ORIGINAL ARTICLE

The Ammonium Channel NOD26 is the Evolutionary Innovation that Drives the Emergence, Consolidation, and Dissemination of Nitrogen-Fixing Symbiosis in Angiosperms

Romina Frare[1,2] · Nicolás Ayub[1,2] · Karina Alleva[1,3] · Gabriela Soto[1,2]

Abstract

Increasing evidence indicates that N-fixing symbiosis has evolved several times in the N-fixing clade of angiosperms and that this evolution is driven by a single evolutionary innovation. However, the genetics of this ancestral predisposition to N-fixing symbiosis remains unclear. A natural candidate for such molecular innovation is the ammonium channel NOD26, the main protein component of the symbiosome membrane, which facilitates the plant uptake of the nitrogen fixed by symbiotic bacteria. Here, in concordance with the emergence of N-fixing symbiosis in angiosperms but not in ancestral plants, phylogenetic analysis showed that NOD26 belongs to an angiosperm-exclusive subgroup of aquaporins. Integrated genomic, phylogenetic, and gene expression analyses supported NOD26 occurrence in the N-fixing clade, the increase in the NOD26 copy number by block and tandem duplications in legumes, and the low-copy number or even the loss of NOD26 in non-legume species of the N-fixing clade, which correlated with the possibility to lose N-fixing symbiosis in legume and non-legume lineages. Metabolic reconstructions showed that retention of NOD26 in N-fixing precursor could represent an adaptive mechanism to bypass energy crisis during anaerobic stress by ammonium detoxification. Finally, we discuss the potential use of NOD26 to transfer N-fixation to non-N-fixing crops as cereals.

Keywords Ammonium · Channel · Preadaptation · Anaerobic · Symbiosis · Evolution

Introduction

Several plant species belonging to the N-fixing clade, a monophyletic group of angiosperms, are able to establish N-fixing symbiosis in root-derived nodule cells with different genera of the bacteria domain. Both the origin and the complex diversification of N-fixing symbiosis have puzzled biologists for a long time and are still issues under debate (Doyle 2011; Geurts et al. 2016; Griesmann et al. 2018). The dispersed occurrence of N-fixing symbiosis in the N-fixing clade can be explained by the non-homologous origins of

N-fixing symbiosis (Svistoonoff et al. 2013; Werner et al. 2014). This point of view is coherent with the predisposition hypothesis, which states that the common ancestor of the N-fixing clade evolved a predisposition towards the evolution of N-fixing symbiosis (Soltis et al. 1995). Under this paradigm, which is in agreement with the traditional concept of deep homology (Shubin et al. 2009), in various lineages of the N-fixing clade, the emerged predisposition has been largely retained in a non-N-fixing symbiosis role (e.g., stress response), and the non-homologous origins of N-fixing symbiosis represent convergent evolutionary events selected by the limitation of fixed nitrogen in many ecosystems (Doyle 2016). This is in accordance with the emergence of complex traits through incrementally key transitions (Oyserman et al. 2016), and with the extraordinary diversity and ecological success of plant species that obtain fixed nitrogen from N-fixing symbiosis (Lavin et al. 2005). Evolutionary reconstructions are normally based on observations, and not on the design and performance of experiments, and consequently, different scenarios for very early evolution events are, by nature, speculative (Cleland

✉ Gabriela Soto
  gabysoto80@gmail.com

1   Instituto de Genética Ewald A. Favret (IGEAF), De los Reseros S/N, Castelar C25(1712), Buenos Aires, Argentina

2   Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

3   Instituto de Química y Fisicoquímica Biológica (IQUIFIB), Buenos Aires, Argentina

and Xa 2002). Exceptionally, the predisposition hypothesis makes predictions that are testable and refutable. Specifically, this hypothesis indicates that it is possible to identify a cryptic evolutionary innovation that drives symbiotic N-fixation evolution by genomic comparisons (Werner et al. 2014), and to then experimentally recreate the acquisition of N-fixing symbiosis by transfer of this genetic novelty to non-N-fixing plants, including major cereal crops. Therefore, the identification of key factors for the emergence of N-fixing symbiosis not only is able to bypass the typical epistemological limitations of evolutionary reconstructions, but also has large ecological and economic consequences.

During N-fixing symbiosis, rhizobia are differentiated from bacteroides, which are N-fixing ammonium-excreting forms of rhizobia (Jones et al. 2007). Bacteroides inside infected root cells are surrounded by a symbiosome membrane, derived from the plant cell plasma membrane, which effectively segregates bacteroides from the host cytoplasm and controls the movement of metabolites between the two symbiotic partners. Rapid movement of ammonium from the bacteroid to the plant cytoplasm is facilitated by the nodule-specific ammonium channel NOD26 (Hwang et al. 2010), the main protein component of the symbiosome membrane in soybean (*Glycine max*) (Fortin et al. 1987). Since ammonium is the principal beneficial compound obtained by plants during rhizobial symbiosis and because its rapid uptake by plants is the outcome of NOD26 activity, this channel is a good candidate to be a key molecular innovation for the emergence of N-fixing symbiosis.

Methods

Search of Protein Databases and Identification of NOD26-Related Proteins

AQP protein sequences of *Arabidopsis thaliana* (AtPIP1;1, AtTIP1;1, AtNIP1;1, and AtSIP1;1) and *Physcomitrella patens* (PpGIP1;1 and PpXIP1;1) were used as query to search against genome databases, including Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), PlantGDB (http://www.plantgdb.org/), NCBI (https://www.ncbi.nlm.nih.gov/), PLAZA (https://bioinformatics.psb.ugent.be/plaza/) by using BLASTP (amino acid identity cut-off ≤ 15%). Thirty green plant genomes, including genomes from unicellular organisms (*Volvox carteri, Coccomyxa subellipsoidea, Ostreococcus lucimarinus*, and *Chlamydomonas reinhardtii*) as well as genomes from multicellular eukaryotes (*Glycine max, Phaseolus vulgaris, Medicago truncatula, Lotus japonicus, Cucumis sativus, Fragaria vesca, Prunus persica, Malus domestica, Ricinus communis, Morchella esculenta, Populus trichocarpa, Linum usitatissimum, Arabidopsis thaliana, Solanum tuberosum, Theobroma cacao,*

*Vitis vinifera, Paspalum virgatum, Scilla italica, Zea mays, Oryza sativa, Brachypodium distachyon, Picea abies, Picea sitchensis, Cryptomeria japonica, Selaginella moellendorffii,* and *Physcomitrella patens*), were explored for the presence of AQPs. Proteins that were not identified by the BLASTP analysis but are within the genomes of these green plants and contain the PF00230 domain (https://pfam.xfam.org/family/PF00230) were also selected.

