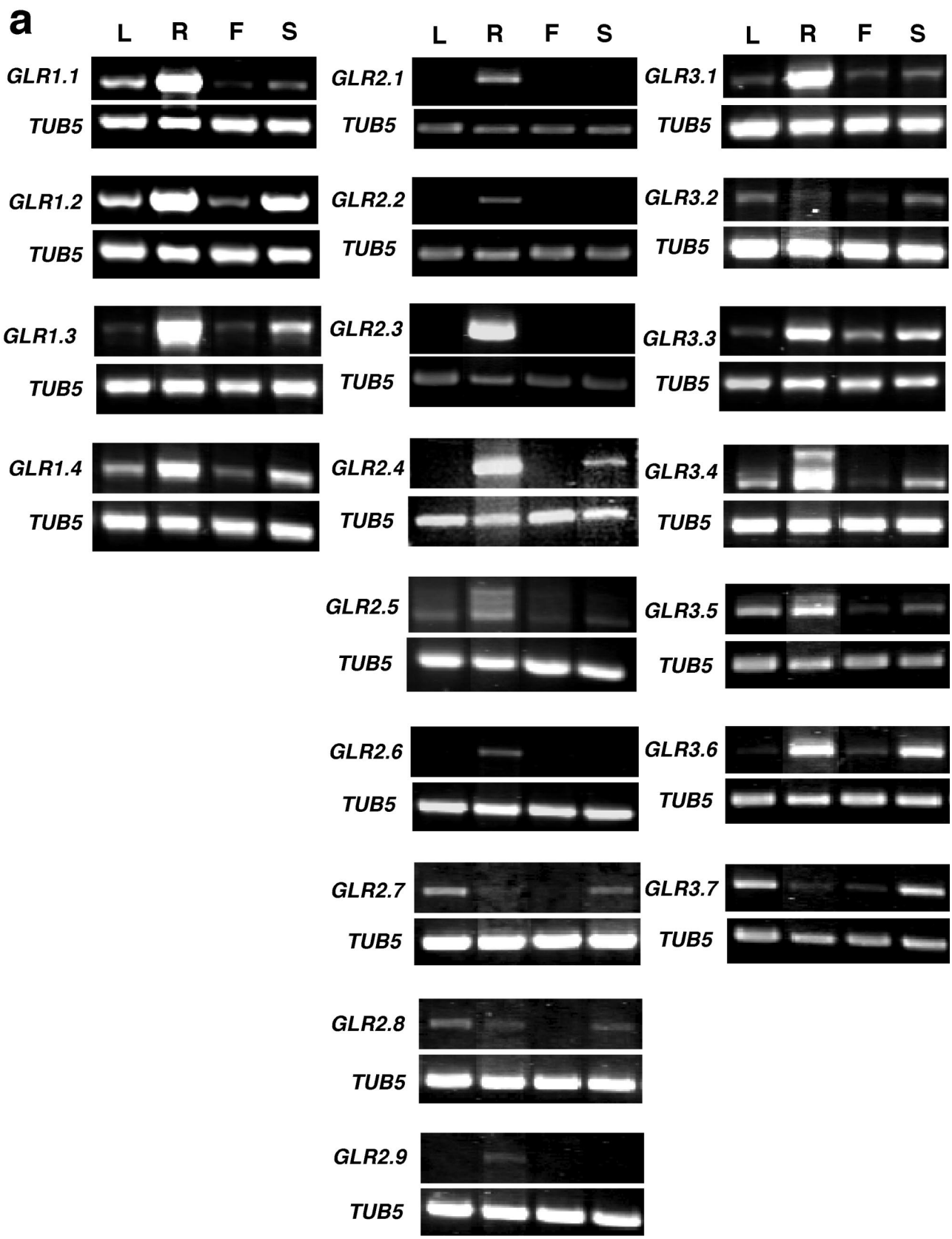

FIG. 4.—RT-PCR mRNA expression analysis reveals overlap in organ expression of *AtGLR* genes from different clades. *a*, RT-PCR analysis of the *AtGLR* gene family using mRNA generated from four different *Arabidopsis* organs: leaf (L), root (R), flower (F), and silique (S). The results shown here represent PCR reactions with 30 amplification cycles. *b*, Graphic representation of the RT-PCR data of the *AtGLR* gene family. Each graph shows the relative intensity of the four *AtGLR* bands in the different organ lanes on the agarose gel shown in figure 4*a* normalized against *TUB5* (accession number M84702).