Synteny, Phylogenetic, and Expression Analyses

Genomic synteny comparisons among chromosomes of plant genomes were performed for dicot species, including species belonging to the N-fixing clade (*Lotus japonicus, Medicago truncatula, Phaseolus vulgaris, Glycine max, Arachis ipaensis, Parasponia andersonii, Trema orientalis,* and *Cucumis sativus*), as well as *Ricinus communis* (a close-related species to N-fixing clade), and the model plant *Arabidopsis thaliana*, by using the LASTZ plugin in the Geneious v10.1 software (Kearse et al. 2012). A strict criterion to identify homologous proteins among chromosomes was used (> 50% of amino acid identity). Protein identities were calculated using MatGAT (Campanella et al. 2003). AQPs were phylogenetically and molecularly analyzed using MEGA version 7.0. (https://www.megasoftware.net/). Protein sequences were aligned using the ClustalW program. Phylogenetic trees were constructed using the Neighbor-Joining (NJ) method with genetic distances computed using the *p* distance model and bootstrap analysis of 500 resamples and root on midpoint. Microarray gene expression data were collected from Mt-Atlas (Benedito et al. 2008), Gm-Atlas (Libault et al. 2010), LjGEA (Verdier et al. 2013), PvGEA (O'Rourke et al. 2014), Ai-Atlas (Chopra et al. 2014), and Pa-Atlas (van Velzen et al. 2018).

Results and Discussion

NOD26 is a member of the aquaporin (AQP) superfamily, a group of highly diversified channel proteins that facilitate the bidirectional transport of water, small solutes, ions, and/or dissolved gases across cell membranes in passive manner (Perez Di Giorgio et al. 2014). To begin to analyze the origin of NOD26, in the present study, we examined the phylogenetic relationships of AQPs from 30 whole genome-sequenced species of plants representing major branches of the plant kingdom. As expected for being the archetype, NOD26 clustered within the NIP (NOD26-like intrinsic proteins) AQP subfamily (Fig. 1), and not into the TIP, PIP, XIP, SIP, and GIP AQP subfamilies. In addition, NOD26 showed high phylogenetic affinity with the angiosperm-exclusive subgroup NIP 1 (Fig. 1), which is consistent with the need to have NOD26 to evolve N-fixing symbiosis. Thus,
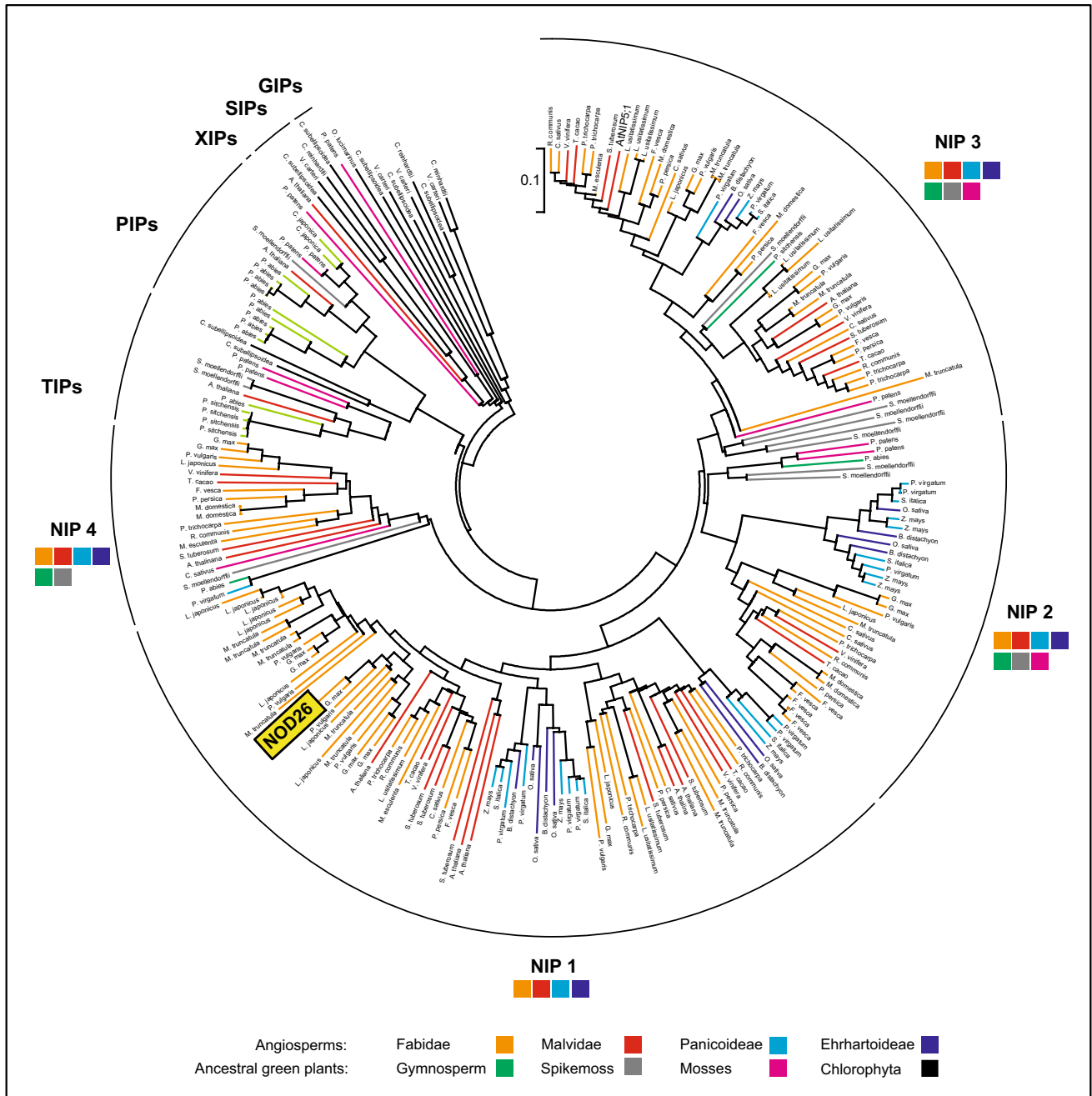
Fig. 1 NOD26 belongs to NIP 1, an angiosperm-exclusive subgroup of the NIP AQP subfamily. Large-scale phylogenetic analysis of NOD26 from *Glycine max* (boxed) based on the NJ method. The GIP, SIP, XIP, PIP, TIP, and NIP subfamilies of AQPs with focus on the classical NIP 1 to NIP 4 subgroups of NIPs (bold types) and the taxonomy of species (different colors) are shown. Angiosperms: Fabids (orange): *Lotus japonicus, Medicago truncatula, Phaseolus vulgaris, Glycine max, Cucumis sativus, Ricinus communis, Fragaria vesca, Prunus persica, Malus domestica, Morchella esculenta, Populus trichocarpa,* and *Linum usitatissimum*; Malvidae (red): *Arabidopsis thaliana, Solanum tuberosum, Theobroma cacao,* and *Vitis vinifera*; Panicoideae (light blue): *Paspalum virgatum, Scilla italic,* and *Zea mays*; Ehrhartoideae (blue): *Oryza sativa* and *Brachypodium distachyon*; Ancestral green plants: Gymnosperms (green): *Picea abies, Picea sitchensis,* and *Cryptomeria japonica*; Spikemosses (gray): *Selaginella moellendorffii*; Mosses (fuchsia): *Physcomitrella patens*; and Chlorophyta (black): *Volvox carteri, Coccomyxa subellipsoidea, Ostreococcus lucimarinus*, and *Chlamydomonas reinhardtii*. (Color figure online)

N-fixing symbiosis might not have emerged in ancestral green plants because these plants did not have NOD26.