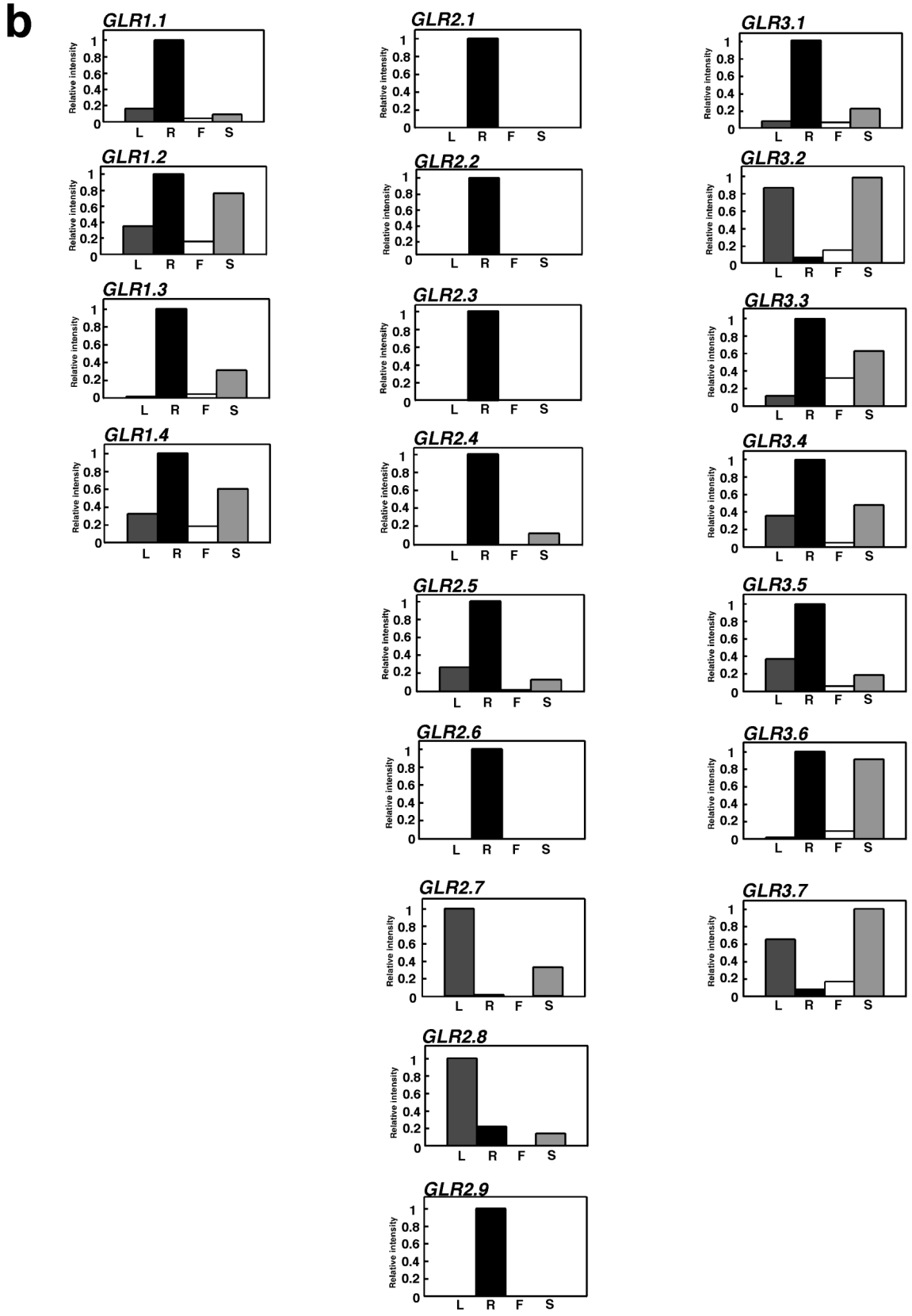

FIG. 4.—Continued.

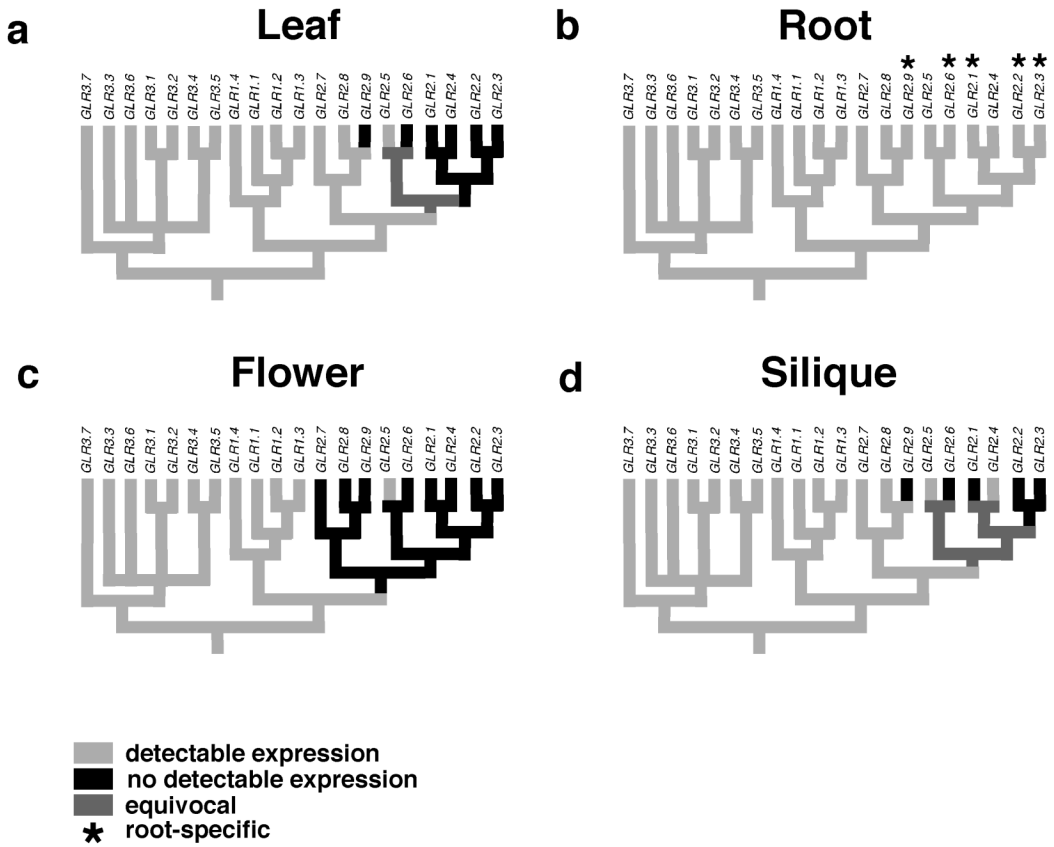

FIG. 5.—Character-mapping of RT-PCR expression data reveals ancestral and derived states of three *AtGLR* clades. *AtGLR* mRNA expression in (*a*) leaves, (*b*) roots, (*c*) flowers, and (*d*) siliques as detected by RT-PCR (fig. 4) are converted into binary format (0 = undetected and 1 = detected) and mapped onto the phylogenetic tree generated from our parsimony analysis (only the *AtGLR* part of the tree is shown in this figure). Light gray: expression is detected by RT-PCR; Black: expression is not detected; Dark gray: equivocal; Asterisks: root-specific *AtGLR* genes as detected by RT-PCR performed on 8-week-old *Arabidopsis* plants (see *Materials and Methods* for growth conditions).

erties. Genes from different animal iGluR classes encode proteins that perform different functions in the animal nervous system (Sprengel and Seeburg 1995). In our phylogenetic analyses in this and our previous paper (Chiu et al. 1999), the animal iGluR clade division we obtained using parsimony and Neighbor-Joining analyses of animal iGluR amino acid sequences agreed with the widely accepted class division based on biochemistry (except with the placement of RNMDAR1 in the current analysis). The well-studied animal iGluR genes thus represent a test case for linking phylogenetic analysis and physiological function. Our results showed that phylogenetic analysis using sequence data is a useful tool for dividing a large gene family into potential functionally distinct classes, when anatomical, biochemical, and other functional data are not yet available. By applying this approach to the *GLR* gene family in *Arabidopsis*, the 20 *AtGLR* genes can be divided into three clades (also shown in Lacombe et al. 2001). The functional significance of this classification sets the stage for clade-specific testing of function using expression, biochemical, electrophysiological, and in planta physiological data.

In this paper, we have begun to address whether the three *AtGLR* clades represent functionally distinct protein classes by examining the mRNA expression patterns of all 20 *AtGLR* genes and promoter activities of representatives of each clade. Because iGluRs exist as tetramers in animals (Rosenmund, Stern-Bach, and Stevens 1998), determining the organ and cell-type–specific expression patterns of each *AtGLR* gene will also enable us to predict which gene products could potentially interact in vivo. Heteromultimeric coassembly can profoundly influence protein function (Sprengel and Seeburg 1995; Nitabach et al. 2001). By using RT-PCR, we obtained an mRNA expression profile for the entire *AtGLR* gene family in four different organ types, namely leaves, roots, flowers, and siliques. As a cautionary note because the phylogenetic analysis was performed using the coding sequences of the genes although the mRNA expression patterns are controlled by regulatory noncoding sequences, they might not have direct correlation. Our RT-PCR results show that there is overlap in organ expression profiles between genes from different *AtGLR* clades. Whereas genes from *AtGLR* clades I and III are expressed more ubiquitously, clade-II genes are largely root specific (fig. 4*a* and *b*). We conclude that the three *AtGLR* clades are not distinct based on organ-level expression patterns. Thus, the three *AtGLR* clades may contain genes with overlapping functions in vivo. Interestingly, five of the nine clade-II genes are root-specific in 8-week-old *Arabidopsis* plants, and may potentially