To better understand when NOD26 appeared into the angiosperm lineage, we restricted the evolutionary studies to well-characterized species of the Fabids clade and the model species *Arabidopsis thaliana* belonging to the Fabids sister clade Malvids (Fig. 2). Regarding the Fabids clade, we selected both N-fixing species (*Lotus japonicus, Medicago truncatula, Phaseolus vulgaris, Glycine max, Arachis ipaensis*, and *Parasponia andersonii*) and non-N-fixing species (*Trema orientalis* and *Cucumis sativus*) from the N-fixing clade, and their related species *Ricinus communis* from the Malpighiales (Fig. 2). N-fixing species were selected based on the availability of both robust genomic sequences and global expression analyses which endorsed the integration of bioinformatic and empirical studies. Comparison of the nucleotide sequence of the region around the NOD26 (Glyma08g12650) of *G. max* with other genomes of phylogenetically related plant species showed a conserved synteny without large-scale DNA rearrangements (Fig. 3). This chromosomal region, which we named the Biological Nitrogen Fixation (BNF) region, is duplicated in *L. japonicus, M. truncatula, P. vulgaris,* and *A. ipaensis*, and quadrupled in *G. max*, which is in accordance with ancestral and recent events of genome duplications in the ancestors of Legumes and *G. max*, respectively (Bertioli et al. 2009; Schmutz et al. 2010). In addition to the increase in the NOD26 copy number (paralogous genes) in Legumes via block duplication events, tandem duplications of NOD26 were observed in the BNF regions of species belonging to the families Legumes (*L. japonicus, M. truncatula, P. vulgaris, G. max,* and *A. ipaensis*) and Rosales (*P. andersonii* and *T. orientalis*), but not in *C. sativus*, which belongs to the family Cucurbitales, and control species *R. communis* and *A. thaliana* (Fig. 3). In agreement with the predisposition hypothesis, the most parsimonious explanation for the synteny results is the occurrence of a single AQP in the BNF region in ancestral angiosperms, the emergence of NOD26 in the N-fixing clade by a tandem duplication event, the amplification of NOD26 in the N-fixing clade by both block and tandem duplication events, and the loss of NOD26 in the ancestor of *C. sativus*. Analogously, we have recently reported the emergence of the pollen-specific AQPs NIP4;1 and NIP4;2 in angiosperms by a tandem duplication event, which is associated with accelerated pollen tube growth rates (Di Giorgio et al. 2016), a critical novelty for the rapid diversification of angiosperms (Williams 2008).

Although the synteny analysis is consistent with the predisposition hypothesis, an integrated phylogenetic and expression analysis is important to support the origin of NOD26 in the N-fixing clade, and necessary to better understand the diversification of NOD26 within Legumes. In line with this, we examined the expression and evolution patterns of all paralogous copies of NOD26 included in the BNF regions of N-fixing species (Table 1; Fig. 4). The expression analysis showed that at least one copy of NOD26 is highly upregulated in inoculated roots or nodules of Legumes (LJ4g029020 in *L. japonicus*, Medtr8g087710 and Medtr8g087720 in *M. truncatula*, Phvul.002G242300 in *P. vulgaris*, Glyma08g12650 in *G. max,* and LOC107621874 in *A. ipaensis*) and non-Legume (PanWU01×14_248030 in *P. andersonii*) N-fixing species, supporting an essential and ancestral role of NOD26 in N-fixing symbiosis (Table 1). Considering that these active copies of NOD26 are in different localizations within BNF regions in Legume species (Fig. 3), we inferred the coexistence of different active copies of NOD26 in the ancestor of Legumes, and the inactivation of most of these copies during the evolution of the N-fixing clade. On the other hand, refined phylogenetic analysis supported the emergence of NOD26 in the N-fixing clade, and clarified the order of NOD26 duplications in Legumes (Fig. 4). Specifically, the phylogenetic tree suggests that the
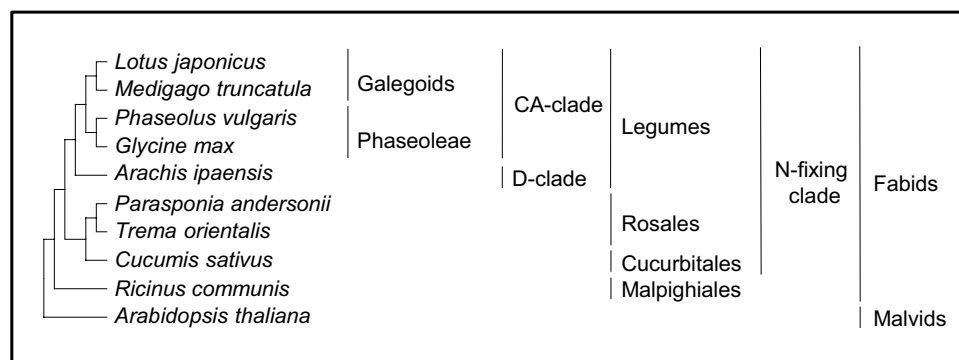
Fig. 2 Schematic representation of the phylogenetic relationship among angiosperm species. This scheme shows the largely accepted evolutionary relationship among different angiosperm species, where the monophyletic N-fixing clade, which contains both N-fixing (*Lotus japonicus, Medicago truncatula, Phaseolus vulgaris, Glycine max, Arachis ipaensis*, and *Parasponia andersonii*) and non-N-fixing species (*Trema orientalis* and *Cucumis sativus*), is closely related to *Ricinus communis* and related to the model species *Arabidopsis thaliana*
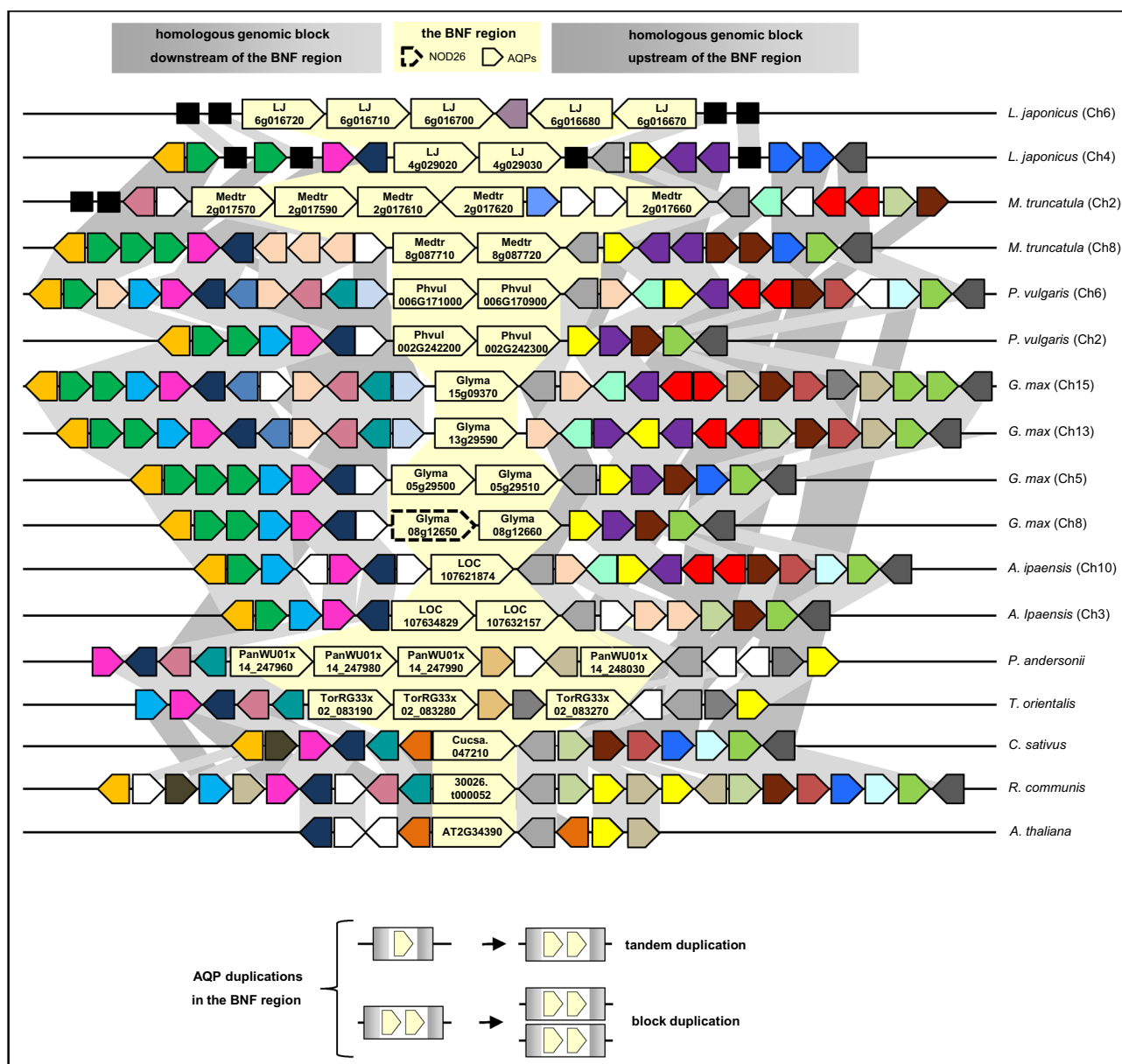
Fig. 3 Emergence of NOD26 in the N-fixing clade and its large expansion in Legumes. The NOD26 from *Glycine max* is situated within a conserved region of the angiosperm genome. This genomic region, which is particularly expanded in Legumes by duplicated blocks on the different chromosomes (Ch6 and Ch4 in *Lotus japonicus*; Ch2 and Ch8 in *Medicago truncatula*; Ch6 and Ch2 in *Phaseolus vulgaris*; Ch15, Ch13, Ch5, and Ch8 in *Glycine max*; and Ch10 and Ch3 in *Arachis ipaensis*), contains only one AQP in control species related to the N-fixing clade (*Ricinus communis* and *Arabidopsis thaliana*), and large tandem duplications of AQPs in all species belonging to the N-fixing clade (*Lotus japonicus, Medicago truncatula, Phaseolus vulgaris, Glycine max, Arachis ipaensis, Trema orientalis,* and *Parasponia andersonii*), with the exception of the non-N-fixing species *Cucumis sativus*. Orthologous genes are shown in the same color (orange = Pfam:00810; green = Pfam:02469; light blue = Pfam:04564; pink = Pfam:04844; dark blue = Pfam:00447; light yellow = Pfam:00230; white = hypothetical protein; gray = Pfam:06775; yellow = Pfam:11721; purple = Pfam:01031; brown = Pfam:00854; very light blue = Pfam:01680; light green = Pfam:01486; dark gray = Pfam:00657; olive green = Pfam:01823; blue = Pfam:07823; light orange = Pfam:00931; very light green = Pfam:01398; dark green = Pfam:03031; dark pink = Pfam:01363; ocean blue = Pfam:05915; cinnamon = Pfam:00069; red = Pfam:00012; light gray = Pfam:00560; light brown = pfam09262). The orthologous block of genes is represented with gray shadows. Arrows and squares represent gene orientation and transposons, respectively. The genomic blocks containing NOD26 and their related AQPs, which is named the BNF (Biological Nitrogen Fixation) region, are shown in light yellow, whereas homologous genomic blocks upstream and downstream of the BNF region are shown in gray. In the bottom panel, schematic representations of both block and tandem duplications are shown. (Color figure online)