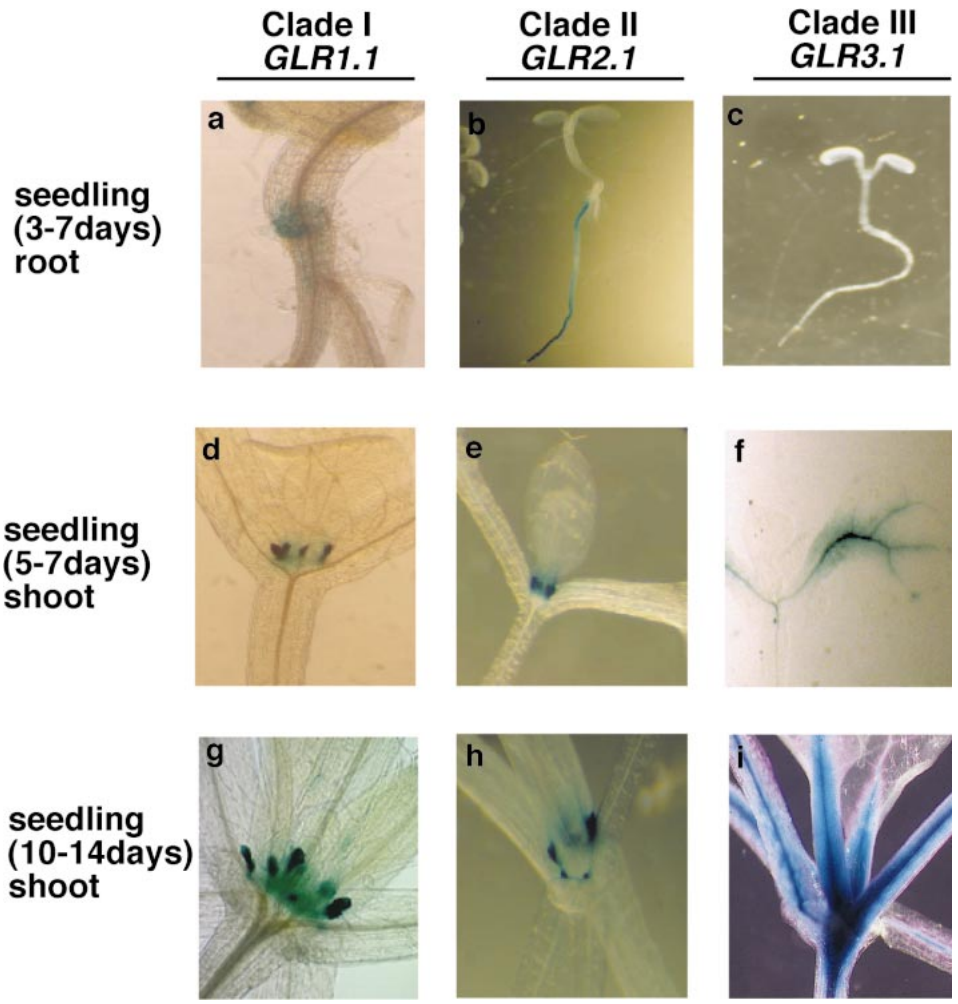

FIG. 6.—Three *AtGLR* clades exhibit cell-specific expression patterns in young seedlings. Anatomical and developmental expression of *GLR 1.1, GLR2.1,* and *GLR3.1* as seen in transgenic *Arabidopsis* seedlings harboring promoter::GUS fusions. Early after germination (3–5 days), *GLR2.1* shows strong expression in all cell types of the root except the apex (b). Similarly, weak expression of *GLR1.1* can be seen in the root (a). *GLR3.1* expression cannot be detected at this early stage (c). In more mature seedlings (5–14 days), *GLR1.1* and *GLR2.1* are expressed in stipules (d, e, g, and h), whereas *GLR3.1* expression is found in the vasculature and in cells neighboring the vasculature (f and i).

represent a functional class. Additional data from physiological studies are necessary to verify this proposition because genes that have the same expression pattern may not share the same function.

By mapping the RT-PCR mRNA expression data onto the iGluR parsimony tree (fig. 5), we are able to make phylogenetic inferences relating organ-specific expression and *AtGLR* gene evolution. Ubiquitous expression in leaves, roots, and reproductive organs appears to be the ancestral character states of the *AtGLR* gene family, but multiple independent character state changes occurred specifically in clade-II lineages, resulting in the loss of expression in leaves, flowers, and siliques in some of the clade-II genes. It is interesting to note that all changes in expression character state occurred specifically in clade-II genes. Perhaps the gene duplication event leading to the divergence of *GLR* clades I and II provided the opportunity for one of these two *AtGLR* clades (clade II) to diverge in terms of function, as indicated by the change in expression patterns.

To test the possibility that mRNA of different *AtGLR* genes that are detected in the same organs may be present in different cell types, we examined transgenic plants harboring promoter::reporter gene constructs to determine the cell types in the organs in which different *AtGLR* genes are expressed. For this, we analyzed one representative gene from each of the three *AtGLR* clades. Although genes from both clades I and III are expressed in all organs tested as shown by RT-PCR, promoter::GUS analyses of representative clade-I (*GLR1.1*) and clade-III (*GLR3.1*) genes indicate that genes from clades I and III may be expressed in different cell types. The expression of *GLR3.1* (a clade-III gene) is limited to the vasculature in all organs tested (as is a second clade-III gene, *GLR3.2* [data not shown]). In contrast, *GLR1.1* (a clade-I gene) is not highly expressed in the vasculature but is expressed instead in distinct cell types (figs. 6–8). Future experiments (protein fusions and antibody staining) are necessary to increase the resolution of this analysis. Cur-

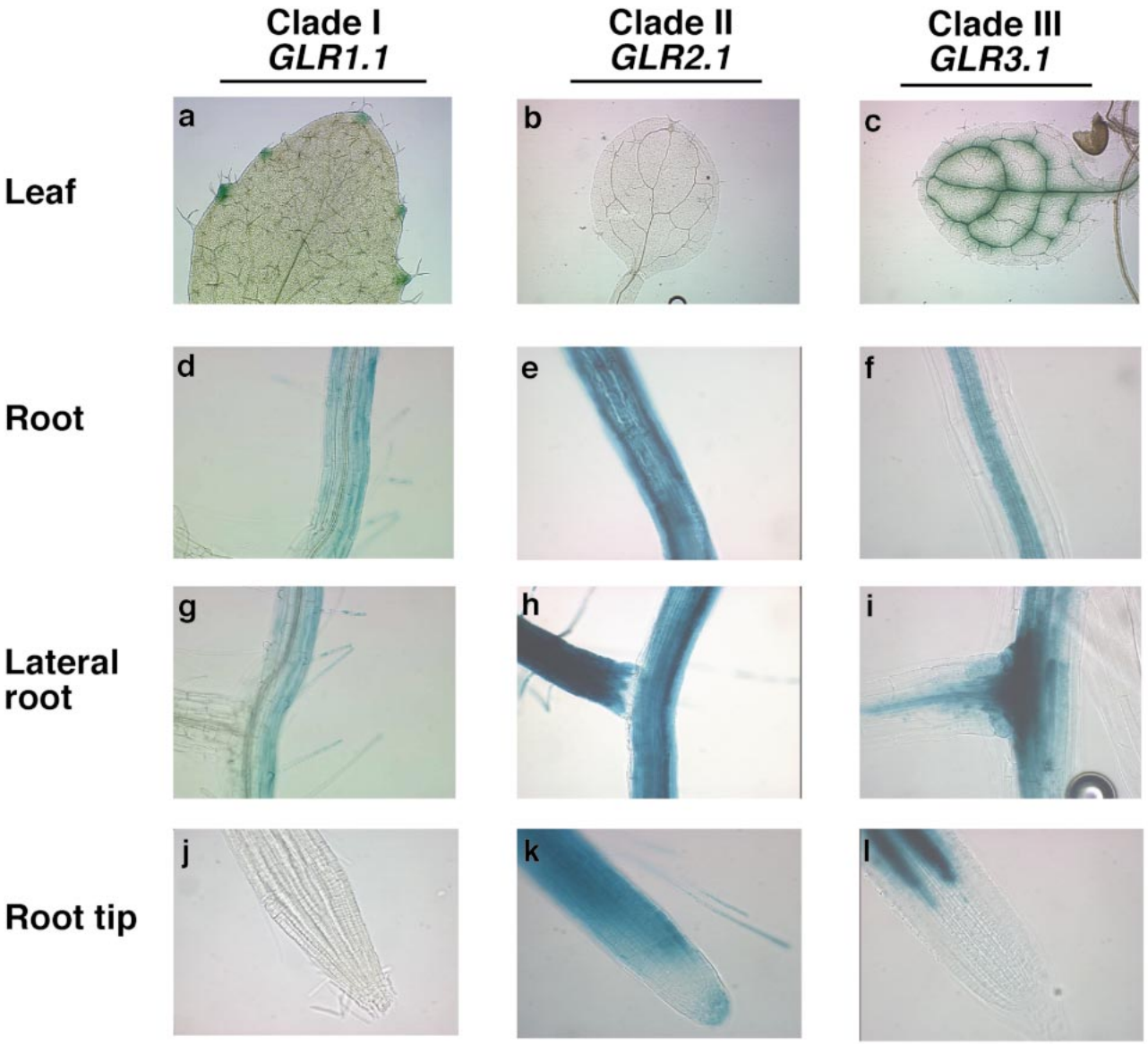

FIG. 7.—Cell-type–specific expression patterns of three *AtGLR* clades in 3- to 4-week-old plants. Ultrastructural expression of *AtGLR* genes in vegetative organs using GUS as a reporter gene. Compound micrographs of the leaf show GUS expression in the leaf margins of *GLR1.1* (a). *GLR2.1* is not detected in the leaf (b). *GLR3.1* expression is seen in the vasculature and in cell types associated with the vasculature in the leaf (c), the root (f), and emerging lateral roots (i). Expression of *GLR1.1* is detected in most layers of the root and the root hairs (d and g). *GLR2.1* is strongly expressed throughout the root (e) and in root hairs (h and k) and emerging lateral roots (h). Expression of all three genes is missing in the apex of the root (j, k, and l).

rently, we can only conclude that *GLR1.1* and *GLR3.1* are expressed in distinct cell types, although they are expressed in the same organs. More experiments are needed to examine if this holds true for all other genes from clades I and III. If *AtGLR* genes from clades I and III are indeed expressed in distinct cell types, it will be more likely for clades I and III to represent distinct functional classes.

This paper focuses on class-specific mRNA expression as being an important mechanism for the regulation of heteromeric coassembly between proteins within the same functional class. It is important to note that other regulatory mechanisms might be involved as well. Moreover, it is also important to keep in mind that genes that have the same expression pattern may not have the same function. In addition, it is possible that unlike the case of animal iGluRs, in which phylogenetically defined clades coincide with true functional classes, phylogenetically defined *AtGLR* clades may not represent functional *AtGLR* classes. Future biochemical and physiological experiments will help to confirm the validity of the root-specific class of five clade-II genes, as well as to define other functional classes.

Both RT-PCR and promoter::GUS analyses have shown that all *AtGLR* genes are strongly expressed in roots as a general rule. If *AtGLR* genes indeed function as ion channels, their high expression in roots may potentially be important for regulating ion uptake from soil. It is interesting to note that glutamate-gated calcium fluxes across the plasma membrane have been observed