Table 1 Expression of AQPs within the BNF region in different tissues of N-fixing plants

	Locus	Seed	Flower	Pod	Petiole	Shoot	Leaf	uninoc.root	inoc.root	Nodule
L. japonicus	LJ6g016720	192	862	1721	1757	1175	534	59	70	43
	LJ6g016710	22	319	25	26	29	31	30	42	31
	LJ6g016700	156	6773	174	559	488	309	51	43	27
	Lj6g016680	1435	507	976	15	14	14	13	14	14
	Lj6g016670	22	319	25	26	29	31	30	42	31
	LJ4g029020	19	16	22	20	17	18	17	19	*153*
	LJ4g029030	14	14	14	12	16	14	13	13	14
M. truncatula	Medtr2g017570	520	358	162	482	611	299	28	61	69
	Medtr2g017590	79	39	26	74	26	162	72	37	20
	Medtr2g017610	520	358	162	482	611	299	28	61	69
	Medtr2g017620	79	39	26	74	26	162	72	37	20
	Medtr2g017660	11	14	32	26	20	49	83	47	15
	Medtr8g087710	18	130	233	725	372	1351	335	452	*19,156*
	Medtr8g087720	7192	344	449	124	119	257	1863	1004	*18,568*
P. vulgaris	Phvul.006G171000	19	ND	171	ND	1133	155	109	825	118
	Phvul.006G170900	28	ND	7	ND	67	14	11	26	3
	Phvul.002G242200	16	ND	8	ND	19	4	30	68	18
	Phvul.002G242300	60	ND	2	ND	7	2	139	*342*	54
G. max	Glyma15g09370	ND	11	ND	ND	ND	16	ND	2	<1
	Glyma13g29590	ND	48	ND	ND	ND	30	ND	4	<1
	Glyma05g29500	ND	1	ND	ND	ND	<1	ND	3	13
	Glyma05g29510	ND	2	ND	ND	ND	<1	ND	4	2
	Glyma08g1265 (NOD26)	ND	<1	ND	ND	ND	<1	ND	90	*3686*
	Glyma08g12660	ND	2	ND	ND	ND	<1	ND	2	<1
A. ipaensis	*LOC107621874*	<1	2	<1	ND	<1	<1	36	ND	*119*
	LOC107634829	4	<1	<1	ND	<1	<1	<1	ND	<1
	LOC107632157	1	1	2	ND	4	2	2	ND	1
P. andersonii	PanWU01×14_247960	ND	ND	ND	ND	ND	ND	<0.1	<0.1	<0.1
	PanWU01×14_247980	ND	ND	ND	ND	ND	ND	<0.1	<0.1	<0.1
	PanWU01×14_247990	ND	ND	ND	ND	ND	ND	<0.1	<0.1	<0.1
	PanWU01×14_248030	ND	ND	ND	ND	ND	ND	<0.1	0.47	*6.21*