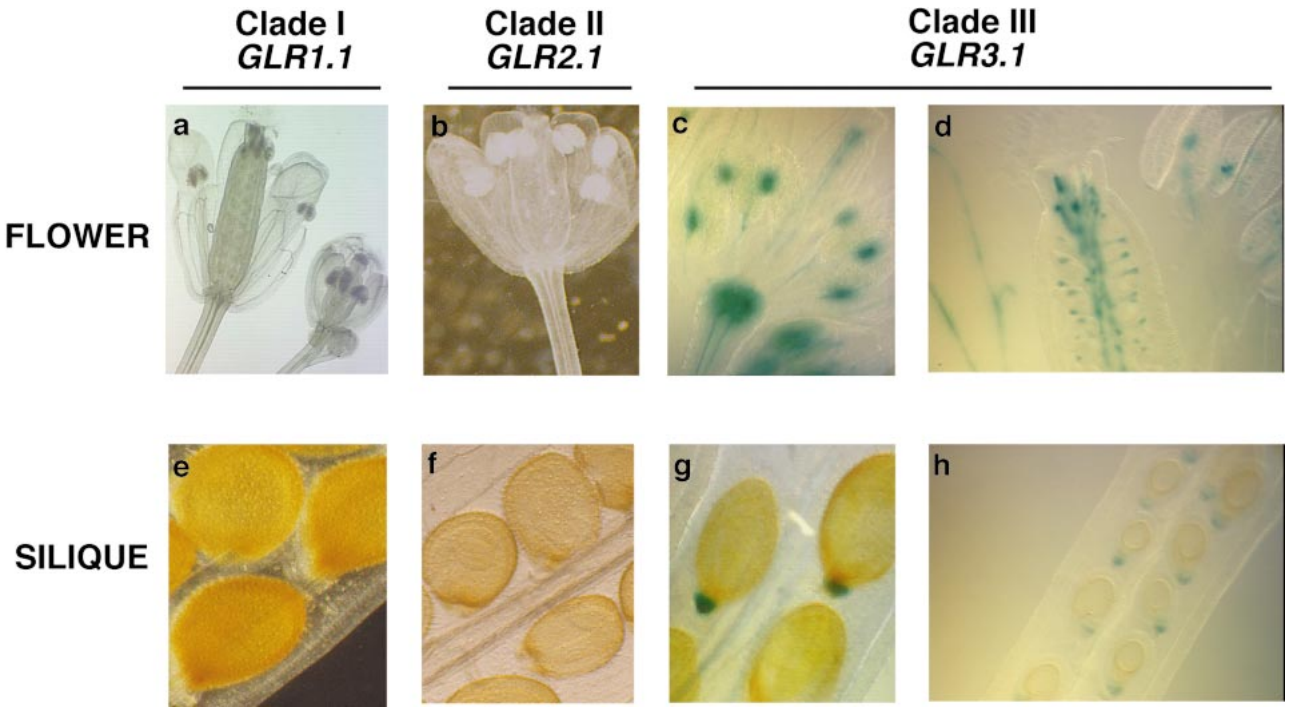

FIG. 8.—*AtGLR* genes show distinct expression patterns in flowers and reproductive organs. Expression of *AtGLR* genes in reproductive organs using GUS as a reporter gene. Expression of *GLR1.1* (a and e) and *GLR2.1* (b and f) are not readily detectable in reproductive organs (slight and transient GUS expression has been observed in anthers and ovules for *GLR2.1*). *GLR3.1* GUS expression is found in the vasculature and neighboring organs, including the petal, the ovary, and the filament (c, d, g, and h). In the vasculature, *GLR3.1* expression is particularly strong at contact sites between the anther and the filament, the receptacle and the flower, and the funiculus and developing seed (c, d, g, and h).

in *Arabidopsis* roots (Dennison and Spalding 2000). Future experiments using mutagenesis and reverse genetics approaches will be necessary to investigate the possible link between this phenomenon and specific *AtGLR* genes. In addition to the strong root expression for the majority of *AtGLR* genes, which suggest a role in ion uptake from soil, the strong vascular expression of clade-III genes (*GLR3.1* and *GLR3.2*) suggests that *AtGLR* genes may potentially play a role in regulating amino acid transport in the phloem. It has been shown that glutamate is transported in the phloem (Lam et al. 1995) as well as to siliques (H. M. Lam, personal communication) in *Arabidopsis*, as in other higher plants.

In addition to testing the hypothesis that the division of the *AtGLR* gene family into three clades has functional significance, we have identified potential functionally important amino acid residues that are invariant in both eukaryotic and prokaryotic iGluR genes in conserved functional domains (fig. 3*a*). Mutagenesis experiments have been conducted for most of these residues in animal iGluRs or bacterial periplasmic-binding proteins, and they have indeed been shown to be functionally important. R505 has been shown to interact with all known animal iGluR agonists and is therefore crucial for the ligand-binding capability of iGluR proteins (Armstrong et al. 1998). F511, G703, and W736 have been shown to be important to the structural integrity of iGluR and bacterial periplasmic-binding proteins (Chen et al. 1999). A621 is part of the YTANLAAL motif that is believed to shape iGluR ligand-gated channel kinetics (Kohda, Wang, and Yuzaki 2000). This analysis shows that invariant residues identified in this manner are almost always functionally important residues.

We have also identified amino acid residues in functional domains that are conserved in eukaryotic iGluRs (rat and *Arabidopsis*) but not in the functional prokaryotic *Synechocystis GluR0* (fig. 3*b*). *Anabaena* GluR has not been shown to be a functional ion channel to date and is therefore excluded in this discussion. The sequence divergence at these particular amino acid residues may play a part in creating the observed differences in properties between eukaryotic iGluR and *Synechocystis GluR0. Synechocystis GluR0* has a different ligand selectivity profile when compared with animal iGluRs (Chen et al. 1999). It does not bind subtype-specific eukaryotic iGluR agonists, such as kainate, NMDA, and AMPA, but does bind a wider variety of amino acids in addition to L-glutamate, such as L-glutamine and L-serine, as compared with animal iGluRs. In addition, *Synechocystis GluR0* encodes a protein that forms a homomeric channel permeable only to potassium ions as tested in heterologous systems (Chen et al. 1999), whereas animal iGluR channels are permeable to sodium, potassium, as well as calcium depending on their subunit composition (Sprengel and Seeburg 1995). It is possible that the amino acid residues identified here might be involved in the difference in ligand selectivity and channel kinetics between eukaryotic iGluRs and *Synechocystis GluR0.* In fact, one of the residues, Y619, has been shown to be involved in influencing animal iGluR channel kinetics (Kohda, Wang, and Yuzaki 2000).