NOD26 copies highly upregulated in inoculated roots or nodules are in italics

All values are the mean of three to six independent experiments

ND not determined

block duplication event (originating clusters G1.1.1 and G1.1.2) preceded the tandem duplication event (originating clusters G1.1.2A and G1.1.2B), indicating that genome duplication in the common ancestor of Legumes could be critical for the consolidation of a stable N-fixing precursor. In addition, clusters G1.1.2A and G1.1.2B exhibited a complete congruence with the organismal tree at species levels, whereas cluster G1.1.1 showed congruence only with the evolution of the CA- and D-clades of Legumes (Fig. 4). This partial congruence pattern has been described in other AQP clusters and is normally the consequence of the differential and large duplications of paralogous copies within different plant species or plant lineages (Abascal et al. 2014). In the case described here, the presence of several transposons next

to the AQP grouping within cluster G1.1.1 (Fig. 3) suggests that the high rate of duplication of these AQPs could be mediated by these mobile elements. Figure 5 schematically summarizes the integration of syntenic, gene expression, and phylogeny analyses of all the AQPs localized within the BNF region in species of the N-fixing clade and in their phylogenetically related species *R. communis* and *A. thaliana*, focusing on the emergence of NOD26 in the N-fixing clade and highlighting the prediction of at least three functional copies of NOD26 in the ancestor of Legumes.

After defining that NOD26 emerged in the N-fixing clade via tandem duplication and was consolidated in Legumes via block and tandem duplications, and under the premise that this symbiotic ammonium channel has an angiosperm origin
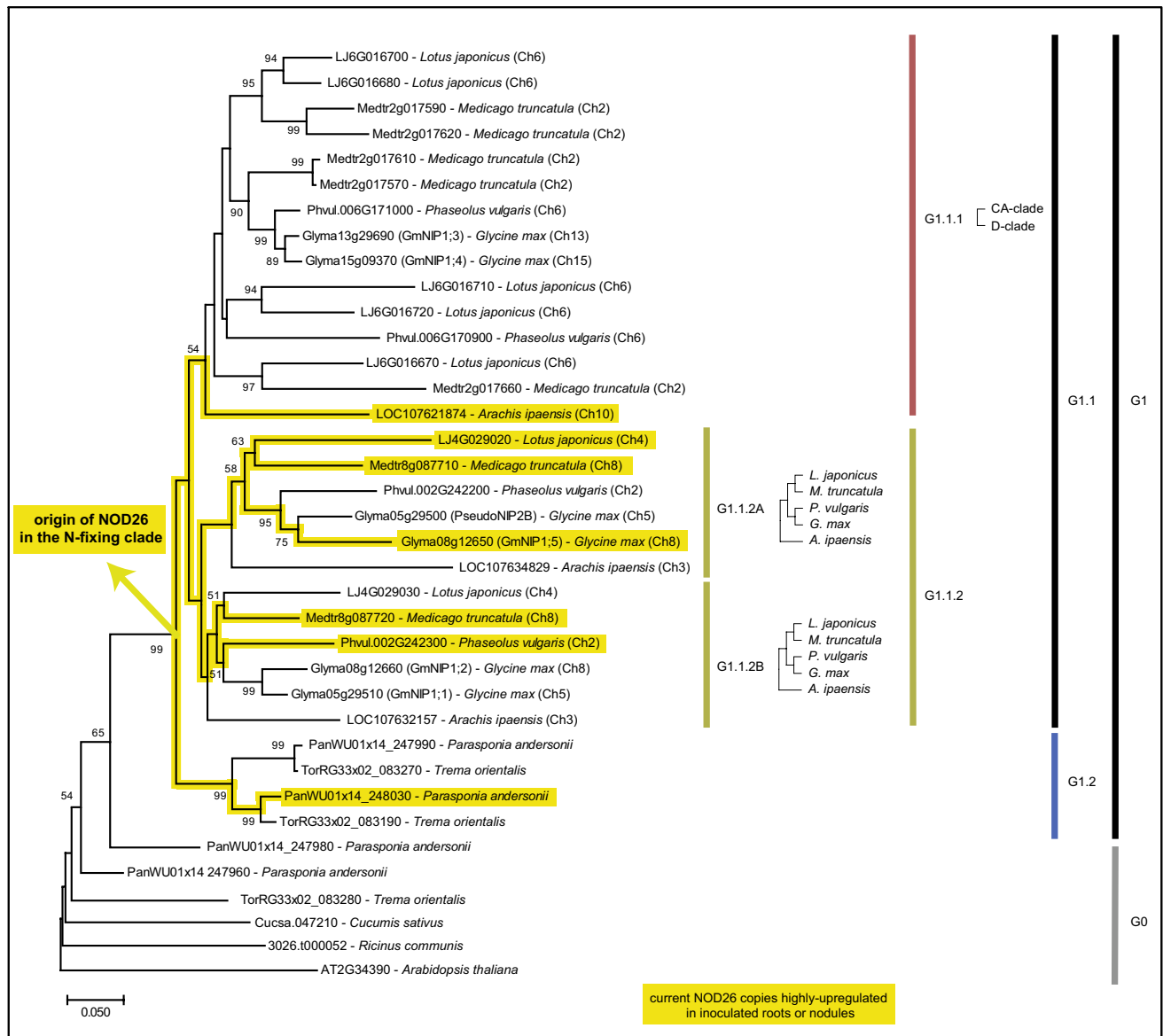
Fig. 4 Origin of NOD26 in the N-fixing clade and its complex evolution in Legumes. Phylogeny of all AQPs included in the BNF region from angiosperms based on the NJ method. Bootstrap percentages are indicated at the branch points (only values above 50% are indicated). Current NOD26 copies highly upregulated in inoculated roots or nodules in N-fixing species (LJ4g029020 in *L. japonicus*, Medtr8g087710 and Medtr8g087720 in *M. truncatula*, Phvul.002G242300 in *P. vulgaris*, Glyma08g12650 in *G. max*, and LOC107621874 in *A. ipaensis* and PanWU01×14_248030 in *P. andersonii*, see Table 1) and branches connecting these sequences are

highlighted, showing the origin of NOD26 in the N-fixing clade. The tree topology shows that the increase in the NOD26 copy number in Legumes occurred firstly by block duplication (clusters G.1.1.1 and G.1.1.2) and secondly by tandem duplication (clusters G.1.1.2A and G.1.1.2B). A congruent pattern between orthologous clusters and organisms (see Fig. 2) is found on the right. Whereas clusters G.1.1.1 and G.1.1.2 showed a complete congruence with the organismal tree, cluster G.1.1.1 showed congruence only with the CA- and D-clades, which is in accordance with the particular expansion of AQPs in cluster G.1.1.1 probably mediated by mobile elements (see Fig. 3)

related to non-symbiotic plant-beneficial functions, we here recreated the genomic and functional transitions that can lead to the emergence of the original N-fixing precursor and its stabilization in Legumes (Fig. 6). Under aerobic conditions, root respiration provides the proton-motive force for the flux of large solutes, including sugars and nitrogen compounds (Fig. 6a). Under anaerobic conditions, angiosperm

roots maintain some ATP synthesis through catabolic processes, including lactate and ammonium fermentations (Fig. 6a). Thus, the expression of genes involved in this type of anaerobic processes, such as lactate dehydrogenase (Shingaki-Wells et al. 2014) and nitrate reductase (Allegre et al. 2004), is a critical factor for the tolerance of plants to root anoxia. However, under long-term root anoxia, fermentation
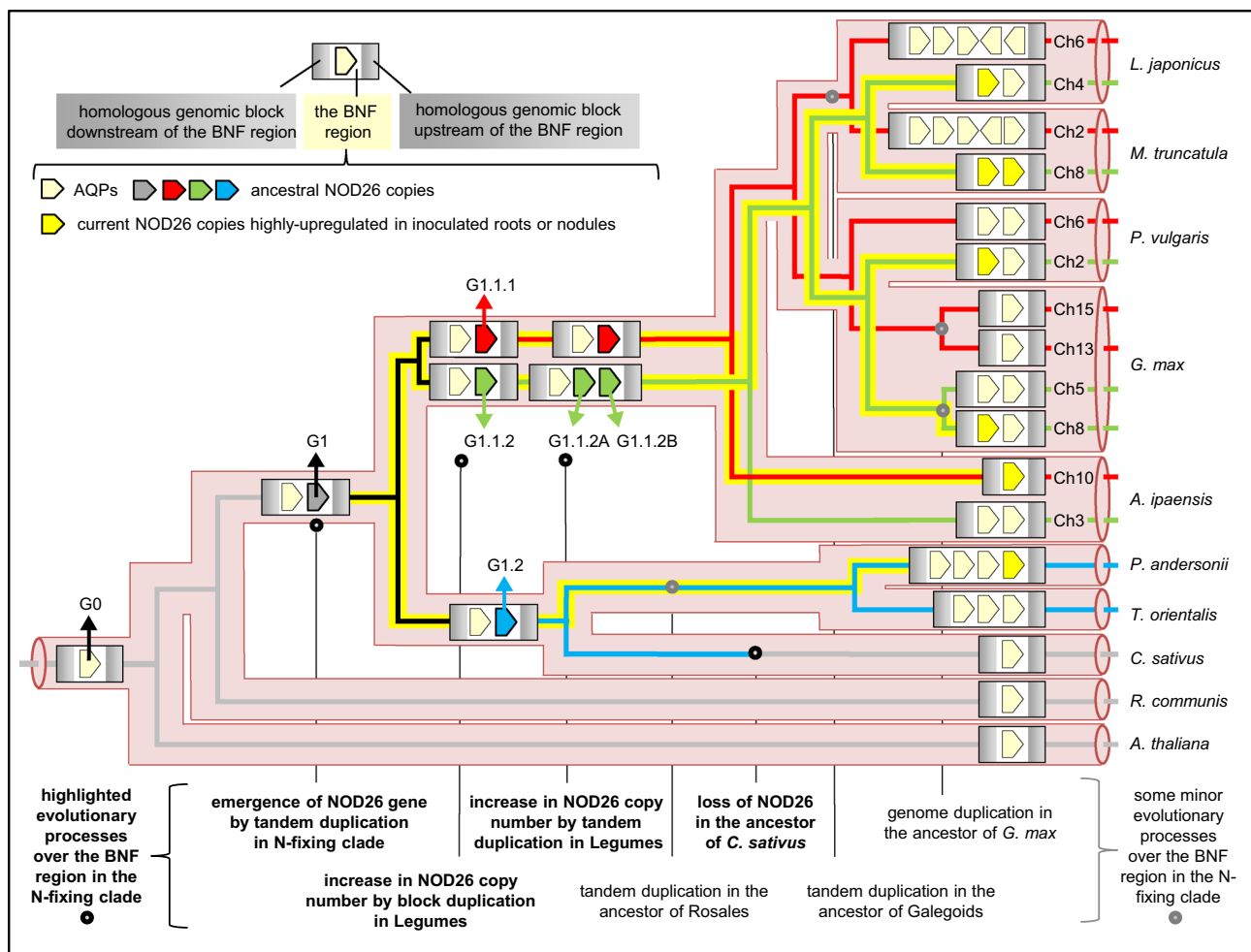
Fig. 5 Schematic representation of the evolution of angiosperm species and their BNF region containing different aquaporins. The integration of synteny (see Fig. 3), gene expression (see Table 1), and phylogeny (see Fig. 4) analyses of AQPs within the BNF region from angiosperm species suggests the origin of NOD26 in the N-fixing clade by tandem duplication, a particular expansion of NOD26 in legumes by both block and tandem duplications, and the loss of NOD26 in the non-N-fixing species *Cucumis sativus*. Similar to that shown in Fig. 4, current NOD26 copies highly upregulated in inoculated roots or nodules in N-fixing species and branches connecting these sequences are highlighted

products are accumulated in the cytoplasm, reducing the rate of glycolysis and further decreasing ATP production, which can cause an energy crisis and consequently cell death. These lethal effects can be accelerated by the intrinsic toxicity of ammonium on cytoplasmic enzymes (Esteban et al. 2016) (Fig. 6a).

During the energy crisis induced by anoxia, the cost of the active export of fermentation products is unsustainable, and, consequently, a selection pressure is exerted for the emergence of passive channels (such as AQPs) induced in anaerobic root cells able to facilitate the diffusion of these reduced compounds in favor of the concentration gradient (from the cytoplasm to the extracellular space). Arabidopsis AQP AtNIP2;1, a lactate channel specifically induced in root cells during anoxic stress (Choi and Roberts 2007), and the ammonium channel NOD26 probably emerged under this