Finally, we have identified amino acid residues in important functional domains that are invariant in all members of the *AtGLR* gene family. These include amino acid residues that are: (1) conserved in both eukaryotes (rat and *Arabidopsis*) and prokaryotes (fig. 3*a*), (2) those that are conserved only in eukaryotes (rat and *Arabidopsis*, fig. 3*b*), and (3) those that are only conserved in the *AtGLR* gene family (fig. 3*c*). In addition, *AtGLR* invariant residues will naturally include residues that are conserved in prokaryotic iGluRs but not in all rat iGluRs (not illustrated in fig. 3). Results of mutagenesis experiments for residues that are conserved in both eukaryotes and prokaryotes were discussed earlier, and they represent obvious targets in future mutagenesis experiments that can help to elucidate the function of *GLR* genes in plants. Mutagenesis experiments for *AtGLR* invariant residues that are conserved in some rat iGluR genes as well as in prokaryotic iGluRs have been performed in animal iGluR genes and bacterial periplasmic amino acid–binding proteins, and results showed that they are functionally important. For example, P514 has been found to be critical to the structural integrity of the bacterial amino acid–binding protein (Chen et al. 1999). The F583A mutation in animal AMPA subunit GluR3 (F605 in GluR3) produces a killer subunit that has a dominant negative effect when expressed with wild-type subunits (Wo and Oswald 1995). Mutagenesis studies on the *AtGLR* invariant residues identified in this study will help us to understand the electrophysiological properties of these *AtGLR* channels if these are indeed channels, as well as their function in planta.

Conclusions

We have conducted the first comprehensive mRNA expression analysis of the *Arabidopsis GLR* gene family. This analysis is the first step to examining whether the three *AtGLR* clades, defined by parsimony analysis, represent functionally distinct classes. On the basis of our RT-PCR analysis, we conclude that genes from the three *AtGLR* clades do not show distinct organ expression profiles. However, five of the nine clade-II genes are root-specific in 8-week-old *Arabidopsis* plants and may represent a functional class. Future physiological and biochemical experiments are necessary to verify this suggestion. In examining cell-type–specific expression, we point out that genes that are expressed in the same organs may be expressed in different cell types within the organs. As a result, the question of whether clades I and III indeed have distinct expression patterns still remains open.

It is important to understand that genes that show the same expression pattern may not necessarily have the same in vivo function. This is why future expression, biochemical, electrophysiological, and in planta physiological experiments are necessary to increase the resolution of the expression analyses presented in this paper and continue to classify *AtGLR* genes into distinct functional classes as well as elucidating their function in vivo. It is important to note that organ or cell-type–specific expression may only be one of the mechanisms to restrict protein coassembly within each *GLR* functional class, assuming GLR proteins are capable of forming heteromultimers. The phylogeny and expression studies presented here will help to drive future biochemical and electrophysiological analyses. *AtGLR* genes that are coexpressed in specific cell or organ types will be targeted for tests of coassembly and cofunction in vivo.

Supplementary Material

The processed data matrix (after culling and gap-coding) used for our parsimony analysis is available online at the MBE website.

Acknowledgments

This work was supported by NIH grant GM32877 to Gloria M. Coruzzi. R.D. is supported in part by the Lewis B. and Dorothy Cullman Program for molecular systematics and the Ambrose Monell collection for molecular and microbial research. E.D.B. is supported by an NIH postdoctoral fellowship. M.N.N. is supported by an NIH NRSA individual fellowship.

LITERATURE CITED

AGI. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**:796–815.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

ARAKI, K., H. MEGURO, E. KUSHIYA, C. TAKAYAMA, Y. INOUE, and M. MISHINA. 1993. Selective expression of the glutamate receptor channel delta 2 subunit in cerebellar Purkinje cells. Biochem. Biophys. Res. Commun. **197**:1267–1276.

ARMSTRONG, N., Y. SUN, G. Q. CHEN, and E. GOUAUX. 1998. Structure of a glutamate-receptor ligand-binding core in complex with kainate. Nature **395**:913–917.

BRENNER, E. D., N. MARTINEZ-BARBOZA, A. P. CLARK, Q. S. LIANG, D. W. STEVENSON, and G. M. CORUZZI. 2000. Arabidopsis mutants resistant to S(+)-β-Methyl-α, β-Diaminopropionic acid, a cycad-derived glutamate receptor agonist. Plant Physiol. **124**:1615–1624.

CHEN, G. Q., C. CUI, M. L. MAYER, and E. GOUAUX. 1999. Functional characterization of a potassium-selective prokaryotic glutamate receptor. Nature **402**:817–821.

CHIU, J., R. DESALLE, H. M. LAM, L. MEISEL, and G. M. CORUZZI. 1999. Molecular evolution of glutamate receptors: a primitive signaling mechanism that existed before plants and animals diverged. Mol. Biol. Evol. **16**:826–838.

CLOUGH, S. J., and A. F. BENT. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J. **16**:735–743.

DENNISON, K. L., and E. P. SPALDING. 2000. Glutamate-gated calcium fluxes in Arabidopsis. Plant Physiol. **124**:1511–1514.

DESALLE, R., and A. BROWER. 1997. Process partitions, congruence and the independence of characters: inferring relationships among closely related Hawaiian *Drosophila* from multiple gene regions. Syst. Biol. **46**:752–765.