selective pressure, improving the NAD$^+$ recycling necessary to continue with the glycolysis process (Fig. 6a, b). Interestingly, the gene that codifies for AtNIP2;1 (AT2G34390) is within the BNF region (Fig. 3), and, consequently, AtNIP2;1 is the most well-characterized orthologous protein of NOD26 in angiosperms. Considering that the phylogenetic analysis supported the emergence of AtNIP2;1 before that of NOD26 (Fig. 4), we can explain the retention, long-term persistence, and conservation of NOD26 in the N-fixing clade as a mechanism to maximize the amount of recycled NAD$^+$, improving the tolerance to anaerobic stress (Fig. 6b). It is important to note that the emergence of an ammonium channel induced in anaerobic roots, i.e., the critical evolutionary innovation for the emergence of N-fixing plants proposed in this work, needs few evolutionary events, which consist of a tandem duplication of an AQP (including its promoter)
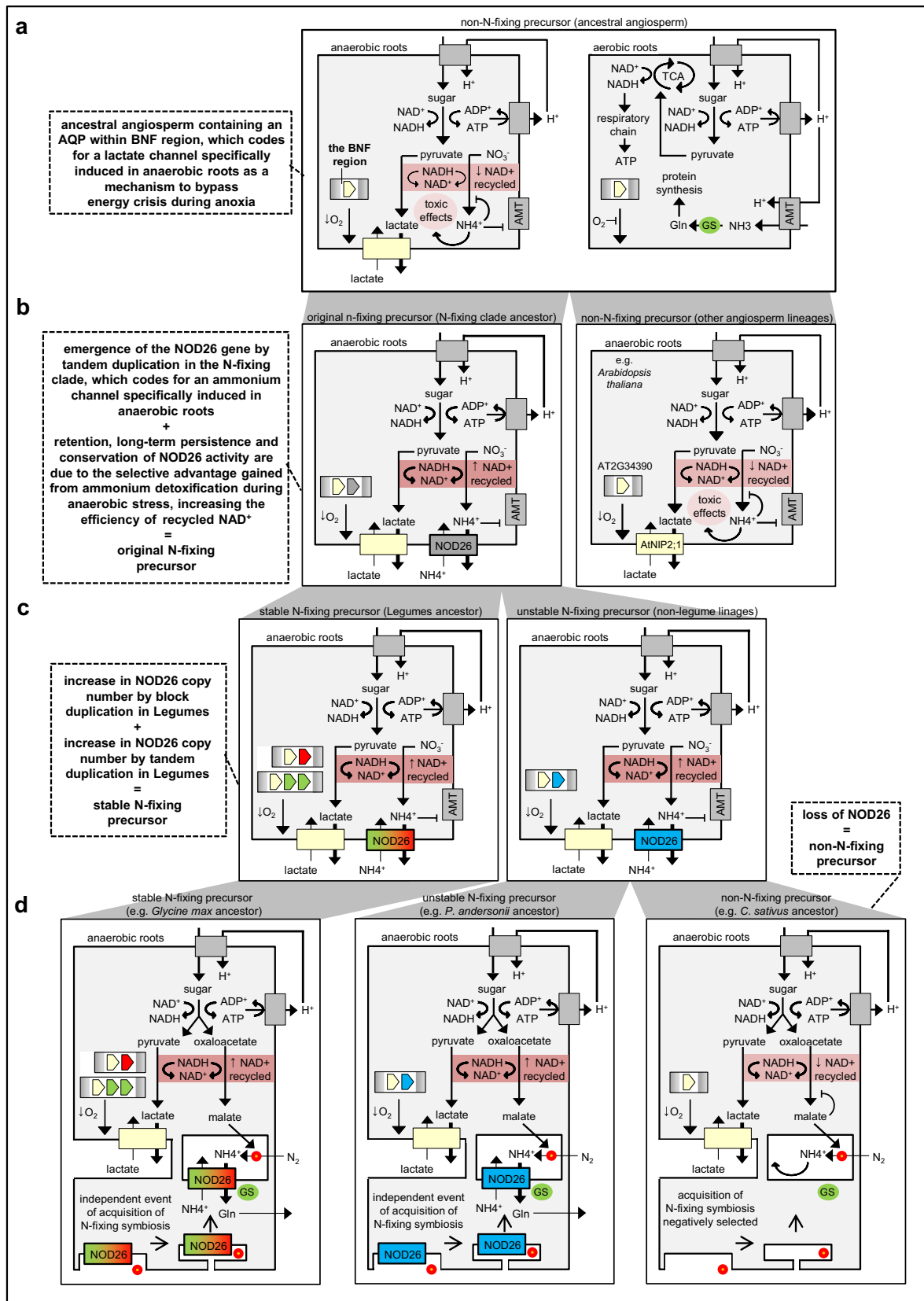
Fig. 6 Metabolic reconstructions of the origin of the N-fixing precursor in angiosperms and the hypothetical ancestral non-symbiotic role of NOD26 in ammonium detoxification in anoxic roots. This picture is largely explained in the text

and low amino acid (80% similar) and functional (small protonated solutes) divergences between homologous protein AtNIP2;1 and NOD26 in the ancestor of the N-fixing clade (Fig. 6b). In addition, AtNIP2;1 and NOD26 show the same pattern of transport activation induced by acidification (e.g., from pH 7.6 to pH 5.0–6.0) (Choi and Roberts 2007), which is consistent with the rapid cytoplasmic acidification during anoxic stress and the relevance of exporting fermented products under this stress condition (Felle 2005).

Returning to the general aspects of the evolutionary reconstruction of N-fixing symbiosis genesis, the evolution of the original N-fixing precursor (Fig. 6b) would have resulted in stable (Legumes) and unstable (non-legumes) N-fixing precursors, which differ mainly in the number of copies of NOD26 (Fig. 6c). Specifically, the ancestor of Legumes increased the number of copies of NOD26, thus improving the stability of this key genetic preadaptation for the emergence of N-fixing symbiosis (Fig. 6c, d), whereas non-legume lineages maintained only the original number or a low number of copies of NOD26, and therefore, were more susceptible to lose this N-fixing symbiosis prerequisite by chance (Fig. 6c, d). Lastly, different lineages from the N-fixing clade that retained at least one copy of NOD26 were able to acquire N-fixing symbiosis independently, whereas this acquisition is negatively selected in plants from the N-fixing clade that lost NOD26 (Fig. 6d). In this context, the probable induction of cell division by ancient N-fixing symbionts (Geurts et al. 2016), with the consequent decrease in free oxygen in root cells (Ott et al. 2005), possibly played a critical role in the initial induction of NOD26 expression.

Considering that bacterial N-fixation is an anaerobic process and that common free-living N-fixing strains do not release the fixed nitrogen to the environment, the emergence of the current and well-known ammonium-excreting N-fixing symbiotic rhizobia seems to need the same single evolutionary innovation as their plant hosts: the acquisition of an efficient ammonium channel. Horizontal gene transfer is the main mechanism of bacterial evolution, and genes related to the N-fixation process in rhizobia are normally within conjugative elements able to convert non-symbiotic strains into N-fixing symbionts (Haskett et al. 2016). For example, the genome of the alfalfa-symbiont model strain *Sinorhizobium meliloti* 1021 includes three replicons, consisting of a circular chromosome (SMc) encoding housekeeping functions, the rhizosphere colonization-related megaplasmid B (SMb), and the N-fixation-related megaplasmid A (SMa) (Galardini et al. 2013). Within the SMa replicon, we found an AQP-like protein produced by SMa0627, which is very similar to NOD26 (68% of amino acid similarity) and is next to *fix* genes (SMa0612–SMa0626). Considering that *fix* genes are essential factors for N-fixing symbiosis (Bobik et al. 2006) and that bacterial genes involved in the same pathway are normally co-localized in the genome, SMa0627 is a natural candidate to promote the release of ammonium from bacteria to plants during N-fixing symbiosis.

In biotechnological terms, the results here presented suggest that the NOD26 from N-fixing plants, and maybe the NOD26-like AQPs from their N-fixing symbiont, could be useful tools to improve N-fixation in N-fixing plants and to transfer N-fixation to non-N-fixing plants such as cereals. Constitutive expression of active ammonium channels in almost any microorganism or plant tissue, including rhizobial and root cells, has probably a strong negative fitness. Consequently, ammonium channels for microbial and plant engineering could be used with specific anaerobic-induced promoters, such as the model non-N-fixing plant Arabidopsis expressing NOD26 from *G. max* under the control of the AtNIP2;1 promoter from Arabidopsis. Since there is a strong framework for the deregulation of transgenic crops, these types of transgenic plants with greater capacity for ammonium interchange under anaerobic conditions could be used to improve the uptake of ammonium produced by N-fixing bacteria in future commercial crops. Contrary to sexual organisms like animals and plants, which exhibit a robust containment of transgene flow, microorganisms show extremely promiscuous gene flow, and consequently, there are no genetically modified bacteria in the market. However, domesticated mutant plant-growth-promoting strains are commonly accepted as inoculants to improve crop production. In this framework, and considering that the substitution of a single amino acid residue can alter the transport profiles of some AQPs (Azad et al. 2012; Krenc et al. 2014), we propose bacterial genome editing with CRISPR-Cas9 over current AQPs to mimic the ammonium channel, as a strategy to produce novel ammonium-excreting N-fixing non-transgenic inoculants to improve the nitrogen content in legumes and cereal crops. Finally, future characterizations of transgenic plants and edited N-fixing bacteria overexpressing ammonium channels in roots under anoxia can contribute not only to improve crop production but also to empirically recapitulate the origin of N-fixing symbiosis.