GATESY, J., R. DESALLE, and W. C. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. **2**:152–157.

HOLLMANN, M., and S. F. HEINEMANN. 1994. Cloned glutamate receptor. Annu. Rev. Neurosci. **17**:31–108.

JEFFERSON, R. A. 1989. The GUS reporter gene system. Nature **342**:837–838.

JEFFERSON, R. A., T. A. KAVANAUGH, and M. W. BEVAN. 1987. GUS fusions: glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6**:3901–3907.

KIM, S. A., J. M. KWAK, S. K. JAE, M. H. WANG, and H. G. NAM. 2001. Overexpression of the *AtGluR2* gene encoding an Arabidopsis homolog of mammalian glutamate receptors impairs calcium utilization and sensitivity to ionic stress in transgenic plants. Plant Cell Physiol. **42**:74–84.

KOHDA, K., Y. WANG, and M. YUZAKI. 2000. Mutation of a glutamate receptor motif reveals its role in gating and δ2 receptor channel properties. Nat. Neurosci. **3**:315–322.

LACOMBE, B., D. BECKER, R. HEDRICH et al. (14 co-authors). 2001. The identity of plant glutamate receptor. Science. **292**:1486–1487.

LAM, H. M., J. CHIU, M. H. HSIEH, L. MEISEL, I. C. OLIVEIRA, M. SHIN, and G. M. CORUZZI. 1998. Glutamate receptor genes in plants. Nature **396**:125–126.

LAM, H. M., K. COSCHIGANO, C. SCHULTZ, R. MELO-OLIVEIRA, G. TJADEN, I. OLIVEIRA, N. NGAI, M. H. HSIEH, and G. M. CORUZZI. 1995. Use of Arabidopsis mutants and genes to study amide amino acid biosynthesis. Plant Cell **7**:887–898.

LOMELI, H., R. SPRENGEL, D. J. LAURIE, G. KÖHR, A. HERB, P. H. SEEBURG, and W. WISDEN. 1993. The rat delta-1 and delta-2 subunits extend the excitatory amino acid receptor family. FEBS Lett. **315**:318–322.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.

NGAI, N., F. Y. TSAI, and G. M. CORUZZI. 1997. Light-induced transcriptional repression of the pea AS1 gene: identification of *cis*-elements and transfactors. Plant J. **12**:1021–1034.

NITABACH, M. N., D. A. LLAMAS, R. C. ARANEDA, J. L. INTILE, I. J. THOMPSON, Y. I. ZHOU, and T. C. HOLMES. 2001. A mechanism for combinatorial regulation of electrical activity: potassium channel subunits capable of functioning as Src homology 3-dependent adaptors. Proc. Natl. Acad. Sci. USA **98**:705–710.

PAAS, Y. 1998. The macro- and microarchitectures of the ligand-binding domain of glutamate receptors. Trends Neurosci. **21**:117–125.

RAMEAU, G. A., Y. AKANEYA, L. Y. CHIU, and E. B. ZIFF. 2000. Role of NMDA receptor functional domains in excitatory cell death. Neuropharmacology **39**:2255–2266.

ROSENMUND, C., Y. STERN-BACH, and C. F. STEVENS. 1998. The tetrameric structure of a glutamate receptor channel. Science **280**:1596–1599.

ROST, B. 1996. PHD: predicting one-dimensional protein structure by profile based neural networks. Methods Enzymol. **266**:525–539.

SPRENGEL, R., and P. H. SEEBURG. 1995. Ionotropic glutamate receptors. Pp. 213–263 *in* R. A. NORTH, ed. Handbook of receptors and channels: ligand- and voltage-gated ion channels. CRC Press, Boca Raton, Fla.

STERN-BACH, Y., B. BETTLER, M. HARTLEY, P. O. SHEPPARD, P. J. O'HARA, and S. F. HEINEMANN. 1994. Agonist selectivity of glutamate receptor is specified by two domains structurally related to bacterial amino acid–binding proteins. Neuron **13**:1345–1357.

SUTCLIFFE, M. J., Z. G. WO, and R. E. OSWALD. 1996. Three-dimensional models of non-NMDA glutamate receptors. Biophys. J. **70**:1575–1589.

SWOFFORD, D. L. 1998. PAUP*. Test Version 4d64. Laboratory of molecular systematics, Smithsonian Institution, Washington, D.C., Sinauer, Sunderland, Mass.

TURANO, F. J., G. R. PANTA, M. W. ALLARD, and P. VAN BERKUM. 2001. The putative glutamate receptors from plants are related to two superfamilies of animal neurotransmitter receptors via distinct evolutionary mechanisms. Mol. Biol. Evol. **18**:1417–1420.

WHEELER, W. C., J. GATESY, and R. DESALLE. 1995. Elision: a method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. Mol. Phylogenet. Evol. **4**:1–9.

WO, Z. G., and R. E. OSWALD. 1995. Unraveling the modular design of glutamate-gated ion channels. Trends Neurosci. **18**:161–168.

WOLLMUTH, L. P., T. KUNER, C. JATZE, P. H. SEEBURG, N. HEINTZ, and J. ZUO. 2000. The Lurcher mutation identifies δ2 as an AMPA/kainate receptor–like channel that is potentiated by Ca^{2+}. J. Neurosci. **20**:5973–5980.

ELIZABETH KELLOGG, reviewing editor

Accepted February 11, 2002