References

Abascal F, Irisarri I, Zardoya R (2014) Diversity and evolution of membrane intrinsic proteins. Biochim Biophys Acta 1840:1468

Allegre A, Silvestre J, Morard P, Kallerhoff J, Pinelli E (2004) Nitrate reductase regulation in tomato roots by exogenous nitrate: a possible role in tolerance to long-term root anoxia. J Exp Bot 55:2625

Azad AK, Yoshikawa N, Ishikawa T, Sawa Y, Shibata H (2012) Substitution of a single amino acid residue in the aromatic/arginine selectivity filter alters the transport profiles of tonoplast aquaporin homologs. Biochim Biophys Acta 1818:1

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK (2008) A gene expression atlas of the model legume *Medicago truncatula*. Plant J 55:504

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimaraes PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM, Sato S, Tabata S, Cannon SB, Stougaard J (2009) An analysis of synteny of Arachis with Lotus and Medicago sheds new light on the structure, stability and evolution of legume genomes. BMC Genom 10:45

Bobik C, Meilhoc E, Batut J (2006) FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of Sinorhizobium meliloti. J Bacteriol 188:4890

Campanella JJ, Bitincka L, Smalley J (2003) MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinform 4:29

Choi WG, Roberts DM (2007) Arabidopsis NIP2;1, a major intrinsic protein transporter of lactic acid induced by anoxic stress. J Biol Chem 282:24209

Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Burow MD (2014) Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-Seq data. PLoS ONE 9:e115055

Cleland C, Xa E (2002) Methodological and epistemic differences between historical science and experimental science*. Philos Sci 69:447

Di Giorgio JA, Bienert GP, Ayub ND, Yaneff A, Barberini ML, Mecchia MA, Amodeo G, Soto GC, Muschietti JP (2016) Pollen-specific aquaporins NIP4;1 and NIP4;2 are required for pollen development and pollination in *Arabidopsis thaliana*. Plant Cell 28:1053

Doyle JJ (2011) Phylogenetic perspectives on the origins of nodulation. Mol Plant Microbe Interact 24:1289

Doyle JJ (2016) Chasing unicorns: nodulation origins and the paradox of novelty. Am J Bot 103:1865

Esteban R, Ariz I, Cruz C, Moran JF (2016) Review: mechanisms of ammonium toxicity and the quest for tolerance. Plant Sci 248:92

Felle HH (2005) pH regulation in anoxic plants. Ann Bot 96:519

Fortin MG, Morrison NA, Verma DP (1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. Nucleic Acids Res 15:813

Galardini M, Pini F, Bazzicalupo M, Biondi EG, Mengoni A (2013) Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. Genome Biol Evol 5:542

Geurts R, Xiao TT, Reinhold-Hurek B (2016) What does it take to evolve a nitrogen-fixing endosymbiosis? Trends Plant Sci 21:199

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, Roswanjaya YP, Kohlen W, Pujic P, Battenberg K, Alloisio N, Liang Y, Hilhorst H, Salgado MG, Hocher V, Gherbi H, Svistoonoff S, Doyle JJ, He S, Xu Y, Xu S, Qu J, Gao Q, Fang X, Fu Y, Normand P, Berry AM, Wall LG, Ane JM, Pawlowski K, Xu X, Yang H, Spannagl M, Mayer KFX, Wong GK, Parniske M, Delaux PM, Cheng S (2018) Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science https://doi.org/10.1126/science.aat1743

Haskett TL, Terpolilli JJ, Bekuma A, O'Hara GW, Sullivan JT, Wang P, Ronson CW, Ramsay JP (2016) Assembly and transfer of tripartite integrative and conjugative genetic elements. Proc Natl Acad Sci USA 113:12268

Hwang JH, Ellingson SR, Roberts DM (2010) Ammonia permeability of the soybean nodulin 26 channel. FEBS Lett 584:4339

Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: the Sinorhizobium–Medicago model. Nat Rev Microbiol 5:619–633

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647

Krenc D, Song J, Almasalmeh A, Wu B, Beitz E (2014) The arginine-facing amino acid residue of the rat aquaporin 1 constriction determines solute selectivity according to its size and lipophilicity. Mol Membr Biol 31:228

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575

Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD, Stacey G (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiol 152:541

O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX, Hernandez G, Vance CP (2014) An RNA-Seq based gene expression atlas of the common bean. BMC Genom 15:866

Ott T, van Dongen JT, Gunther C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P, Udvardi MK (2005) Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Curr Biol 15:531

Oyserman BO, Moya F, Lawson CE, Garcia AL, Vogt M, Heffernen M, Noguera DR, McMahon KD (2016) Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating bacteria. ISME J 10:2931

Perez Di Giorgio J, Soto G, Alleva K, Jozefkowicz C, Amodeo G, Muschietti JP, Ayub ND (2014) Prediction of aquaporin function by integrating evolutionary and functional analyses. J Membr Biol 247:107

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178

Shingaki-Wells R, Millar AH, Whelan J, Narsai R (2014) What happens to plant mitochondria under low oxygen? An omics review of the responses to low oxygen and reoxygenation. Plant Cell Environ 37:2260

Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. Nature 457:818

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci USA 92:2647

Svistoonoff S, Benabdoun FM, Nambiar-Veetil M, Imanishi L, Vaissayre V, Cesari S, Diagne N, Hocher V, de Billy F, Bonneau J, Wall L, Ykhlef N, Rosenberg C, Bogusz D, Franche C, Gherbi H (2013) The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. PLoS ONE 8:e64515

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, Roswanjaya Y, Wardhani TAK, Kalhor MS, Jansen J, van den Hoogen J, Gungor B, Hartog M, Hontelez J, Verver J, Yang WC, Schijlen E, Repin R, Schilthuizen M, Schranz ME, Heidstra R, Miyata K, Fedorova E, Kohlen W, Bisseling T, Smit S, Geurts R (2018) Comparative genomics of the nonlegume Parasponia reveals insights into evolution of

nitrogen-fixing rhizobium symbioses. Proc Natl Acad Sci USA 115:E4700

Verdier J, Torres-Jerez I, Wang M, Andriankaja A, Allen SN, He J, Tang Y, Murray JD, Udvardi MK (2013) Establishment of the Lotus japonicus Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. Plant J 74:351

Werner GD, Cornwell WK, Sprent JI, Kattge J, Kiers ET (2014) A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. Nat Commun 5:4087

Williams JH (2008) Novelties of the flowering plant pollen tube underlie diversification of a key life history stage. Proc Natl Acad Sci USA 105:11259